Addressing Methodological Issues in the Study of Tiger Metapopulation Dynamics in Western Ghats, India



Arjun Mallipatna Gopalaswamy

Balliol College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2014

This thesis is dedicated to
my dear sister, Lachu,
for bearing family responsibilities
and allowing me to pursue my interests
for over ten long years.

Abstract

The tiger *Panthera tigris* is a globally endangered large carnivore of high conservation importance. It is cryptic, wide-ranging and naturally occurs at low population densities. Any study of its metapopulation in real landscapes is faced with a suite of methodological constraints because the temporal and spatial scale of study is large. The scope presented for developing, testing and reviewing methods, whether they are analytical, practical or conceptual, in such a context is wide: ranging from statistical methods to laboratory techniques to field methodologies to software products and methods, all relevant to tiger research and conservation endeavours.

Framed within the ecological theme of assessing population dynamics of species within patchy environments, and with the source-sink view of tiger metapopulations, this thesis investigates the abundance estimation question of tigers and their prey at multiple scales within a 38,000 km^2 tiger landscape in the Western Ghats, India.

While the emphasis is firmly rooted in the development of statistical methods, software products and field methodologies, it also brings in research in laboratory techniques to bear on the abundance and distribution questions on hand. In specific, this thesis:

1. Advances our understanding of spatial capture-recapture methods - a statistical methodology being widely used over the last five years for density estimation.
2. Develops a new field-based method for estimating the density of tiger prey, specifically when they occur at low densities, using an occupancy-based approach. The results of this exercise show much promise and offers a way of tackling the abundance estimation question in unmarked, rare species.
3. Develops a software product called SPACECAP to make novel spatial capture-recapture methods accessible to wildlife biologists, ecologists and park managers. The software is adequately tested and provides users with the necessary summary statistics of density and

related parameters, along with the necessary diagnostic tools in order facilitate accurate interpretation of parameters.

4. Develops a Bayesian inferential approach to estimating tiger density in areas of low abundance by bringing together data sets from multiple sources (camera trapping images and faecal DNA samples, in our case) to strengthen inference about tiger density. Results from this approach suggests that it takes a relatively small increase in sampling effort to bring about large reductions in the variance of density estimates.

5. Investigates a long-standing controversy of index-calibration experiments at large scales. The theoretical models and the empirical testing with tiger sign-encounter data at macroecological scales demonstrate the relative futility in employing simplified and direct linear models using the R^2 statistics, because there are latent sampling process parameters which considerably weaken inference.

6. Develops a spatial capture-recapture model that facilitates investigation of unanswered questions lying at the interface between behavioural and population ecology of carnivores. This is done by introducing the attraction-repulsion spatial arrangements of carnivores into spatial capture-recapture models.

Nearly 70 percent of the world's tigers remain in less than 100,000 km^2 of habitat today. Yet, we lack sufficient methodological tools to reliably estimate their population dynamic parameters. This thesis provides a toolbox of advanced methods to make these assessments more reliable and, more importantly, accessible. It is envisioned that researchers and practical conservation managers will both benefit with this toolbox.

Acknowledgements

I acknowledge the support from Clarendon Fund, Jason Hu, WWF-Kathryn Fuller and WCS-Clive Marsh scholarships for supporting my doctoral work at the University of Oxford. I thank my supervisor David Macdonald for providing me with this great opportunity to study here at Oxford and for ensuring I crossed every hurdle that came my way. I thank Claudio Sillero-Zubiri, my co-supervisor, for ensuring that things were always on track with this thesis.I thank Ullas Karanth, who has demonstrated to me over the years as to what science-based conservation of tigers really means and has been a source of great inspiration. I thank Jim Nichols, whose breadth of knowledge in quantitative ecology, has shown me how to think about ecological problems. I thank Mohan Delampady, whose deep understanding of mathematical statistics, and who explained deep concepts with a lot of patience and played to vital back bone for the statistical elements in my thesis. I thank Andy Royle, whose mastery with spatial capture-recapture models and with R, made some implementations look like a breeze for me. I thank N. Samba Kumar, whose mastery over implementing large scale field projects at the "large carnivore scale", helped prompt some of the questions in this thesis. I thank Krithi. K. Karanth, who was responsible for providing me valuable insight into posing questions in the realm of human dimensions of wildlife conservation. I thank Seb Shimeld for being a very supportive college advisor.

I thank all research and field assistants of the Wildlife Conservation Society, India program, the Centre of Wildlife Studies, Bangalore for doing much of the work with the collection of data in trying circumstances in the field. I thank all members of WildCRU for providing me with a well-organized office to facilitate completion of this thesis. I thank the Balliol college whose excellent facilities made sure I had an incredible time over the past few years. I thank the University Counselling Centre for doing an excellent job during my lows.

Specifically, chapter-wise:

Chapter 2: I thank Andy Royle, Jim Nichols, Beth Gardner, Murray Efford and David Borchers for discussing SECR related ideas with me at various moments. I thank David Macdonald, K. Ullas Karanth and Claudio Sillero-Zubiri for discussing the biology of carnivores with me. I thank Mohan Delampady for statistical advice.

Chapter 3: We are grateful to the Wildlife Conservation Society, New York and for supporting our research work. Additionally we are indebted to the following donors for funding support: Save the Tiger Fund of National Fish and Wildlife Foundation, and, Rhinoceros and Tiger Conservation Fund of US Fish and Wildlife Service, Washington DC. In India, we acknowledge facilitation of the study by Karnataka State Forest Department, National Tiger Conservation Authority, New Delhi, and Centre for Wildlife Studies, Bangalore. We thank M. Delampady, Andy Royle and Jim Nichols for useful discussions and D. Jathanna, N. Patil and Raghavendra Mogaroy for analytical support. We thank V. Srinivas for GIS support. We thank the numerous field assistants and volunteers who enthusiastically participated in our field surveys. We thank the Associate Editor and two anonymous reviewers for providing helpful comments to the manuscript.

Chapter 4: We would like to thank Wildlife Conservation Society and Centre for Wildlife Studies for funding and administrative support. We would like to thank Patuxent Wildlife Research Center, USGS for administrative support. We thank Vishal Gokhale, Shreya Dasgupta, Rashmi Bhat, Sumanta Mukherjee, Deepthi Bharadwaj and Vatsaraj for programming assistance. We thank Mohan Delampady for statistical advice. We thank Raghavendra Mogaroy for testing the software on many capture-recapture data sets. We thank two anonymous reviewers and the associate editor for their suggestions to significantly improve this manuscript and the software product.

Chapter 5: We thank N. Samba Kumar and Anish Andheria for overseeing the camera trapping and fecal DNA field work, respectively. We thank Narendra Patil and Saw Htun for assistance in GIS. We thank Samrat Mondal and Uma Ramakrishnan for genotyping tiger individuals from fecal DNA samples. We thank Claudio Sillero-Zubiri for general support. We thank Evan Cooch and two anonymous reviewers for providing useful comments on an earlier version of this manuscript.

Chapter 6: We thank Wildlife Conservation Society, New York and Centre for Wildlife Studies, Bangalore for sharing some unpublished data

that helped frame the questions of this manuscript. We thank Bill Link, Jim Nichols, Krithi K. Karanth, Paul Johnson, Alex Braczkowski, Susana Rostro-Garcia, the Associate Editor and two anonymous reviewers for some helpful comments. DWM thanks Recanati-Kaplan Foundation for their support.

Chapter 7: I thank the Advanced Research Computing centre resources, University of Oxford, for facilitating computations during earlier development of this manuscript. I also thank Andy Royle and Jim Nichols. I thank Peter Diggle for some early ideas with spatial point processes.

Finally, I thank all my well wishers to ensure I completed this doctoral work. I thank Appa, Amma and Lachu for being an incredibly supportive family for whatever I chose to do over the last decade. I thank the Mallips and the Channarayapatnas. I thank Krithi Karanth for all the support especially during my time at the Patuxent Laboratories. I thank all my old friends, especially Faizal, Sachin, Mohan, Fantastic 4 - Krishna, Muneesh, Kanishka and Avish, and Valleyites whose collective knowledge of me as a person, lifted me when I faltered. I thank all my good friends in Balliol. And finally, I thank all those in WildCRU for pushing me over the final hurdle. A special thanks to Michelle, Nic, Alex B, Alex Z, Emre, Joelene and Susy.

List of Publications and Author Contributions

Table 1: Citations of published pieces of work from this thesis along with author contributions

In thesis	Citation	Authors	Contributions
Chapter 2	Gopalaswamy (2013)	AMG	Sole author
Chapter 3	Gopalaswamy *et al.* (2012b)	AMG	C, D, F, A, W
		KUK	C, F, W
		NSK	C, A, W
		DWM	W
Chapter 4	Gopalaswamy *et al.* (2012d)	AMG	C, D, P, A, W
		JAR	P, W
		JEH	P
		PS	P, A
		DJ	A, W
		NSK	A, W
		KUK	C, W
Chapter 5	Gopalaswamy *et al.* (2012c)	AMG	C, D, A, W
		JAR	D, A, W
		MD	D, A, W
		JDN	C, D, W
		KUK	D, F, W
		DWM	D, W
Chapter 6	Gopalaswamy *et al.* (In review)	AMG	C, D, S, F, A, W
		MD	D, S, A, W
		KUK	C, D, F, W
		NSK	D, F
		DWM	W
Appendix A	Hiby *et al.* (2009)	LH	C, D, A, W
		PL	C, D, A, W
		NP	D, A, F
		NSK	D, A, F
		AMG	D, A, F
		KUK	C, D, A, F, W
Appendix B	Mondol *et al.* (2009)	SM	C, G, A, W
		KUK	C, A, F, W
		NSK	F, W
		AMG	A, W
		AA	F
		UR	C, G, W

Continued on next page

Table 1 – continued from previous page

In thesis	Citation	Authors	Contributions
Appendix C	Royle *et al.* (2009b)	JAR	C, D, A, W
		JDN	C, D, A, W
		KUK	C, F, W
		AMG	A, W
Appendix D	Royle *et al.* (2009a)	JAR	C, D, A, W
		KUK	C, F, W
		AMG	A, W
		NSK	F, A, W
Appendix E	Singh *et al.* (2010)	PS	C, D, F, A, W
		AMG	C, D, F, A, W
		KUK	C, D, A, W
Appendix F	Hines *et al.* (2010)	JEH	C, D, A, W
		JDN	C, D, A, W
		JAR	D, A, W
		DIM	D, A, W
		AMG	D, A, F, W
		NSK	D, F, W
		KUK	D, F, W
Appendix G	Karanth *et al.* (2011b)	KUK	C, D, F, A, W
		AMG	D, F, A, W
		NSK	D,F,W
		SV	D,F,W
		JDN	C,D,A,W
		DIM	C,D,A,W
Appendix H	Karanth *et al.* (2011a)	KUK	C,A,W
		AMG	C,A,W
		NSK	C,A,W
		MD	A,W
		JDN	C,W
		JS	C,W
		BRN	C,W
		SLP	C,W
Appendix I	Patil *et al.* (2011)	NP	C,A,W
		NSK	C,A,W
		AMG	C,A,W
		KUK	C,A,W
Appendix J	Karanth *et al.* (2012)	KKK	C,D,F,A,W
		AMG	C,D,A,W
		RD	C,D,A,W
		NB	F
Appendix K	Lachish *et al.* (2012)	SL	C,D,A,L,W
		AMG	D,A,W
		SCLK	D,L,W
		BCS	C,D,W
		KUK	C,A,W

Continued on next page

Appendix L Karanth *et al.* (2013a)

Table 1 – continued from previous page

In thesis	Citation	Authors	Contributions
		AMG	C,A,W
		KKK	C,A,W
		JG	W
		JS	W
		JGR	W
Appendix M	Karanth *et al.* (2013b)	KKK	C, D, F, A, W
		AMG	A, W
		PKP	F, A
		SD	F, A
Appendix N	Karanth *et al.* (2013c)	KKK	C, D, F, A, W
		LNT	C, D, W
		RD	C, D, W
		AMG	A, W
Appendix O	Mondol *et al.* (In press)	SM	C, D, A, W
		NSK	F, W
		AMG	F
		KS	A, W
		KUK	D, F, W
		UR	C, D, A, W

Author abbreviation: Arjun. M. Gopalaswamy (AMG), K. U. Karanth (KUK), N. Samba Kumar (NSK), David. W. Macdonald (DWM), J. Andrew Royle (JAR), James E. Hines (JEH), Pallavi Singh (PS), Devcharan Jathanna (DJ), Mohan Delampady (MD), James. D. Nichols (JDN), Lex Hiby (LH), Phil Lowell (PL), Narendra Patil (NP), Samrat Mondol (SM), Anish Andheria (AA), Uma Ramakrishnan (UR), Priya Singh (PS), Darryl. I. MacKenzie (DIM), Srinivas Vaidyanathan (SV), John Seidensticker (JS), Barry. R. Noon (BRN), Stuart. L. Pimm (SLP), Krithi. K. Karanth (KKK), Ruth DeFries (RD), Natasha Ballal (NB), Shelly Lachish (SL), Sarah. C. L. Knowles (SCLK), Ben. C. Sheldon (BCS), John Goodrich (JG), John Seidensticker (JS), John. G. Robinson (JGR), Parvathi. K. Prasad (PKP), Shreya Dasguptha (SD), Lisa Naughton-Treves (LNT) and Karthik Sunagar (KS).

Contributions code: Concept (C), Design (D), Field sampling (F), Analysis (A), Writing (W), Programming (P), Statistical development (S), Genetic analysis (G) and Laboratory work (L).

Abstract

The tiger *Panthera tigris* is a globally endangered large carnivore of high conservation importance. It is cryptic, wide-ranging and naturally occurs at low population densities. Any study of its metapopulation in real landscapes is faced with a suite of methodological constraints because the temporal and spatial scale of study is large. The scope presented for developing, testing and reviewing methods, whether they are analytical, practical or conceptual, in such a context is wide: ranging from statistical methods to laboratory techniques to field methodologies to software products and methods, all relevant to tiger research and conservation endeavours.

Framed within the ecological theme of assessing population dynamics of species within patchy environments, and with the source-sink view of tiger metapopulations, this thesis investigates the abundance estimation question of tigers and their prey at multiple scales within a 38,000 km^2 tiger landscape in the Western Ghats, India.

While the emphasis is firmly rooted in the development of statistical methods, software products and field methodologies, it also brings in research in laboratory techniques to bear on the abundance and distribution questions on hand. In specific, this thesis:

1. Advances our understanding of spatial capture-recapture methods - a statistical methodology being widely used over the last five years for density estimation.
2. Develops a new field-based method for estimating the density of tiger prey, specifically when they occur at low densities, using an occupancy-based approach. The results of this exercise show much promise and offers a way of tackling the abundance estimation question in unmarked, rare species.
3. Develops a software product called SPACECAP to make novel spatial capture-recapture methods accessible to wildlife biologists, ecologists and park managers. The software is adequately tested and provides users with the necessary summary statistics of density and

related parameters, along with the necessary diagnostic tools in order facilitate accurate interpretation of parameters.

4. Develops a Bayesian inferential approach to estimating tiger density in areas of low abundance by bringing together data sets from multiple sources (camera trapping images and faecal DNA samples, in our case) to strengthen inference about tiger density. Results from this approach suggests that it takes a relatively small increase in sampling effort to bring about large reductions in the variance of density estimates.

5. Investigates a long-standing controversy of index-calibration experiments at large scales. The theoretical models and the empirical testing with tiger sign-encounter data at macroecological scales demonstrate the relative futility in employing simplified and direct linear models using the R^2 statistics, because there are latent sampling process parameters which considerably weaken inference.

6. Develops a spatial capture-recapture model that facilitates investigation of unanswered questions lying at the interface between behavioural and population ecology of carnivores. This is done by introducing the attraction-repulsion spatial arrangements of carnivores into spatial capture-recapture models.

Nearly 70 percent of the world's tigers remain in less than 100,000 km^2 of habitat today. Yet, we lack sufficient methodological tools to reliably estimate their population dynamic parameters. This thesis provides a toolbox of advanced methods to make these assessments more reliable and, more importantly, accessible. It is envisioned that researchers and practical conservation managers will both benefit with this toolbox.

Contents

1 Introduction **1**

2 Spatially-explicit Capture-Recapture Models **7**

2.1 Citation . . . 7
2.2 Abstract . . . 7
2.3 Introduction . . . 8
2.4 Spatially-Explicit Capture-Recapture Models . . . 8
2.4.1 Concepts . . . 8
2.4.1.1 Simulations and Inverse Prediction . . . 9
2.4.1.2 Likelihood-based methods . . . 10
2.4.1.3 Bayesian methods . . . 11
2.4.2 Software . . . 13
2.4.3 Future directions . . . 13

3 Estimating tropical forest ungulate densities from sign surveys using abundance models of occupancy **15**

3.1 Citation . . . 15
3.2 Abstract . . . 15
3.3 Introduction . . . 16
3.3.1 Study species and objectives . . . 17
3.3.2 Study Area . . . 17
3.4 Methods . . . 18
3.4.1 Design of simulations . . . 18
3.4.1.1 Model component for the spatial distribution of ungulates and their signs . . . 18
3.4.1.2 Model component for the field survey sampling process 19
3.4.1.3 Model validation using simulations . . . 20
3.4.2 Field surveys . . . 22
3.4.2.1 Choice of grid cell size . . . 22
3.4.2.2 Line transect surveys of ungulates . . . 23
3.5 Results . . . 23
3.5.1 Simulations of occupancy models . . . 23

3.5.2 Estimates from Field Surveys . . . 24
3.5.2.1 Influence of heterogeneity in detection probabilities on estimates of occupancy . . . 24
3.5.2.2 Density estimates from AOS model . . . 25
3.5.2.3 Density estimates from line transect surveys . . . 25
3.6 Discussion . . . 26
3.7 Figures . . . 28
3.8 Tables . . . 33
3.9 Appendices . . . 39

4 Program SPACECAP: Software for estimating animal density using spatially-explicit capture-recapture models 49
4.1 Citation . . . 49
4.2 Abstract . . . 49
4.3 Introduction . . . 50
4.4 Material and methods . . . 51
4.4.1 Package development . . . 51
4.4.2 Model description . . . 52
4.4.3 Models implemented . . . 53
4.4.4 Parameters reported . . . 53
4.4.5 Assessment of model adequacy . . . 54
4.4.6 Summary statistics and output . . . 55
4.5 Results . . . 55
4.6 Discussion . . . 56
4.7 Tables . . . 58
4.8 Figures . . . 60

5 Density estimation in tiger populations: combining information for strong inference on tiger density 63
5.1 Citation . . . 63
5.2 Abstract . . . 63
5.3 Introduction . . . 64
5.4 Methods . . . 65
5.4.1 Field surveys . . . 65
5.4.1.1 Photographic surveys . . . 65
5.4.1.2 Fecal DNA surveys . . . 66
5.4.2 Analytical methods . . . 66
5.4.2.1 The direct approach . . . 66
5.4.2.2 The step-wise approach . . . 67
5.4.2.3 The spatially explicit capture-recapture model . . . 68

5.4.3 Implementation for tiger survey data in Bandipur 70
5.4.3.1 A priori models 71
5.4.3.2 MCMC implementation and convergence diagnostics 73
5.4.4 Assessment of model adequacy 73
5.5 Results 75
5.6 Discussion 76
5.7 Tables 78
5.8 Figures 82
5.9 Appendices 83

6 An examination of index-calibration experiments: counting tigers at macroecological scales 89
6.1 Citation 89
6.2 Abstract 89
6.3 Introduction 90
6.4 Material and Methods 92
6.4.1 Sampling situation 92
6.4.2 Objective 1: Proposing models to uncover the parameters of the R^2 statistic in index-calibration experiments 93
6.4.2.1 Model 2: When p^* is a variable 96
6.4.3 Objective 2: Application to large scale, macro-level assessments of tiger distribution and abundance in India 97
6.4.3.1 Establishing the relationship between the detection probability parameters of two occupancy models . . 98
6.4.3.2 Standardizing the scale of detection probability estimates across multiple studies to enable a valid comparison 98
6.4.3.3 Predicting R^2 statistics using the detection probability estimates from the tiger occupancy studies . . 99
6.4.3.4 Empirical estimation of R^2 from tiger index-calibration experiments and assessments against predictions of R^2 from detection probability estimates. 100
6.5 Results 101
6.5.1 Objective 1: Uncovering the sampling process parameters of the R^2 statistic in index-calibration experiments 101
6.5.1.1 The R^2 statistic when p^* is a constant 101
6.5.1.2 The R^2 when p^* is a variable 101
6.5.2 Objective 2: Empirical assessments of tiger index-calibration experiments in India 102

6.5.2.1 Relationship between the detection probability parameters of two occupancy models 102
6.5.2.2 Standardizing scale of detection probability estimates across multiple studies to enable a valid analysis . . 102
6.5.2.3 Predicting R^2 statistics for index-calibration experiments, from detection probability estimates from large-scale tiger distribution studies 103
6.5.2.4 Assessing predictions of R^2 statistics from our proposed models against empirical estimates of R^2 from known tiger index-calibration exercises 103
6.6 Discussion 104
6.7 Figures 108
6.8 Tables 113
6.9 Appendices 117
6.9.1 Derivation of R^2 when p^* is constant 117
6.9.2 Derivation for lower and upper bounds of R^2 when p^* is constant 117
6.9.3 R code for predicting R^2 when p^* is a constant 120
6.9.4 Derivation for R^2 when p^* is variable 122
6.9.5 Derivation for lower and upper bounds for R^2 when p^* is variable 124
6.9.6 R code for predicting R^2 when p^* is variable 127
6.9.7 R code for parametric bootstrapping to obtain standard errors for p(1) and calculations of predicted R^2 values 130

7 An attraction-repulsion type Bayesian spatially-explicit capture-recapture model to assess behavioural plasticity in carnivore populations 139
7.1 Introduction 139
7.2 Material and methods 143
7.2.1 Notations and variables used 143
7.2.2 Sampling scheme 143
7.2.3 Model development 143
7.2.4 Implementation of the model 147
7.2.5 Simulations 147
7.3 Results 148
7.4 Discussion 148
7.5 Tables 150
7.6 Figures 151
7.7 Appendices 153
7.7.1 Attraction-repulsion SECR model - R code 154
7.7.2 Attraction-repulsion SECR model - R code for simulation . . . 176

8 Conclusions 181

A A tiger cannot change its stripes: using a three-dimensional model to match images of living tigers and tiger skins 187

B Evaluation of non-invasive genetic sampling methods for estimating tiger population size 193

C A hierarchical model for estimating density in camera-trap studies 205

D Bayesian inference in camera trapping studies for a class of spatial capture-recapture models 217

E Factors influencing densities of striped hyenas (Hyaena hyaena) in arid regions of India 231

F Tigers of trails: occupancy modeling for cluster sampling 241

G Monitoring carnivore populations at the landscape scale: occupancy modelling of tigers from sign surveys 253

H Counting India's tigers reliably 263

I Dispersing tiger makes a point 265

J Assessing patterns of Human-Wildlife Conflicts and Compensation around a Central Indian Protected Area 267

K Site-occupancy modelling as a novel framework for assessing test sensitivity and estimating wildlife disease prevalence from imperfect diagnostic tests 281

L Sinks as Saviours: Why flawed inference cannot assist tiger recovery 293

M Patterns of human-wildlife conflicts and compensation: Insights from Western Ghats protected areas 295

N Living with Wildlife and Mitigating Conflicts around three Indian Protected Areas 307

O Identifying species, sex and individual tigers and leopards in the Malenad-Mysore Tiger Landscape 321

Bibliography 333

List of Figures

3.1 Map of Bhadra Tiger Reserve with a system of line transects (inset map shows the location of the park in India). The study area (Muthodi) is shaded grey 29

3.2 An example of a simulation output in R describing the ecological process of sign deposition (black circles) around animal group activity centres (pink dots), within 100 sampling grid cells. 30

3.3 Ungulate sign survey conducted in 55 grid cells (each of size 3.24 km^2) within Muthodi (Bhadra Tiger Reserve) 31

3.4 Effect of daily movement range R on the relative RMSE value of the estimate of $\hat{\lambda}$ from simulation exercises. Graph demonstrates that low or high values of R produces large relative RMSE values. 32

4.1 The Graphic User Interface(GUI) of SPACECAP 60

4.2 A pixellated density map produced in SPACECAP showing relative animal densities. This map reports estimated tiger densities per pixel of size 0.336 km^2. White regions represent unsuitable habitat. 61

5.1 Camera trap locations used in Bandipur Tiger Reserve during 2006. The area shaded grey is the approximate area of sampling that includes a 2 km buffer around the convex polygon. 82

5.2 Fecal DNA sampling routes used in Bandipur Tiger Reserve in 2006. Blue dots represent mid-points of the 1 km segments along the sampling routes. These blue dots serve as the fecal DNA trapping devices in this study.The area shaded grey represents the area that was sampled using photographic sampling to represent the extent of overlap between the two sampling procedures. 83

6.1 Plot of R^2 upper bound values when, p, V and Q are varied using Model 1: p^* = constant. 108

6.2 Plot of R^2 lower bound values when, p, V and Q are varied using Model 1: p^* = constant. 109

6.3 Plot of R^2 upper bound values when, a and b, V, Q and μ are varied using Model 2: p^* = variable. Note, here μ is represented as three categories - $\mu < 33$ is represented by an empty square, $33 < \mu < 66$ is represented by a crossed-out square and $66 < \mu < 100$ is represented by a dark square 110
6.4 Plot of R^2 lower bound values when, a and b, V, Q and μ are varied using Model 2: p^* = variable. Note, here μ is represented as three categories - $\mu < 33$ is represented by an empty square, $33 < \mu < 66$ is represented by a crossed-out square and $66 < \mu < 100$ is represented by a dark square. 111
6.5 Plot of the relationship between tiger sign encounter rates and estimated tiger densities from 8 different sites in India in IC-Karanth. The line represents the estimated regression line. 112

7.1 A simulation of the state space with a repulsive field $\omega = 0$. Here, $N = 200$, $\sigma = 0.7$ and $\kappa = 0.4$ 151
7.2 A simulation of the state space with an attractive field $\omega = 1$. Here, $N = 200$, $\sigma = 2.3$ and $\kappa = 0.4$ 152

Chapter 1

Introduction

Early theoretical work in ecology demonstrates the importance of patchy environments in determining the extent of stability in model ecosystems, both at the level of individuals (MacArthur & Pianka, 1966) and at a broader population level (Levin, 1976). The basic idea is well illustrated in a host-parasitoid system, an important sub-class of predator-prey systems, by May (1978), where the notion of patchiness is summarised into a single overdispersion parameter k with a negative binomial model. Most importantly, May (1978) shows that among the three basic parameters described in the system, viz., the parasitoid 'area of discovery' a, and the host rate of increase F, and the clumping parameter k, it is the parameter k that dominantly determines the stability of such a host-parasitoid system, fairly independent of the other two parameters a and F. The generality of this mathematical development in describing patchiness via the negative binomial model is widely applicable to most predator-prey systems. Interestingly, it is shown theoretically that such clumping of resources also provides for increased opportunities for species co-existence among prey communities (Holt, 1993).

The population dynamics involved in such patchy environments (Levin, 1976) have been theoretically explored under two types of arrangements: (1) the metapopulation view (Hanski & Hanski, 1999) and the (2) source-sink view (Pulliam, 1988). A 'metapopulation' structure in its most general sense can be defined as a system of local populations linked by dispersal. And the empirical systems which motivated much of the development of theoretical metapopulation dynamics models involved highly ephemeral systems, such as landscapes of shifting mosaics of patches at different stages of succession. Consequently, environment or demographic stochasticity plays a vital role in many metapopulation models (Hanski & Hanski, 1999). In contrast, if in a habitat or patch, the death rate always exceeds the birth rate then the population deterministically goes extinct if immigration is prevented. Hence, a source-sink structure can simply be viewed as a less ephemeral system of local populations, where the regional population persists as long as productive source populations (where birth rates exceed death rates) can successfully offset failing

sink populations (where death rates exceed birth rates) (Holt, 1993). It logically then follows that if natural selection favours organisms which select their habitats to maximize their relative fitness (Fretwell, 1972) then the source-sink structure must be the more pervasive structure describing natural regional population structures.

Perhaps, both these theoretical views about how regional populations structure themselves and evolve are equally relevant and it is only an issue of the ecological scale (Wiens, 1989) or system in consideration that would define one view strongly over the other. For example, the source-sink structure may be one 'act' of the ecological theatre (Hutchinson, 1965) within a larger metapopulation dynamics play at the evolutionary level. For example, in volatile environments typified by host-parasitoid systems, associations of hosts to habitat types may be less tight, leading to spatial self-organization (Hassell *et al.*, 1984) and providing a greater justification for metapopulation type of regional population models in order to assess population persistence (Hanski & Hanski, 1999). These models may further gain importance for terrestrial wildlife populations where we have empirical evidences demonstrating, on occasions, decoupling in the species density-habitat (or patch) quality relationships (Van Horne, 1983). Such decoupling can also occur as a result of over-hunting of wildlife populations by humans, leading to patches of 'empty forests' (Redford, 1992), and making it necessary to factor in the effect of hunting in assessing, what are otherwise, reasonable representations of source-sink systems (Novaro *et al.*, 2005).

Large carnivore populations tend to structure themselves as typical source-sink systems (Woodroffe & Ginsberg, 1998) owing to their large dependence on prey availability (Macdonald, 1983; Carbone & Gittleman, 2002; Karanth *et al.*, 2004). The understanding of and assessing the persistence of such predator-prey systems based on source-sink ideas (Pulliam, 1988; Watkinson & Sutherland, 1995) requires the estimation of state variables (principally density or abundance at each patch, in order to provide an assessment of the overdispersion parameter k), as well as vital rates, such as survival, recruitment, mortality, immigration and emigration rates (Williams *et al.*, 2002). And regardless of whether the broader view takes the source-sink form or the metapopulation form, estimation of dispersal parameters becomes essential (Holt, 1993).

However, these estimation goals are very difficult to meet in empirical studies owing to the classical sampling problems of spatial sampling (Thompson, 2002) and the estimation of detection probability (Williams *et al.*, 2002). These basic problems are compounded by field logistics when the target species is elusive, wide-ranging and when they naturally tend to occur in low densities, such as most large carnivores (Long & MacKay, 2008). Hence, the methodological issues in reliable estimation of such parameters become broad in scope but urgent in need in the context of increased habitat destruction leading to high extinction debts (Tilman *et al.*, 1994; Kuussaari

et al., 2009). In this thesis, I look into some methodological issues that come to light in studying the metapopulation dynamics of tigers *Panthera tigris*, a globally endangered species of high conservation importance.

A broad topic such as "Addressing Methodological Issues in the Study of Tiger Metapopulation Dynamics in the Western Ghats, India" as a doctoral project, comes with an obvious set of concerns. In terms of spatial scale, it is massive - we are referring to a study area size of about 38,000 km^2 within the state of Karnataka, India. Any field survey calls for large investments of money and manpower. The temporal scales of such a study will also extend well beyond the time line of a typical doctoral study considering that the species of interest, the tiger, is a secretive, wide-ranging carnivore occurring at low population densities. Methodological research questions can also span across sub-disciplines, from ecology to applied ecology and conservation, because across the landscape, tigers do come in contact with humans, resulting in tiger mortality and varying situations of conflict. Hence, addressing methodological challenges across such a spectrum of problems will require collaboration with scientists of different specializations across this landscape.

In this context, this thesis is organized broadly into two parts:

1. Primary contributions - These are presented as the main chapters in the thesis and represent my work as the lead or sole author.

2. Secondary contributions - These are presented in the appendices in the thesis and represent my work as the secondary author.

I have been involved in the development of methodological ideas relevant to the estimation of tiger density a few months prior to my doctoral work. These early forays led to the development of some of the first few papers on Bayesian spatially-explicit capture recapture models (see Royle *et al.*, 2009b,a, Appendices C and D, respectively). The following years led to an explosive development of these ideas in the field of statistical ecology. Hence, in *Chapter 2* of the thesis, I review the various milestones during the development of spatial capture-recapture models to lay the foundation for the other chapters in this thesis.

I take up specific questions related to the broader thesis topic from *Chapter 3*. The high predictive power of tiger prey species on the abundance of tigers has been well established (Karanth *et al.*, 2004). When prey densities are high, well established methods, such as line-transect sampling (Karanth *et al.*, 2004; Kumar, 2011). However, estimating tiger prey densities when they are at low levels has proven to be a major challenge. And prey populations are at expected to be at very levels at over 94% of the tiger's current range (Walston *et al.*, 2010). Therefore, in this chapter, I propose an alternative technique of estimating tiger prey density

by applying some of the more advanced occupancy approaches (Royle & Nichols, 2003).

An important aspect in the field of methodological developments is that of taking cutting-edge models to field ecologists and conservationists. The gap is often so large that such models may take years to reach applied ecologists. To bridge this gap, one very useful method has been to develop user friendly software packages that the take latest methodological developments to the application. This was also a topic of much discussion in the International Statistical Ecology Conference 2014 in Montpeliier, France. Motivated by my peripheral involvement in the development of a software package called ExtractCompare (Hiby *et al.*, 2009, Appendix A), and also appreciating its practical application, in *Chapter 4* I develop and describe a free software program called SPACECAP, which implements the Bayesian SECR analysis of Royle *et al.* (2009a).

Continuing with the theme of spatial capture-recapture model development, in *Chapter 5*, I extend this idea to address the problem of estimating tiger density when they are at low densities. In most landscapes, large carnivores occur at densities that can be too low to permit collection of sufficient data for accurate density estimation; therefore, there is a premium on developing approaches that make full use of data from multiple sources and using an inferential framework that permits the use of ecological constraints and prior knowledge in density estimation. In *Chapter 5* I develop such an approach to estimate abundance and density of a tiger population in the Western Ghats, India, utilizing a recently-developed spatially explicit capture-recapture model (Royle *et al.*, 2009a, 2011a) to combine information about from two pre-designed but distinct sources of tiger capture-recapture survey data: (1) photographic captures of tigers from camera-traps and (2) faecal samples of tigers containing DNA collected during the same season at a single site. The motivation for this study comes from three earlier studies I participated in: two studies involving the application of a conventional capture-recapture analysis of tiger individuals identified by faecal DNA sampling and photographic capture-recapture (Mondol *et al.*, 2009, In press, Appendix B and Appendix O) and an application of capture-recapture study of striped hyenas, (*Hyaena hyaena* Singh *et al.*, 2010, Appendix E). It appealed to me from these studies that when camera trap surveys are conducted, it is also possible, in parallel, to sample for faecal samples of the target species without the addition of extra sampling effort. This situation prompted me, in this chapter, to ask how effective would it be to combine information from more than a single source of data in such studies.

In *Chapter 6*, I address methodological issues of tiger metapopulation dynamics at large spatial and temporal scales, still focussed though on the basic abundance estimation question. Motivated by the large scale tiger monitoring programs in

India, this chapter focusses on a particular aspect of index-calibration experiments, a technique used in the assessment of large carnivores at large spatial scales. I aim to uncover some of the theoretical underpinnings of a very basic and long-standing controversy in ecology, about the utility of abundance indices as surrogates for true estimates of animal abundance at macroecological scales. I then confront these theoretical findings with empirical analysis from some recent landscape scale surveys of tiger in India (see Karanth & Nichols, 2000; Karanth *et al.*, 2004; Jhala *et al.*, 2011b,a; Karanth *et al.*, 2011b,a, Appendix G, Appendix H). Since, in all these studies huge resources were invested across large spatial scales, it was important to assess the inference from such experiments and, thus, providing the motivation for this chapter.

Finally, in *Chapter 7*, I bring back focus of methodological developments from macroecological scales back to individuals and populations. As a keen participant in the new developments in spatial capture-recapture models from some of the previous chapters, it struck me that the explicit recognition of animal locations during the estimation of animal density in these models provides a great opportunity for asking questions about carnivore spacing patterns and mechanisms (Sandell, 1989). In this chapter, I develop an attraction-repulsion type of a spatially explicit capture-recapture model that estimates the extent of attraction or repulsion between individuals. As a carnivore biologist and, more broadly, as an ecologist, I was motivated by the gap between that exists in our knowledge between individuals and populations in understanding ecology (May, 1999). This development hopes to direct research attention to address this gap.

Chapter 2

Spatially-explicit Capture-Recapture Models

2.1 Citation

Gopalaswamy, A. M. (2013). Capture-Recapture Models, Spatially Explicit. *Wiley StatsRef: Statistics Reference Online*.

2.2 Abstract

Animal density is a parameter of basic interest to field biologists. But owing to logistical and analytical constraints, estimating animal density has been very challenging. Spatially-explicit capture-recapture models (SECR) were developed to address this particular, long-standing problem. In this article, some of the important 'milestones' in the SECR literature are discussed. The basic idea in SECR models is to include information about locations of animal captures into the modeling so that density is estimated directly. The early SECR type of approach involved estimation of density via simulations and inverse prediction. The first direct SECR estimator was likelihood-based, which was more flexible and allowed for inclusion of covariates. In parallel began the development of Bayesian versions of the SECR models. These models allowed for greater flexibility and easily allowed for an extension, which could, in addition, also estimate population vital rates. SECR models have already witnessed several applications: ranging from analysis of camera trap data on large cats to acoustic signal data from birds. The future of SECR models is expected to go far beyond density estimation. It has great potential to answer some long-standing questions about how behavioural processes drive population dynamics in wild animals.

2.3 Introduction

Animal density is a parameter of fundamental interest to wildlife ecologists, conservation biologists and park managers. But this important parameter has been very difficult to estimate in the field, especially for rare and elusive species (Thompson, 2004), owing to several logistical and analytical constraints (Williams *et al.*, 2002). Conventional capture-recapture sampling (Williams *et al.*, 2002) has been very popular among field biologists for estimating animal abundance, and consequently density, for animals that can be easily 'trapped' or are naturally marked with a non-invasive survey procedure (Long & MacKay, 2008). A wide variety of capture-recapture tools are now available (Amstrup *et al.*, 2005), accompanied by user-friendly software interfaces (White *et al.*, 1978; White & Burnham, 1999). Hence, the use of capture-recapture approaches has been widespread in animal abundance estimation literature. While developments in capture-recapture theory has been prolific for abundance estimation, the procedure of converting animal abundance to animal density has, essentially, remained ad hoc, although a wide range of estimators exist (see Efford, 2004, for a discussion). Conversion of abundance to density, traditionally, involved estimating the effectively sampled area by applying a notional 'buffer' around the trapping array; the buffer width is set to the full (Dice, 1941) or 1/2 (Wilson & Anderson, 1985) the mean of the maximum distance (MMDM) moved by individual animals in the trapping array. An empirical study (Soisalo & Cavalcanti, 2006) demonstrated that both these approaches provided biased results for density.

Thus, the primary objective during the development of spatially-explicit capture-recapture (SECR) models (Efford, 2004; Borchers & Efford, 2008; Royle *et al.*, 2009b,a) was to directly address the problem of density estimation. This objective is adequately achieved by SECR models because information about the location of animal captures during sampling is incorporated within a single, unifying model that estimates density directly. This entry briefly, in a non-technical manner, summarises some of the important SECR estimators developed. It also discusses some extensions of SECR models that have been applied in practice. Finally, some of the potential directions for future research are outlined.

2.4 Spatially-Explicit Capture-Recapture Models

2.4.1 Concepts

During early capture-recapture experiments with small mammals it was quickly understood that different animals in the population appeared to be differentially caught by trapping devices and simple capture-recapture models (e.g. the null

model) produced biased estimates of population size in the face of such heterogeneity in capture-probabilities. The preferred estimator for animal abundance in the face of such heterogeneity in capture-probabilities was the Mh-jackknife estimator (Otis *et al.*, 1978), and more recently, the mixture models estimator (Pledger, 2000). However, a large part of the heterogeneity in capture probabilities among individuals is related to the relative position of an animal with respect to the trapping array. So animals located at the edge of the trapping area had a lower probability of being captured by the trapping device in comparison with an animal located at the centre. Early attempts to estimate density involved the use of trapping web designs. And the area of 'animal exposure' was modelled as a as a detection function, as used in distance sampling, but doubts remained on the reliability of the estimates (Lukacs, 2002).

2.4.1.1 Simulations and Inverse Prediction

The first idea to address this problem of density estimation via some sort of spatially-explicit process was proposed by Efford (2004). The problem was viewed as a problem of estimating the parameter vector $\theta = (D, g_0, \sigma)$ by sampling animals using a trapping array consisting of r traps. Here, D is the intensity parameter of a spatial point process (representing the animal distribution), g_0 is the overall magnitude of captures and σ is the spatial scale over which the capture probability declines. But at that point, no method for directly estimate the parameter vector θ was available. However, there were plenty of readily available capture-recapture estimators (Otis *et al.*, 1978) that could estimate the population size $\hat{N}$ and average capture-probability $\hat{\bar{p}}$ well. And the estimate of the average distance moved by each animal $\hat{\bar{d}}$ was separately available from the capture-recapture data.

Efford (2004) applied the logic that this parameter estimate, say $\hat{\phi} = \left(\hat{N}, \hat{\bar{p}}, \hat{\bar{d}}\right)$, contains the information necessary to obtain $\hat{\theta}$. Further, Efford (2004) constructs plots (see Efford, 2004, for these plots), by *simulations*, to show the direct, monotonic relationships between the following pairs of parameters: (N, D), (p, g_0) and (d, σ), when the other parameters are kept constant. Hence, if the response y_i (an estimate of parameters N, p and d) is obtained from a particular data set i, then,

$$y_i = F_T(\theta) + \epsilon_i$$

where, T is the trapping array used and ϵ_i is the multivariate normally distributed error term. The estimate $\hat{\theta}$ is then obtained, by *inverse prediction* as follows,

$$\hat{\theta} = F_T^{-1}(y_i)$$

The details of the estimation of $\hat{\theta}$ and its variance-covariance matrix, by numerical procedures, is described in Efford (2004). The model itself can be applied using the

software program DENSITY (Efford *et al.*, 2004). While this approach is generally encouraging, this method of estimation is restrictive and it is inherently difficult to include covariates of interest into the modelling.

2.4.1.2 Likelihood-based methods

One of the first direct SECR estimators was the likelihood-based modelling proposed by Borchers & Efford (2008). This model offers more flexibility so that covariates can easily be incorporated during estimation. The approach is flexible enough to include some variants from conventional capture-recapture theory, such as modelling behavioural response and the effect of time on capture-probability. Further, Borchers & Efford (2008) also account for unmodelled heterogeneity in individual capture probabilities in their approach.

The likelihood model is of the form,

$$L\left(\phi,\theta|\, n, w_1, w_2, w_3, \ldots, w_n\right) = \Pr\left(n|\phi,\theta\right) \Pr\left(w_1, w_2, w_3, \ldots, w_n|n,\phi,\theta\right)$$

Here, θ is the parameter vector related to the capture function. For example, θ could represent the parameters of a half-normal detection function, as used very often in distance sampling (Buckland *et al.*, 2001); ϕ is the parameter vector related to the spatial point process, or the vector of parameters describing animal distribution and abundance in the state space. For example, in their model development Borchers & Efford (2008) use the inhomogeneous Poisson point process on the real plane to describe animal distribution. While, in theory, this point process could take any other form (see Møller & Waagepetersen, 2004, for many examples), constructing the likelihood may be difficult for other forms. Here, n is the total number of animals caught in the study so that w_i, is the vector of trap locations where animal i appeared during sampling, is observed for $i = 1, \ldots, n$.

The model parameters θ and ϕ can be estimated by maximising the above likelihood equation with respect to them. For a fixed area, suppose $D\left(X;\phi\right)$ represents the animal density, where X are the home-range centres within the area. Then, evaluating $D\left(X;\phi\right)$ at the MLE $\hat{\phi}$ provides an estimate of the density surface. Hence, over a specific area, $D\left(X;\hat{\phi}\right)$ is the mean animal density. Consequently, the total number of animals is given by $\hat{N} = \int D\left(X;\hat{\phi}\right) dX$ over the region of interest.

Borchers & Efford (2008) suggest extensions of this model to incorporate M_t (time effect) and M_b (behavioural response) versions (Otis *et al.*, 1978) of conventional capture-recapture models into this spatially-explicit version. Further, unmodelled heterogeneity in individual capture probabilities for reasons other than trap access is also quite easily incorporated in this SECR model using the model M_h (Pledger, 2000). As well, being likelihood-based, competing models can be assessed

using a tool, such as AIC (Burnham & Anderson, 2002), for estimation of density and model selection.

The Borchers & Efford (2008) model can be viewed as a *quasi*-SECR model in the context that this model only seeks to obtain information from the 'caught' animals n in the study for inference about animal density and abundance. But inference about 'uncaught' animals from this model is limited. While this likelihood construction may not have a direct impact on the density estimate per se, it may be quite restrictive in terms of developing extensions for say, open populations.

2.4.1.3 Bayesian methods

The success of conventional capture-recapture models, in part, can be attributed to the fact that the CR framework has been very flexible to test ecological phenomena, such as transience (Pradel *et al.*, 1997) or temporary emigration (Kendall *et al.*, 1997) using simple capture-recapture data matrices (Otis *et al.*, 1978). Additionally, the idea that these models kept track of 'marked' individuals and paid equal attention to 'uncaught' animals made this approach very attractive. For example, the development of open models via the 'robust design' (Pollock, 1982; Kendall *et al.*, 1995) from closed model data sets exploits the idea of tracking individual animals over time in the face of sampling uncertainty, such as, for example, some 'uncaught' animals in year 1 are caught in year 2 of which some may not get caught in year 3 etc.

Bayesian versions of SECR models (Royle & Young, 2008; Royle *et al.*, 2009b,a; Gardner *et al.*, 2009) were developed, motivated by such capture-recapture thinking, keeping also in mind that animals that are located within the trapping array may remain 'uncaught' during sampling. The model developed by Royle *et al.* (2009a) is described in more detail here because it is a refined version of the two earlier developments (Royle & Young, 2008; Royle *et al.*, 2009b). Royle *et al.* (2009a) develop a class of Bayesian SECR models and show how these models can be represented as forms of generalized linear models (GLM) with random effects. However, the analysis itself can become complicated because there are several unknown quantities: how many animals there are, N, and where these N animals are located, S. Suppose, $s_i = (s_{1i}, s_{2i})$ represents a point in a 2-dimensional coordinate system about which movements of an individual i are concentrated. Suppose there are N animals exposed to the trapping area, so the location of each individual activity centre is given by $s_i, i = 1, 2, 3, \ldots, N$. These locations describe the *state process* of animal distribution over space. In the Bayesian implementation of Royle *et al.* (2009a), this state process is modelled as a binomial point process so that,

$$s_i \sim \text{Uniform}\,(S)$$

This uniform point process serves as the *prior* distribution for the location of individual activity centres. The benefits of using the Bayesian implementation is that the posterior distribution of this point process permits that animals are concentrated in patches which reflect the true underlying distribution of animals.

Next, consider J trapping devices with locations $\{x_j, j = 1, 2, 3, \ldots, J\}$. Define $d_{ij} = \|s_i - x_j\|$, where $\|.\|$ is the normal Euclidean distance. Suppose these trapping devices are exposed to sampling for k occasions. Then, the encounters of an individual animal i at a trap j on a particular occasion k is modelled as,

$$y_{ijk} \sim \text{Poisson}\left(\lambda_0 g_{ij}\right),$$

where, λ_0 is defined as the basal encounter rate that represents the number of encounters had the trapping device been located exactly at the home range centre of an animal i. g_{ij} represents the probability of detection of an individual i at trapping device j, and g_{ij} is a function the distance between an individual i from trapping device j. This function could be one of the many detection functions used in distance sampling (Buckland *et al.*, 2001). So, for example, $g_{ij} = e^{\left(-d_{ij}^2/\sigma\right)}$ is the half-normal detection function that can be used to model the rate of decline in encounter probability as the animal is located away from the trap.

The above model is implemented using Bayesian Markov Chain Monte Carlo (MCMC) simulations. The idea is to employ a technique called 'data augmentation' (Royle *et al.*, 2007), whereby, a large number of zero-encounter histories, with their locations, are added to the known number of animals caught, n, in the survey. Suppose M is defined as the size of the augmented population, and suppose ψ represents the proportion of M consisting of real individuals, then the population size is modelled as,

$$N|\psi \sim \text{Binomial}\left(M, \psi\right)$$

So that the true population size, N, lies in between a lower limit of n and an upper limit of M.

Gardner *et al.* (2010a) extend this Bayesian version to develop a model to estimate survival and recruitment in a demographically open population. In their formulation of the observation process, the encounters of an individual animal i at a trap j on a particular 'secondary' occasion k in a particular 'state' (alive or dead) at time t is modelled as,

$$y_{ijkt}|z(i,t) \sim \text{Poisson}(\lambda_0 g_{ij} z(i,t))$$

where, $t = 1, 2, 3, \ldots, T$ are the 'primary sampling periods', as defined in Pollock (1982), which represent sampling periods across which we expect demographic changes. Here, if individual i is alive at a time t then $z(i,t) = 1$, and consequently, the observations are Poisson. And if the individual i is not alive then $z(i,t) = 0$,

then the observations must be zero with probability 1. So the model becomes a version of the zero-inflated Poisson regression model and $z(i,t)$ is then probabilistically modelled to include the survival and recruitment parameters (see Gardner *et al.*, 2010a, for more details).

2.4.2 Software

New methods and models in ecology can get so complicated that practising field biologists find them very difficult to implement with their data. But software packages such as CAPTURE (White *et al.*, 1978) and MARK (White & Burnham, 1999) for analysing capture-recapture data, DISTANCE (Thomas *et al.*, 2010) for analysing line and point transect data and PRESENCE (Hines, 2006) for analysing occupancy data, have helped bridge this gap between model developers and practitioners, because of their user-friendly interfaces. Fortunately, there are currently two user-friendly software packages for analysing spatial capture-recapture data; DENSITY (Efford *et al.*, 2004) and SPACECAP(Gopalaswamy *et al.*, 2012d). Program DENSITY implements the likelihood-based models discussed above along with several variants. And program SPACECAP, a software-package built within the R environment (R Development Core Team, 2009), implements some of the Bayesian versions of the SECR models. Most new models developed are published along with pieces of codes, written either in R, or WinBUGS (Lunn *et al.*, 2000) which can directly be utilised by more advanced users.

2.4.3 Future directions

This article describes a few of the basic spatially-explicit capture-recapture models developed recently. Without doubt, we will see many extensions to these basic models in the future. We have already witnessed the application of SECR models on data from camera trap surveys of large cats(Royle *et al.*, 2009b,a), mist-netting surveys of birds(Borchers & Efford, 2008), acoustic signal surveys of birds(Dawson & Efford, 2009), hair-snare surveys of large carnivores(Gardner *et al.*, 2010b) and for all kinds of search-encounter surveys (Royle *et al.*, 2011b). We have also seen useful extensions to improve estimation of density from SECR models by incorporating covariates (Sollmann *et al.*, 2011) to some parameters and by combining information from multiple sources of data (Gopalaswamy *et al.*, 2012c).

But the utility of SECR models can go far beyond the objective of simply estimating animal density. The SECR approach now provides a framework to investigate how behavioural mechanisms drive population dynamics in wild animals. This has been a big knowledge gap in ecology (May, 1999) but has so far been very challenging to confront. SECR approaches have opened doors by providing unifying models that estimate parameters about behavioural processes (such as home range sizes) with

population processes (such as survival and recruitment), amidst the field problem of sampling processes (Thompson, 2002; Williams *et al.*, 2002). For example, telemetry studies have been widely used in ecology to study individuals and obtain answers to questions about home range sizes of breeding females and how these sizes change over time (see Goodrich *et al.*, 2010, for an example with tigers). These behavioural mechanisms, almost certainly, influence vital rates, and consequently the population size. But, population level parameters are usually estimated by completely different approaches, such as camera-trapping (see Karanth & Nichols, 1998; Karanth *et al.*, 2006; O'Connell *et al.*, 2010b, for an example with tigers). Now, with the advent of SECR models, we have a convincing way of integrating knowledge from individual-based studies and population-based studies with a combined model.

Consequently, I see enormous potential for the use of sophisticated spatial point processes (Møller & Waagepetersen, 2004) within SECR models. I also expect that more complex home range models (Horne & Garton, 2006) will be integrated into SECR models, as fine scale data from radio- and satellite-telemetry are increasingly made available. I also see the extension of SECR ideas to attack inference on 'unmarked' populations. There will, nevertheless, be a constant need for professional software programmers to step in and develop user-friendly packages for increasing the uptake of such complex models by field biologists. I see great potential for research about designing surveys based on SECR models. We still know very little about SECR model robustness and optimal survey designs for estimating parameters of SECR models. Eventually-perhaps in the next 10 years-I expect conventional capture-recapture models to only be a tool of the past.

Chapter 3

Estimating tropical forest ungulate densities from sign surveys using abundance models of occupancy

3.1 Citation

Gopalaswamy, A. M., Karanth, K. U., Kumar, N. S., Macdonald, D. W. (2012). Estimating tropical forest ungulate densities from sign surveys using abundance models of occupancy. *Animal Conservation* 15(6), 669-679.

3.2 Abstract

Large ungulates critically influence forest structure and functioning besides being seriously threatened by anthropogenic pressures. For assessing their populations, surveys of ungulate sign encounters are widely used because of their practicality. However, these yield unreliable results because of their failure to address the problem of imperfect detection. Here, we present an innovative application to address this key weakness in traditional ungulate sign surveys. We describe the ecological process of ungulate sign deposition as well as the observation process of sign detection in our modeling. We simulate 183 ecological and sampling-related parameter values to first evaluate model performance. Simulation results demonstrate that we can achieve good estimates of animal density when the radius of the animal daily movement range is accounted for during survey design. We design and conduct a field survey of ungulate signs to estimate ungulate densities using both occupancy and distance sampling approaches. For five species of ungulates the densities estimated from our sign survey (number of ungulate clusters/km^2) were 1.46(0.68) chital, 1.42(0.67) sambar, 1.01(0.44) gaur, 0.74(0.39) wild pig and 1.42(1.59) muntjac, were similar to those generated from line transect sampling 2.16(0.76) chital, 2.47(0.56) sambar, 0.94(0.3) gaur, 1.09(0.37) wild pig and 4.03(0.83) muntjac), except for muntjac. The potential utility of this approach extends beyond sign surveys of forest ungulates to

a wider range of animal monitoring contexts, including those based on scent-station surveys and camera trap surveys of elusive mammals.

3.3 Introduction

Most large mammal species are threatened globally by over-hunting and habitat destruction consequent upon rapid human population growth and economic development (Ceballos *et al.*, 2005; Schipper *et al.*, 2008). Large ungulates (>15kg body mass) are particularly vulnerable because of biological traits such as their wide-ranging movements, tendency to forage on croplands and body size that attracts hunters (Macdonald, 2001; Madhusudan & Mishra, 2003). However, densities of most tropical forest ungulate species are now significantly depressed and several species are threatened with extinction because of human impacts, potentially driving major changes in forest ecology. Several species impact agriculture or forestry and many are harvested for food by local populations. Given their critical ecological role, reliable assessment of their populations assumes importance. Effective conservation, or even sustained harvest of ungulate populations requires reliable estimates of their population size (abundance) and densities (Katzner *et al.*, 2011).

In the face of imperfect detection, capture-recapture modelling (Williams *et al.*, 2002) has been successfully used for estimating abundance of ungulate species that are individually identifiable from photographs (Goswami *et al.*, 2007) or from fecal DNA (Eggert *et al.*, 2003). Photo-captures require specialized camera trap surveys (O'Connell *et al.*, 2010a), whereas DNA captures demand sophisticated and costly laboratory support, thus limiting their application in the field. Although estimation methods based on photographic encounter rates have been proposed for individually unidentifiable species (Rowcliffe *et al.*, 2008), these require accurate data on animal movement rates that are often difficult to acquire in the field (Rovero & Marshall, 2009).

Line transect surveys (Buckland *et al.*, 2001) involve visual counts of either ungulates (Karanth & Sunquist, 1992; Jathanna *et al.*, 2003) or their dung (Plumptre & Harris, 1995; Marques *et al.*, 2001) and are widely employed to estimate abundance. However, visual surveys are sometimes rendered impractical by low densities or extreme wariness of animals, dense vegetation or difficult terrain. Often converting dung counts from line or point transect surveys (Buckland *et al.*, 2001) into estimates of ungulate density is also problematic because of difficulties in estimating species-specific rates for daily defecation and dung decay across seasons and habitats (Nchanji & Plumptre, 2001; Laing *et al.*, 2003). Furthermore, the amount of dung or tracks deposited by an individual ungulate is usually so large that establishing a

reliable statistical relationship between such signs and actual animal densities is a futile exercise.

Consequently, there exists a need for reliable, accurate and a cost-effective method for monitoring forest ungulate populations in situations where transect surveys are not feasible - particularly when abundance is low. This is a more general abundance estimation problem in ecology and conservation that has received methodological focus in recent times (see Katzner *et al.*, 2011; Isaac *et al.*, 2011; Rowcliffe *et al.*, 2011; Archaux *et al.*, 2012). For our specific problem with forest ungulates, and more broadly across a range of taxa, we believe that the conceptual framework offered by MacKenzie *et al.* (2006) and Royle & Dorazio (2008) is appropriate to address our specific need. We present here the outcome of a rigorous field study that develops and validates the use of such an approach that shows promise.

3.3.1 Study species and objectives

We develop a practical field sampling and density estimation protocol based on field counts of easily detectable signs (dung or tracks) for five species of Asian forest ungulates: gaur *Bos gaurus*, sambar *Rusa unicolor*, chital *Axis axis*, wild pig *Sus scrofa* and muntjac *Muntiacus muntjak*. These species exhibit a diverse range of body sizes (20–1000 kg), social behaviors (solitary territorial to solitary wide-ranging to social herding), diets (selective feeders to browsers to mixed feeders to grazers), habitat preferences (dense forests to open grasslands; Macdonald, 2001) and variations in population densities (0.89–50 animals/km^2).

Our methodology rests on the application of an abundance model of occupancy (see Royle & Nichols, 2003) on designed sign survey data. Although this model is designed for temporally replicated samples, we explore its relevance for spatially-replicated samples (hereafter called AOS model-for Abundance-Occupancy-Spatial) through simulations. We validate properties of our estimates with density estimates derived from extensive computer simulations, and compare it to density estimates derived from line transect surveys using visual counts.

3.3.2 Study Area

We conducted this study in Bhadra Tiger Reserve, Karnataka, India (Figure 3.1; Longitude 75 °29′ E to 75 °47′ E and Latitude 13 °22′ to 13 °47′ N) from April to June 2007. The line transect survey was carried out between April and May and the occupancy survey was conducted during the month of June when pre-monsoon showers begin. This 492 km^2 moist deciduous forest tract (Jathanna *et al.*, 2003) supports populations of the five forest ungulate species of interest. It also supports several large mammal species such as Asian elephant *Elephas maximus*, sloth bear *Melursus ursinus*, dhole or Asiatic wild dog *Cuon alpinus*, leopard *Panthera pardus*

and tiger *Panthera tigris*, and is considered to be an important conservation site within a global biodiversity hotspot (Das *et al.*, 2006). Most of these ungulate species currently occur well below their optimal densities here (Jathanna *et al.*, 2003; Karanth *et al.*, 2004), because the reserve is still in the process of recovering from past anthropogenic impacts (Karanth, 1982; Jathanna *et al.*, 2003; Karanth, 2007). Within the Bhadra reserve, we chose a specific study area of 178 km^2 size in Muthodi (Figure 3.1).

3.4 Methods

3.4.1 Design of simulations

3.4.1.1 Model component for the spatial distribution of ungulates and their signs

Ungulates are assumed to be spatially distributed in the study area following a random poisson process (Pielou, 1969, pages 81–82). For the simulation exercises we utilized a grid of $20 \times 20 = 400$ cells to overcome potential biases of small sample size on parameter estimates. Point locations of ungulate groups were generated using the software package R (R Development Core Team, 2009), with their x and y coordinates in a Cartesian plane. Locations were drawn from a uniform distribution so that a binomial process can describe the number of individual groups in each cell and be approximated as a poisson variate. Let $S = (S_x, S_y)$ represent points on the Cartesian plane and defined as a collection of activity range centres of all individual ungulate groups present within the grid (Figure 3.2).

For a square grid composed of $Q(= 400)$ cells, S_x and S_y are modeled as below:

$$S_x \sim Uniform(0, \sqrt{Q})$$
$$S_y \sim Uniform(0, \sqrt{Q})$$

We assume that signs (tracks or dung) generated by individual ungulate groups are deposited at a uniform rate over a pre-defined period within a circle of radius R (Figure 3.2).

Using the simulation approach described by Hlynka & Loach (2005), we define two additional random variables:

$$\alpha \sim Uniform(0, R)$$
$$\theta \sim Uniform(0, 2\pi)$$

For an individual ungulate group u, located at x_u, y_u represent one possible location of a sign.

Then the location (x_u, y_u) is generated by:

$$x_u = S_{x(u)} + \sqrt{\alpha} cos(\theta)$$
$$y_u = S_{y(u)} + \sqrt{\alpha} sin(\theta)$$

The purpose of employing the Hlynka & Loach (2005) approach is to ensure that signs are 'uniformly' distributed within a circle of a fixed radius without clustering either at the centre or at the rim. We acknowledge that other distributions may also be relevant in the simulation, such as the bi-variate normal distribution commonly employed in spatially-explicit capture-recapture modeling nowadays (Borchers & Efford, 2008; Royle *et al.*, 2009a).

3.4.1.2 Model component for the field survey sampling process

The survey was designed so that skilled field personnel record species presence by detecting signs (fresh tracks and dung), systematically along S-shaped pre-determined survey routes within each grid cell (Figure 3.3). For performing computer simulations we treated sign detections as occurring over narrow detection strips on such trails. Each trail consists of 8 segments, which form 'spatial replicates'Hines *et al.* (2010) within a cell (Figure 3.3). Observers record detections or non-detections of signs of each ungulate species only once on each replicate.

We simulated two broad scenarios of detection conditions. The first assumed that if signs are present in a spatial replicate, observers detect species presence with certainty. In the second, a more realistic scenario, we permitted the probability that observers could miss some signs present on a replicate.

We modeled the probability of detecting ungulate group presence using the Royle & Nichols (2003) formulation

$$p_{sign(ij)} = 1 - (1 - r_{sign})^{N_{sign(ij)}}$$

where, $p_{sign(ij)}$ = probability of detecting ungulate group presence if at least 1 sign is in the spatial replicate j given presence of ungulate group in the replicate of cell i. r_{sign} = probability of detecting a sign of an individual present along a segment $N_{sign(ij)}$ = total number of signs actually present at spatial replicate j of cell i. Hence, given presence of ≥ 1 sign in the replicate, the detection of sign at replicate j of cell i, d_{ij} was simulated as a Bernoulli random deviate,

$$d_{ij} \sim Bernoulli(p_{sign(ij)})$$

In simulations the following parameters were varied: (1) animal group density per grid cell via the total number of animal group activity centres N of the grid,

(2) the number of signs deposited by each animal group N_{sign}, (3) the radius of the circle that represents the daily movement range R of an individual animal group and (4) the probability of detecting an individual animal groups sign, r_{sign}, on a spatial replicate, although we did not assess for heterogeneity in r_{sign} between replicates. We expected, a priori, that the value of R will play a critical role in the estimates of abundance.

In addition to the initial large sample of 400 grid cells, simulations were also run to investigate model performance with fewer cells to cover a smaller area. We simulate 100 data sets based on 183 plausible scenarios to evaluate the performance of the AOS model. All simulations were coded and executed using software package R (R Development Core Team, 2009). The simulation and the estimation codes are available upon request.

3.4.1.3 Model validation using simulations

Counts of signs generated from the simulated data from spatial replicate j in cell i, were stored in matrix form (number of grid cells [sites] x number of segments [spatial replicates] in each cell). For example, in a survey for animal signs as in (Figure 3.2) that consists of 8 spatial replicates within a cell, the vector representing actual number of signs in a cell $N_{sign(i.)}$ may look as follows.

$N_{sign(i.)} = \{23\ 24\ 0\ 2\ 4\ 68\ 23\ 5\}$

By reducing this numerical vector into a binary detection and non-detection form,

$w_{i.} = \{1\ 1\ 0\ 1\ 1\ 1\ 1\ 1\}$

Detection vector $w_{i.}$ will take this form if the value of $r_{sign} = 1$, because $p_{sign(ij)} = 1$.

However, if the value of r_{sign} takes any other value between 0 and 1 due to imperfect detections, for example if $r_{sign} = 0.5$, then, one possible manifestation of $w_{i.}$ can be,

$w_{i.} = \{1\ 1\ 0\ 0\ 1\ 1\ 1\ 1\}$

The extra zero that now appears in $w_{i.}$ in the 4^{th} column is a chance consequence arising out of the fact that $p_{sign(i4)} = 1 - (1 - 0.5)^2 = 0.75$(which is < 1). Applying the same logic at the cell level:

$$p_{animal(i)} = 1 - (1 - r_{animal})^{N_{animal(i)}}$$

where, $p_{animal(i)}$ is the probability of detecting at least one animal group in cell i in a replicate. r_{animal} is the probability of detecting an individual animal group in a replicate. $N_{animal(i)}$ is the number of animal groups in cell i.

In our sampling situation, the probability of detecting an individual animal group r_{animal} is the result of the product: r_{animal} = (probability of an individual group located in the vicinity of a replicate) x (probability of detecting at least one sign of an individual group on replicate conditional on group located in the vicinity).

Our sampling situation and protocol does not provide sufficient information to estimate the above two probabilities describing r_{animal} separately. Further, the quantity r_{animal} can vary between replicates within a cell as a consequence of the above two probabilities. In reality, we would expect the second probability, i.e., the probability of detecting at least one sign of an individual group on replicate, to very rapidly approach unity because the quantum of signs usually deposited by an ungulate group is very high (usually in several thousands). So our main concern is the possibility of variation in the quantity r_{animal} as a result of the location of a spatial replicate relative to the location of ungulate groups because the AOS model that we apply is based on the assumption that r_{animal} will be a constant. Our objective, is to investigate how the AOS model performs in spite of this assumption violation under varying conditions of these parameters in our simulations (Appendices S1, S2 and S3).

We considered the application of the poisson version of the likelihood (Royle & Nichols, 2003), which takes the form,

$$L(w_{i.}|r_{animal},\lambda) = \prod_{i=1}^{M}\sum_{k=0}^{K}\binom{T_i}{w_{i.}}\left[1-(1-r_{animal})^k\right]^{w_{i.}} \times\left[(1-r_{animal})^k\right]^{T_i-w_{i.}}\left[\frac{e^{-\lambda}\lambda^k}{k!}\right]$$

where $w_{i.}$ is the number of detections in cell i (i.e. the sum of the vector $d_{i.}$). λ is the average number of animal groups in each cell M is the total number of cells. T_i is the number of trials (sampled replicates) in cell i. K is the upper limit of the number of animal groups in a cell.

We assessed the quality of the above estimator for estimating λ against simulated data (see Appendices S1, S2 and S3) using the Root Mean Squared Error (RMSE; Casella & Berger, 1990):

$$RMSE(\hat{\lambda}) = \sqrt{Var(\hat{\lambda}) + \left[\bar{\hat{\lambda}} - \lambda_{true}\right]^2}$$

where λ_{true} = true value of λ defined in the simulation. We also fit the overdispersed version (Pielou, 1969, pages 94-95) of the above likelihood to the real data (Royle & Nichols, 2003) using the form:

$$L(w_{i.}|r_{animal},\lambda) = \prod_{i=1}^{M}\sum_{k=0}^{K}\binom{T_i}{w_{i.}}\left[1-(1-r_{animal})^k\right]^{w_{i.}} \times\left[(1-r_{animal})^k\right]^{T_i-w_{i.}}\left[\frac{\Gamma(k+a)}{k!\Gamma(a)}\right] \times\left[\frac{(1/a)\mu}{1+(1/a)\mu}\right]^k\left[\frac{1}{1+(1/a)\mu}\right]^a$$

where, μ is the average number of animal groups in each cell and a is the overdispersion parameter.

Throughout our formulation we describe $N_{animal(i)}$ and λ or μ as number of animal groups rather than number of animals, deviating from the original Royle & Nichols (2003) formulation, because our study species occur in groups of varying sizes (Karanth & Sunquist, 1992). Failure to recognize this behavioral trait will seriously distort our interpretation of the results, because the AOS model assumes that the probability $1 - r_{animal}$ can be multiplied $N_{animal(i)}$ times, implying *independence*.

3.4.2 Field surveys

3.4.2.1 Choice of grid cell size

We choose a grid cell size of 3.24 km^2, which is sufficiently large to circumscribe expected daily movement of the study species based on home range size and movement rates reported in field studies. For sambar (see Dinerstein, 1979; Sankar & Acharya, 2004; Schaller, 1967); for gaur (see Conry, 1989; Sankar *et al.*, 2000); for chital (see Mishra, 1982; Mishra & Wemmer, 1987; Moe & Wegge, 1997; Schaller, 1967); for muntjac (see Barrette, 1977, 1987; Odden & Wegge, 2007); for wild pig (see Baber & Coblentz, 1986; Caley, 1997; Singer, 1981).

The study area in Muthodi (178 km^2) is partitioned into 55 grid cells. Each cell contains nine equally spaced point locations, called "sampling destinations" and the distance between any two adjacent points is 600 m (Figure 3.3). Each spatial replicate is visualized as a 300 m trail segment along the sampling route passing through these 9 sampling destinations. Consequently, each grid cell contained 16 spatial replicates (Figure 3.3).

Survey teams consisting of 2-3 skilled field personnel searched sampling routes in each cell to locate fresh ($<$ 24 hrs old) tracks or dung of ungulates. Teams optimized sign detections by deviating from the survey route by $<$ 50 m on either side to search logging roads, river banks, stream beds, water holes, mineral licks or wallows, all the

while ensuring proximity to nine sampling destinations and uniform spatial coverage of the cell.

The survey was conducted after light showers when fresh signs were clearly visible and detectable. Thus, we could reasonably assume that the individual animal depositing a sign was present within the cell. Field teams surveyed each cell within $< 8 - 10$ hrs to limit potential biases induced by animal movements between cells (Figure 3.3).

3.4.2.2 Line transect surveys of ungulates

To derive independent estimates of ungulate density, we also used the standard line transect sampling method (Buckland *et al.*, 2001) based on visual detections. This method has been effectively employed since 1988 (Karanth & Sunquist, 1992; Jathanna *et al.*, 2003; Karanth *et al.*, 2004) and this study followed field and analytical protocols described elsewhere (Thomas & Karanth, 2002; Karanth *et al.*, 2002). We used a subset of 11 line transects used for long term monitoring of ungulate populations (Karanth *et al.*, 2008). The survey generated using design options in program DISTANCE 5.0 (Thomas *et al.*, 2010) employed systematic transect placements at 3-km spacing, with a random start (Figure 3.1). Each 3.2 km long transect sampler is of a square geometry, with each side of 800m in length. Each spatial replicate was sampled eight times - thus a total walk-effort of 272.8 km was invested.

Line transect data were analyzed using program DISTANCE for each ungulate species. We fitted several plausible detection probability models generated under hazard-rate, half-normal and uniform detection functions to the observed distance data, to select the most appropriate models (see Buckland *et al.*, 2001). We generated estimates of detection probability, animal cluster densities, cluster size and animal densities (Thomas & Karanth, 2002).

3.5 Results

3.5.1 Simulations of occupancy models

Results of the various simulation runs are tabulated in Appendices S1, S2 and S3. We report only the key findings from these simulations. The RMSE values indicates that the model performed optimally (least relative RMSE value) when the number of animal groups per cell varied from 0.5-4 (Appendices S1 and S2). The RMSE values indicate (Appendices S1 and S2) that the model performance improved when the number of signs deposited by each animal group is increased. Similarly, in the face of imperfect detection of signs ($r_{sign} = 0.5$), the quality of the estimation drops, albeit only marginally. However, because the number of signs deposited is usually so

large that the effect of a decrease in r_{sign} is relatively negligible because p_{sign} would asymptote to a value very close to 1. We note here that p_{sign} is also influenced by the width of the strip considered in this simulation, which was maintained as a constant in our study.

We observe that the estimate of cell-specific abundance is most strongly influenced by the choice of R, the radius of the daily movement range, as indicated by a positive bias. We conclude that accuracy of the specification of R is critical for estimating animal abundance under the AOS model. We do note that the direction of the bias is consistent (Appendices S1 and S2). From the estimated RMSE values, we observe that the model performs most optimally when the radius of the daily movement range of an animal group is set between 6-15% of the length of a grid cell (Figure 3.4). Alternatively, if we have an accurate estimate of R, this bias can always be corrected.

We examined the effect of reduced sample sizes of grid cells because,our field occupancy survey of large ungulates contain fewer than the 400 cells used in all other simulations. These results show that while the mean estimate of the parameter λ is positively biased by reduction in sample size, the median value of the estimate remains relatively unbiased. This is the classical small-sample size bias. These results confirm the desirability of having as many grid-cells (say 200+) of the right size with > 8 replicates for reliable estimation of λ (also see Royle & Nichols, 2003). However, when the number of signs deposited is very large, the quality of the estimation improves drastically (see RMSE results in Appendix S1).

3.5.2 Estimates from Field Surveys

3.5.2.1 Influence of heterogeneity in detection probabilities on estimates of occupancy

The standard occupancy model (MacKenzie *et al.*, 2002) assumes that sign detection probabilities do not vary among cells even if animal abundance varies across them, whereas the AOS model permits such variation and attributes it to variation in cell-specific abundances. For four of the five study species (chital, sambar, gaur and wild pig), the occupancy models of AOS model fit the data better than the standard MacKenzie *et al.* (2002) model (ΔAIC = +36.43, +38.05, +59.38 and +5.80 for chital, sambar, gaur and wild pig, respectively; Table 3.1). For muntjac, the smallest ungulate, the reverse was true (ΔAIC = -2.01). These results indicate that heterogeneity in detection probabilities among cells, presumably due to varying cell-specific abundances does exist for all ungulates except muntjac, although the wide standards errors weakens the inference.

3.5.2.2 Density estimates from AOS model

The parameter estimates, $r_{\hat{animal}}$ and $\hat{\lambda}$, along with the computed correlation coefficient between the two parameters are reported (Table 3.2) where $\hat{\lambda}$ refers to the average number of animal groups present within one grid cell of size 3.24 km^2 and $r_{\hat{animal}}$ refers to group-specific detection probabilities for each species. Due to the relatively small sample size attained in the field survey (number of cells = 55 < 200+ suggested by the simulations), we obtained a correlation coefficient of less than -0.9 between the estimated parameters $r_{\hat{animal}}$ and $\hat{\lambda}$ for all species surveyed. Hence, we view the estimate with some caution, although the median estimate of λ in our simulations (Appendix S1) for low sample sizes suggests these estimates are reliable. However, in the extreme case where the correlation coefficient between $r_{\hat{animal}}$ and $\hat{\lambda}$ is as extreme as -0.98 (for muntjac data), and estimated r_{animal} is also very low (0.03), we are concerned with the estimate of λ in our survey. This is because we are not sure if r is indeed very low or whether it is an outcome due to high correlation with λ during estimation.

Our results also provide evidence that the study area was fairly homogenous, and consistent with our *a priori* assumption, because the over-dispersed version of the Royle & Nichols (2003) received considerably lower support (ΔAIC> 2) relative to the poisson model for all species, except gaur (ΔAIC= 1.432) (Table 3.2). However, the correlation coefficient between $r_{\hat{animal}}$ and $\hat{\lambda}$ was very high ($\rho_{r\mu}$=-0.98) for the gaur analysis, indicating that the parameters r_{animal} and μ are virtually non-identifiable causing concern in the parameter estimates with the over-dispersed model.

3.5.2.3 Density estimates from line transect surveys

The line transect surveys yielded 20 to 59 visual detections for chital, sambar, gaur and muntjac, enabling us to fit detection functions reasonably well and generate density estimates (Table 3.3). The number of detections for wild pigs (n=12) was inadequate for reliable density estimation (Buckland *et al.*, 2001). Therefore, we borrowed density estimates derived from a larger set of 24 transects from an expanded area around the study site and assumed pig densities to be uniform across the wider area. Our results show that despite a fairly high sampling effort (273 km), sample sizes from visual detections were very low.

Half-normal detection key function models in combination with cosine adjustment of orders produced the best fit for sambar and chital sighting data while the uniform + cosine adjustment models fitted muntjac and gaur detection data best (Table 3). However, a combination of few spatial replicates ($n = 11$) and detection events ($n < 60$) produced wide standard errors. Comparisons (Table 3.4) between

the two survey methods show that cluster density estimates, despite being derived from two different approaches, matched well for all species except muntjac.

3.6 Discussion

Understanding the processes influencing forest structure and dynamics by large ungulates requires reliable information on their densities. However, consequent upon anthropogenic pressures, large ungulate densities have declined drastically - reducing them to occur in very low numbers in most remaining forests. Additionally, logistical issues preclude the application of standard methods such as distance sampling or capture-recapture sampling to estimate their densities. Therefore, there is a premium in developing approaches that enable assessments of absolute or relative densities that directly address the issue of imperfect detections.

Our study provides one possible solution to this problem. We estimate densities of four large ungulate species from simple, cost-effective sign surveys using the AOS model in a region where ungulate densities are relatively low. It is important for investigators to incorporate biological knowledge of a target species in the design of such a sign survey. Of critical importance is the reasonably accurate specification of expected daily range sizes of animals to establish cell sizes and consideration of the length of spatial replicates. However, simulation results show that although underestimation or overestimation of values assumed for radius of activity R can bias estimates of density, the direction of this bias appears predictable. This consistency in direction of bias makes our approach a useful tool, at least as an 'index', for monitoring ungulate densities across space and time. On a related note, if accurate information on R, or the number of cells expected to be visited by each animal group is based on prior knowledge (e.g., via radio-telemetry), this information can be profitably incorporated into the survey design. In our specific study, the specification of cell size was targeted primarily for larger ungulate species, and turned out too large for muntjac, the smallest among five species considered. As a result our survey data may have failed to capture the evidence of hetereogeneity in detection probability among cells for muntjac because the cell specific abundance value encountered (> 12 animal groups per cell) may have been too high. We emphasize that the AOS model approach works better when expected densities of ungulates are low enabling the capturing of heterogeneous detection rates among cells that is critical for the model to work.

Our study is a preliminary step towards reliably estimating absolute abundance of large ungulate species from field counts of signs along trails using spatial replications. While our approach employs survey design and protocols based on species biology, and the approach is flexible to accommodate for varying group sizes, we

recognize that supplementary information in the form of prior data on daily movement ranges, sizes of groups and other covariates influencing ungulate abundance and detection rates across space, can all improve the reliability of this approach. At another large site in the same region Rayar (2010) has also confirmed the reliability of this approach more recently using additional covariates to describe cell-specific λ. In addition to density, environmental factors (substrate type, moisture and local weather) may often be determinants of encounter rates of animal signs (tracks, dung) during surveys. These factors can vary widely across or within cells. However, our approach of aggregation of survey data into simple binary counts, combined with a systematic sampling regime ensures that fluctuations in sign encounter rates do not unduly affect the estimation process. Further, if required all these environmental factors can be incorporated in the modeling of r_{animal} directly Royle & Nichols (see 2003) under this approach. Future research in this domain can focus on models that account for unmodeled heterogeneity in detection probability and spatial autocorrelation between spatial replicates, which are not accounted for in this study. From a practical standpoint, this approach is most justified when the aim is to find a cost-effective approach when animal abundance is low. Unlike in transect surveys, whereby "pre-marking" of transect lines is sometimes essential (in our case, about 3 months in the field), the AOS method requires little preparation time and is thus cost-effective.

This approach can also potentially be extended from sign surveys to other animal detection techniques such as camera traps, scent stations, and track-plates, which are currently used to assess populations of other animal taxa (e.g., Nag, 2008). It has recently been applied to assess better prevalence in wildlife disease (Lachish *et al.*, 2012) Therefore, the potential utility of our approach may extend to include a wide range of animal monitoring contexts. But, more immediately, we foresee our approach to be utilized for studies of ungulate species of high conservation priority, including the Sao La *Pseudoryx vukongensis*, Malay Tapir *Tapirus indicus* or the Sumatran Rhinoceros *Dicerorhinus sumatrensis*, which are very rare species and unlikely to be surveyed well with other methods.

3.7 Figures

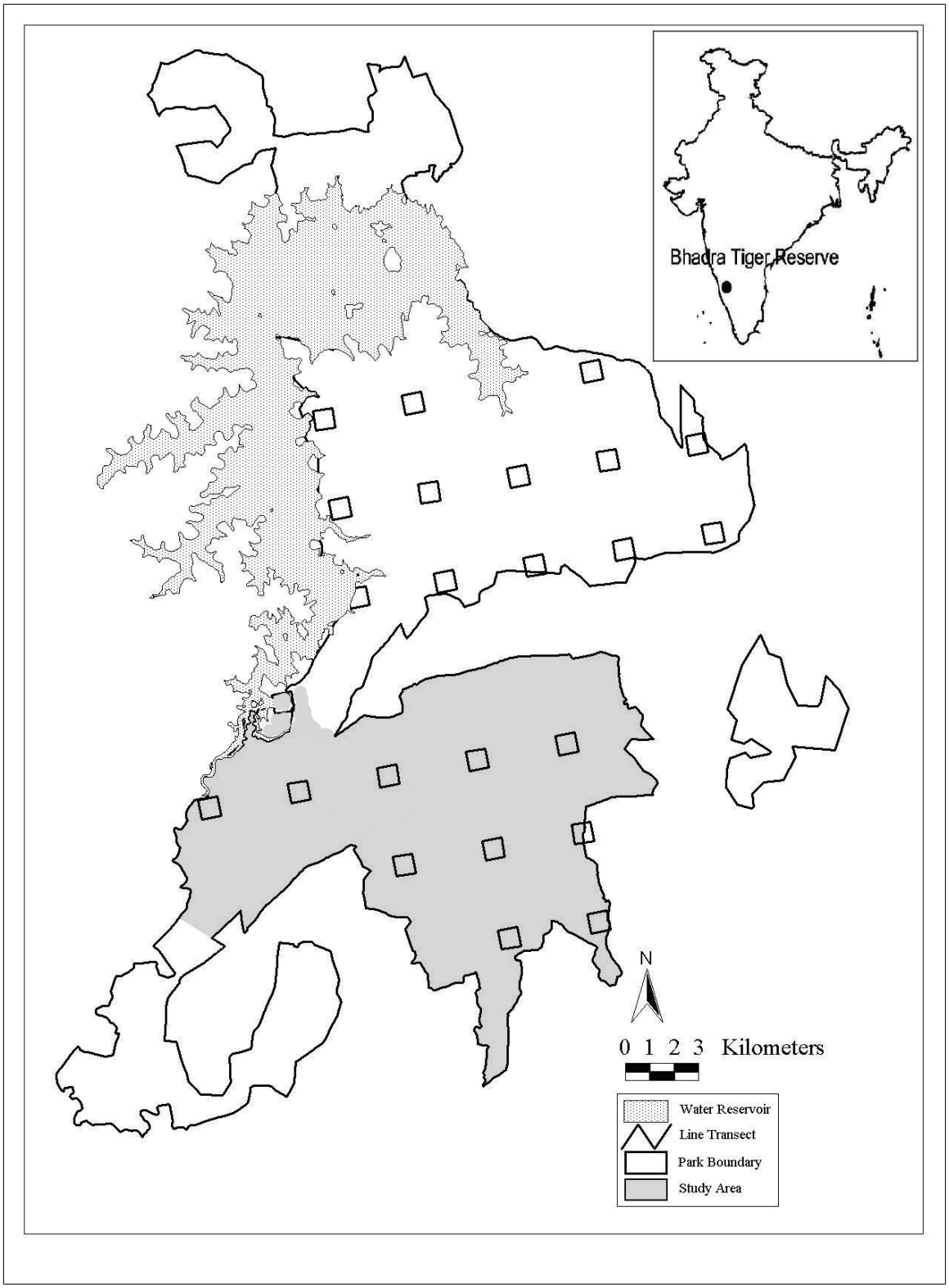


Figure 3.1: Map of Bhadra Tiger Reserve with a system of line transects (inset map shows the location of the park in India). The study area (Muthodi) is shaded grey

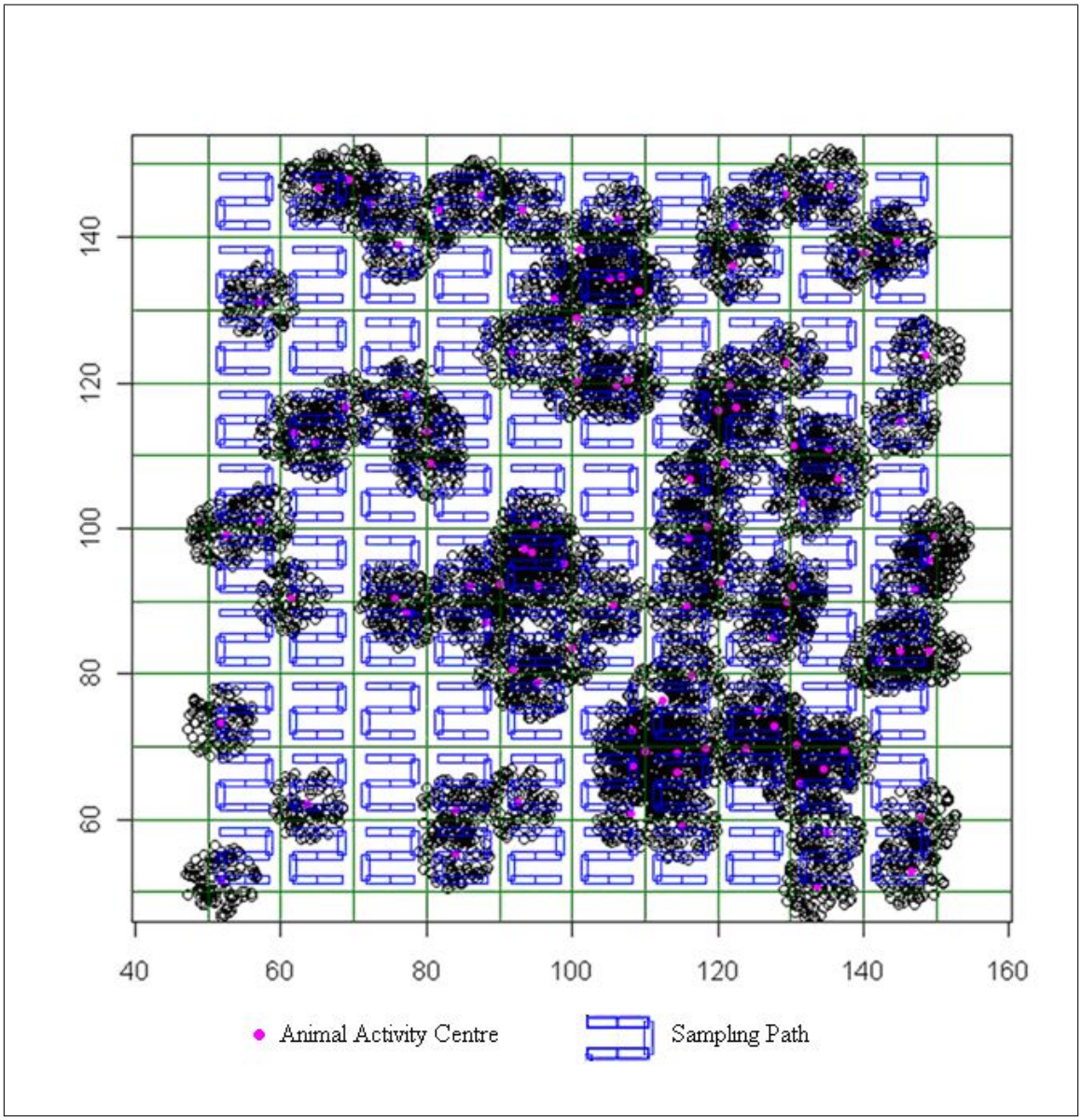


Figure 3.2: An example of a simulation output in R describing the ecological process of sign deposition (black circles) around animal group activity centres (pink dots), within 100 sampling grid cells.

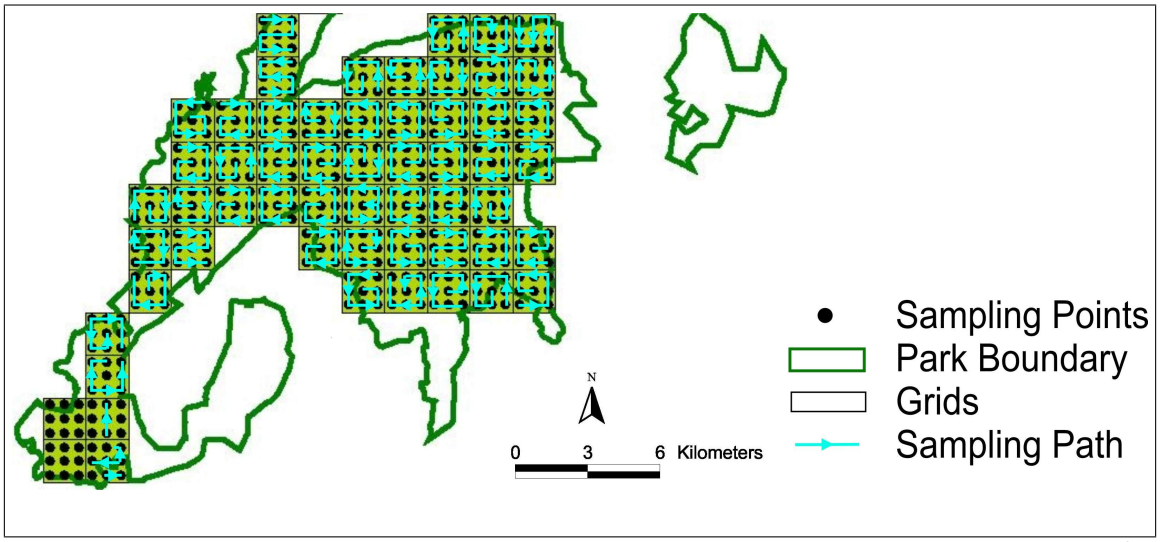


Figure 3.3: Ungulate sign survey conducted in 55 grid cells (each of size 3.24 km^2) within Muthodi (Bhadra Tiger Reserve)

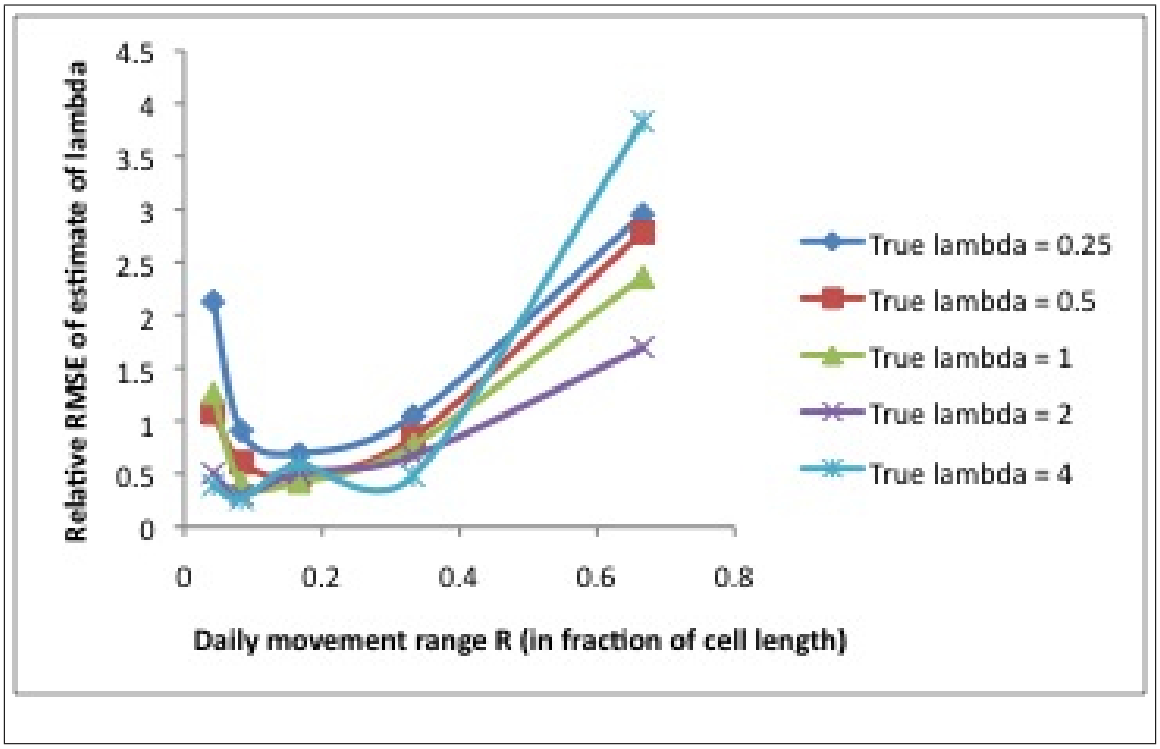


Figure 3.4: Effect of daily movement range R on the relative RMSE value of the estimate of $\hat{\lambda}$ from simulation exercises. Graph demonstrates that low or high values of R produces large relative RMSE values.

3.8 Tables

Table 3.1: Comparison of performances of habitat occupancy models of MacKenzie *et al.* (2002) model and AOS model using data from sign surveys of ungulate species in Bhadra Tiger Reserve, India

Species	Number of spatial replicates per cell	AIC value from MacKenzie *et al.* (2002) model	AIC value from AOS model	ΔAIC
Chital	16	888.10	851.67	+36.43
Sambar	16	1147.01	1108.96	+38.05
Gaur	16	1187.56	1128.18	+59.38
Pig	16	765.26	759.46	+5.80
Muntjac	16	688.78	690.79	-2.01

Table 3.2: Estimates of individual group-specific detection probability and cell-specific animal group abundance generated from AOS model using data from sign surveys of ungulate species in Bhadra Tiger Reserve, India. $\hat{r}$ is the estimate of animal group-specific detection probability, $\hat{\lambda}$ or $\hat{\mu}$ is the estimate of average number of animal groups in a cell, $\hat{a}$ is the overdispersion parameter, $\hat{\rho}$ is the estimate of the correlation coefficient.

		AOS-Poisson			AOS-Overdispersed Poisson				
Species	No. of spatial replicates	$\hat{r}[\hat{SE}(\hat{r})]$	$\hat{\lambda}[\hat{SE}(\hat{\lambda})]$	$\hat{\rho_{r\lambda}}$	$\hat{r}[\hat{SE}(\hat{r})]$	$\hat{\mu}[\hat{SE}(\hat{\mu})]$	$\hat{a}[\hat{SE}(\hat{a})]$	$\hat{\rho_{r\mu}}$	Δ AIC
Chital	16	0.31 (0.07)	4.74 (1.23)	-0.96	0.30 (0.01)	4.98 (1.93)	71.6 (379.05)	-0.98	+2.03
Sambar	16	0.20 (0.05)	4.61 (1.21)	-0.96	0.08 (0.47)	12.13 (74.65)	6.12 (21.11)	-1.0	+2.07
Gaur	16	0.20 (0.04)	3.30 (0.80)	-0.93	0.03 (0.02)	23.71 (12.92)	3.04 (0.92)	-0.98	-1.43
Pig	16	0.08 (0.02)	2.41 (0.70)	-0.91	0.07 (0.025)	2.47 (0.875)	71.60 (525.49)	-0.94	+2.02
Muntjac	16	0.03 (0.02)	4.60 (2.86)	-0.99	0.03 (0.02)	4.84 (3.509)	71.59 (363.08)	-0.99	+2.04

$^{+}$ A group here is defined as a 'social group', while a cluster is defined as a momentary congregation of animals. Group size $\geq$ cluster size. For the species we refer to here, group size $\simeq$ cluster size.

* Estimates for wild pigs were derived from a larger study area.

Table 3.3: Results of density estimation for ungulate species using line transect sampling data at Bhadra Tiger Reserve. Model choice (HNcos=Half-normal model with cosine adjustment terms; UNcos=Uniform model with cosine adjustment terms), number of sightings (n after right truncation of data), and estimates of detection probability ($\hat{p}$), cluster size ($\hat{S}$), cluster density (D_s) and animal density (D) along with their standard errors (SE) are reported.

Species	Model	n	$\hat{p}$	$\hat{S}$ $\hat{SE}(\hat{S})$	$\hat{D_s}$ $\hat{SE}(\hat{D_s})$	$\hat{D}$ $\hat{SE}(\hat{D})$
Sambar	HN-cos	59	0.49	1.78(0.12)	2.47(0.56)	4.40(1.03)
Chital	HN-cos	47	0.33	4.11(0.81)	2.16(0.76)	8.88(3.58)
Muntjac	UN-cos	49	0.52	1.08(0.04)	4.03(0.83)	4.35(0.91)
Gaur	UN-cos	21	0.59	4.10(0.90)	0.94(0.30)	3.86(1.49)
Wild pig	UN-cos	29	0.61	2.24(0.39)	1.10(0.37)	2.46(0.94)

Table 3.4: Comparison of sampling effort and density estimates for ungulate species in Bhadra Tiger Reserve generated from sign surveys using AOS model and from line transect surveys based on visual detections

	AOS approach		Distance sampling	
Species	Sampling effort (in km)	Density of animal groups^{+} $\hat{D}_s \left[\hat{SE}(\hat{D}_s)\right]$	Sampling effort (in km)	Density of animal clusters^{+} $\hat{D}_s \left[\hat{SE}(\hat{D}_s)\right]$
Chital	256.80	1.46 (0.68)	272.78	2.16 (0.76)
Sambar	256.80	1.42 (0.67)	272.78	2.47 (0.56)
Gaur	256.80	1.01 (0.44)	272.78	0.94 (0.30)
Wild Pig	256.80	0.74 (0.39)	523.68*	1.09 (0.37)*
Muntjac	256.80	1.42 (1.59)	272.78	4.03 (0.83)

3.9 Appendices

Appendix S1

Results of simulations for 90 scenarios run on 100 datasets assuming an underlying Poisson distribution where sign-specific detection probability for the sign of an individual group was set at equal to 0.999. The simulations were run assuming different values for the radius of movement within which signs are deposited *R* (represented as a fraction of the length of a grid cell), the number of signs deposited by each animal groups (N_{sign}) and the total number of animal groups *S* in the grid. The width of the detection strip *d* is represented as the fraction of the length of a grid cell. See *Methods* for description of other parameters.

S	N_{sign}	*R*	No. of cells	r_{sign}	*d*	Data sets	Mean ($\hat{r}_{animal}$)	SD [Mean ($\hat{r}_{animal}$)]	LCL ($\hat{r}_{animal}$)	UCL ($\hat{r}_{animal}$)	Median ($\hat{r}_{animal}$)	Mean ($\hat{\lambda}$)	SD [Mean($\hat{\lambda}$)]	LCL ($\hat{\lambda}$)	UCL($\hat{\lambda}$)	Median ($\hat{\lambda}$)	True λ	RMSE ($\hat{\lambda}$)
100	50	0.042	400	0.999	0.02	100	0.041	0.020	0.007	0.077	0.038	0.309	0.254	0.111	1.288	0.235	0.250	0.507
100	**50**	**0.083**	**400**	**0.999**	**0.02**	**100**	**0.066**	**0.021**	**0.021**	**0.106**	**0.065**	**0.307**	**0.159**	**0.164**	**0.738**	**0.257**	0.250	0.403
100	50	0.167	400	0.999	0.02	100	0.094	0.018	0.059	0.130	0.094	0.384	0.068	0.287	0.557	0.376	0.250	0.294
100	50	0.333	400	0.999	0.02	100	0.143	0.018	0.105	0.179	0.140	0.408	0.050	0.326	0.529	0.403	0.250	0.274
100	50	0.667	400	0.999	0.02	100	0.108	0.016	0.077	0.135	0.107	0.677	0.102	0.527	0.882	0.675	0.250	0.533
100	200	0.042	400	0.999	0.02	100	0.046	0.019	0.014	0.082	0.046	0.268	0.216	0.116	0.694	0.216	0.250	0.465
100	**200**	**0.083**	**400**	**0.999**	**0.02**	**100**	**0.081**	**0.019**	**0.044**	**0.119**	**0.080**	**0.259**	**0.063**	**0.169**	**0.428**	**0.247**	0.250	0.251
100	200	0.167	400	0.999	0.02	100	0.143	0.016	0.112	0.172	0.143	0.326	0.034	0.269	0.397	0.324	0.250	0.199
100	200	0.333	400	0.999	0.02	100	0.268	0.021	0.221	0.305	0.267	0.421	0.034	0.368	0.497	0.417	0.250	0.251
100	200	0.667	400	0.999	0.02	100	0.283	0.016	0.251	0.318	0.285	0.780	0.052	0.701	0.891	0.775	0.250	0.577
100	800	0.042	400	0.999	0.02	100	0.050	0.021	0.012	0.090	0.049	0.273	0.282	0.115	0.792	0.206	0.250	0.532
100	**800**	**0.083**	**400**	**0.999**	**0.02**	**100**	**0.086**	**0.019**	**0.050**	**0.124**	**0.085**	**0.251**	**0.052**	**0.168**	**0.375**	**0.247**	0.250	0.227
100	800	0.167	400	0.999	0.02	100	0.158	0.014	0.131	0.183	0.158	0.317	0.025	0.271	0.369	0.319	0.250	0.173
100	800	0.333	400	0.999	0.02	100	0.311	0.022	0.270	0.354	0.313	0.436	0.033	0.379	0.502	0.434	0.250	0.261
100	800	0.667	400	0.999	0.02	100	0.443	0.017	0.412	0.474	0.443	0.950	0.054	0.861	1.062	0.936	0.250	0.738
200	50	0.042	400	0.999	0.02	100	0.039	0.015	0.010	0.070	0.039	0.642	0.671	0.266	1.670	0.492	**0.500**	0.832
200	**50**	**0.083**	**400**	**0.999**	**0.02**	**100**	**0.064**	**0.015**	**0.033**	**0.093**	**0.064**	**0.597**	**0.213**	**0.401**	**1.052**	**0.537**	**0.500**	0.472

200	50	0.167	400	0.999	0.02	100	0.093	0.014	0.064	0.121	0.094	0.766	0.117	0.587	1.090	0.747	**0.500**	0.433
200	50	0.333	400	0.999	0.02	100	0.144	0.014	0.114	0.168	0.145	0.808	0.081	0.688	1.003	0.798	**0.500**	0.419
200	50	0.667	400	0.999	0.02	100	0.109	0.014	0.083	0.139	0.108	1.326	0.175	1.062	1.712	1.307	**0.500**	0.926
200	200	0.042	400	0.999	0.02	100	0.045	0.015	0.015	0.075	0.045	0.524	0.286	0.271	1.454	0.464	**0.500**	0.536
200	**200**	**0.083**	**400**	**0.999**	**0.02**	**100**	**0.081**	**0.014**	**0.057**	**0.108**	**0.082**	**0.512**	**0.100**	**0.396**	**0.747**	**0.482**	**0.500**	0.317
200	200	0.167	400	0.999	0.02	100	0.144	0.013	0.123	0.170	0.143	0.644	0.050	0.551	0.738	0.644	**0.500**	0.266
200	200	0.333	400	0.999	0.02	100	0.269	0.015	0.238	0.301	0.268	0.837	0.048	0.743	0.960	0.836	**0.500**	0.402
200	200	0.667	400	0.999	0.02	100	0.291	0.017	0.262	0.323	0.292	1.524	0.101	1.337	1.709	1.520	**0.500**	1.072
200	800	0.042	400	0.999	0.02	100	0.047	0.016	0.014	0.075	0.047	0.523	0.289	0.270	1.473	0.459	**0.500**	0.538
200	**800**	**0.083**	**400**	**0.999**	**0.02**	**100**	**0.087**	**0.014**	**0.059**	**0.116**	**0.088**	**0.498**	**0.092**	**0.388**	**0.716**	**0.480**	**0.500**	0.303
200	800	0.167	400	0.999	0.02	100	0.159	0.012	0.137	0.186	0.158	0.625	0.041	0.562	0.705	0.623	**0.500**	0.238
200	800	0.333	400	0.999	0.02	100	0.316	0.017	0.283	0.346	0.316	0.861	0.048	0.769	0.962	0.855	**0.500**	0.422
200	800	0.667	400	0.999	0.02	100	0.453	0.018	0.415	0.482	0.454	1.854	0.101	1.694	2.059	1.853	**0.500**	1.391
400	50	0.042	400	0.999	0.02	100	0.040	0.014	0.010	0.067	0.040	1.115	0.858	0.514	3.425	0.932	1.000	0.933
400	**50**	**0.083**	**400**	**0.999**	**0.02**	**100**	**0.064**	**0.014**	**0.037**	**0.093**	**0.062**	**1.186**	**0.407**	**0.774**	**1.856**	**1.127**	1.000	0.664
400	50	0.167	400	0.999	0.02	100	0.091	0.013	0.068	0.117	0.091	1.557	0.219	1.228	2.071	1.536	1.000	0.728
400	50	0.333	400	0.999	0.02	100	0.145	0.015	0.116	0.172	0.145	1.610	0.161	1.339	1.995	1.605	1.000	0.730
400	50	0.667	400	0.999	0.02	100	0.110	0.015	0.084	0.137	0.111	2.640	0.369	2.032	3.368	2.588	1.000	1.749
400	200	0.042	400	0.999	0.02	100	0.045	0.014	0.017	0.070	0.044	1.076	1.199	0.557	2.654	0.863	1.000	1.098
400	**200**	**0.083**	**400**	**0.999**	**0.02**	**100**	**0.080**	**0.013**	**0.053**	**0.105**	**0.079**	**1.032**	**0.191**	**0.763**	**1.486**	**1.011**	1.000	0.438
400	200	0.167	400	0.999	0.02	100	0.141	0.013	0.119	0.167	0.139	1.324	0.121	1.112	1.554	1.323	1.000	0.476
400	200	0.333	400	0.999	0.02	100	0.272	0.016	0.244	0.302	0.272	1.662	0.103	1.479	1.869	1.656	1.000	0.735
400	200	0.667	400	0.999	0.02	100	0.309	0.017	0.277	0.338	0.310	2.853	0.185	2.515	3.228	2.839	1.000	1.903
400	800	0.042	400	0.999	0.02	100	0.048	0.014	0.018	0.075	0.047	1.105	1.576	0.567	2.464	0.847	1.000	1.260
400	**800**	**0.083**	**400**	**0.999**	**0.02**	**100**	**0.085**	**0.013**	**0.060**	**0.112**	**0.085**	**1.001**	**0.177**	**0.732**	**1.359**	**0.983**	1.000	0.420
400	800	0.167	400	0.999	0.02	100	0.158	0.013	0.133	0.189	0.156	1.274	0.104	1.078	1.487	1.271	1.000	0.424
400	800	0.333	400	0.999	0.02	100	0.320	0.017	0.288	0.352	0.319	1.705	0.102	1.541	1.915	1.691	1.000	0.774
400	800	0.667	400	0.999	0.02	100	0.495	0.020	0.457	0.531	0.493	3.323	0.174	3.013	3.672	3.328	1.000	2.361
800	50	0.042	400	0.999	0.02	100	0.039	0.012	0.013	0.062	0.041	2.486	2.941	1.255	5.722	1.830	2.000	1.782
800	**50**	**0.083**	**400**	**0.999**	**0.02**	**100**	**0.062**	**0.013**	**0.038**	**0.084**	**0.062**	**2.411**	**0.558**	**1.714**	**3.816**	**2.288**	2.000	0.853

800	50	0.167	400	0.999	0.02	100	0.087	0.015	0.060	0.113	0.086	3.342	0.595	2.473	4.514	3.247	2.000	1.547
800	50	0.333	400	0.999	0.02	100	0.147	0.015	0.115	0.172	0.147	3.176	0.338	2.646	4.054	3.148	2.000	1.312
800	50	0.667	400	0.999	0.02	100	0.116	0.018	0.085	0.149	0.116	5.027	0.816	3.857	6.976	4.957	2.000	3.159
800	200	0.042	400	0.999	0.02	100	0.044	0.013	0.018	0.063	0.045	2.192	2.194	1.273	4.369	1.720	2.000	1.494
800	**200**	**0.083**	**400**	**0.999**	**0.02**	**100**	**0.076**	**0.013**	**0.054**	**0.102**	**0.077**	**2.177**	**0.388**	**1.618**	**3.030**	**2.116**	2.000	0.648
800	200	0.167	400	0.999	0.02	100	0.130	0.018	0.098	0.166	0.128	2.907	0.420	2.276	3.788	2.929	2.000	1.114
800	200	0.333	400	0.999	0.02	100	0.283	0.021	0.246	0.321	0.284	3.188	0.264	2.789	3.678	3.179	2.000	1.294
800	200	0.667	400	0.999	0.02	100	0.347	0.027	0.294	0.391	0.343	4.994	0.422	4.315	5.850	4.959	2.000	3.064
800	800	0.042	400	0.999	0.02	100	0.047	0.013	0.020	0.069	0.048	1.964	1.015	1.235	3.885	1.719	2.000	1.008
800	**800**	**0.083**	**400**	**0.999**	**0.02**	**100**	**0.081**	**0.013**	**0.059**	**0.104**	**0.081**	**2.112**	**0.354**	**1.624**	**2.817**	**2.095**	2.000	0.605
800	800	0.167	400	0.999	0.02	100	0.145	0.018	0.103	0.175	0.147	2.805	0.390	2.307	3.844	2.741	2.000	1.019
800	800	0.333	400	0.999	0.02	100	0.335	0.021	0.290	0.374	0.332	3.252	0.229	2.877	3.764	3.248	2.000	1.340
800	800	0.667	400	0.999	0.02	100	0.587	0.032	0.530	0.660	0.585	5.334	0.366	4.596	5.968	5.350	2.000	3.388
1600	50	0.042	400	0.999	0.02	100	0.038	0.014	0.016	0.070	0.038	4.656	3.602	2.088	9.201	3.841	4.000	2.008
1600	**50**	**0.083**	**400**	**0.999**	**0.02**	**100**	**0.059**	**0.012**	**0.036**	**0.085**	**0.060**	**5.021**	**1.193**	**3.336**	**7.882**	**4.775**	**4.000**	1.495
1600	50	0.167	400	0.999	0.02	100	0.086	0.017	0.053	0.117	0.086	6.814	1.513	4.794	10.632	6.528	4.000	3.071
1600	50	0.333	400	0.999	0.02	100	0.151	0.021	0.118	0.185	0.149	6.215	0.953	4.970	8.042	6.161	4.000	2.420
1600	50	0.667	400	0.999	0.02	100	0.118	0.023	0.079	0.155	0.115	10.053	2.073	7.313	14.709	9.849	4.000	6.222
1600	200	0.042	400	0.999	0.02	100	0.043	0.014	0.020	0.070	0.045	4.508	4.871	2.223	7.568	3.508	4.000	2.265
1600	**200**	**0.083**	**400**	**0.999**	**0.02**	**100**	**0.074**	**0.014**	**0.044**	**0.102**	**0.072**	**4.535**	**0.982**	**3.144**	**7.450**	**4.484**	**4.000**	1.126
1600	200	0.167	400	0.999	0.02	100	0.124	0.021	0.084	0.161	0.124	6.188	1.115	4.640	8.828	6.041	4.000	2.429
1600	200	0.333	400	0.999	0.02	100	0.314	0.030	0.252	0.374	0.316	5.751	0.611	4.761	7.181	5.652	4.000	1.918
1600	200	0.667	400	0.999	0.02	100	0.361	0.069	0.227	0.478	0.356	9.857	2.343	7.045	15.557	9.468	4.000	6.054
1600	800	0.042	400	0.999	0.02	100	0.046	0.014	0.021	0.071	0.048	4.093	2.447	2.230	7.657	3.449	4.000	1.567
1600	**800**	**0.083**	**400**	**0.999**	**0.02**	**100**	**0.078**	**0.014**	**0.050**	**0.107**	**0.077**	**4.468**	**0.878**	**3.149**	**6.845**	**4.347**	**4.000**	1.047
1600	800	0.167	400	0.999	0.02	100	0.136	0.022	0.098	0.172	0.135	6.092	1.040	4.664	8.157	5.976	4.000	2.328
1600	800	0.333	400	0.999	0.02	100	0.377	0.038	0.292	0.456	0.378	5.757	0.638	4.686	7.293	5.696	4.000	1.930
1600	800	0.667	400	0.999	0.02	100	0.571	0.263	0.100	0.838	0.683	18.680	18.945	6.624	61.946	8.664	4.000	15.31
3200	50	0.042	400	0.999	0.02	100	0.038	0.013	0.011	0.064	0.039	9.392	5.211	4.575	25.522	7.758	8.000	2.674
3200	**50**	**0.083**	**400**	**0.999**	**0.02**	**100**	**0.063**	**0.016**	**0.032**	**0.098**	**0.062**	**9.825**	**2.912**	**5.915**	**18.064**	**9.231**	**8.000**	2.498

3200	50	0.167	400	0.999	0.02	100	0.089	0.024	0.045	0.136	0.086	13.845	4.558	8.172	24.942	13.112	8.000	6.223
3200	50	0.333	400	0.999	0.02	100	0.163	0.030	0.092	0.216	0.164	11.796	2.700	8.678	20.238	11.204	8.000	4.136
3200	50	0.667	400	0.999	0.02	100	0.127	0.036	0.065	0.196	0.126	19.448	6.382	11.651	34.452	17.656	8.000	11.72
3200	200	0.042	400	0.999	0.02	100	0.043	0.013	0.020	0.066	0.045	8.241	3.293	4.692	15.952	7.184	8.000	1.831
3200	**200**	**0.083**	**400**	**0.999**	**0.02**	**100**	**0.076**	**0.018**	**0.044**	**0.115**	**0.075**	**8.910**	**2.291**	**5.582**	**14.567**	**8.582**	**8.000**	1.766
3200	200	0.167	400	0.999	0.02	100	0.134	0.027	0.086	0.186	0.133	11.675	2.533	8.124	17.124	11.291	8.000	4.004
3200	200	0.333	400	0.999	0.02	100	0.360	0.072	0.234	0.490	0.358	10.345	2.330	7.015	15.402	9.794	8.000	2.798
3200	200	0.667	400	0.999	0.02	100	0.251	0.217	0.099	0.822	0.139	42.113	19.901	7.557	62.053	50.465	8.000	34.404
3200	800	0.042	400	0.999	0.02	100	0.045	0.013	0.020	0.069	0.045	8.167	3.539	4.722	16.184	7.258	8.000	1.889
3200	**800**	**0.083**	**400**	**0.999**	**0.02**	**100**	**0.081**	**0.018**	**0.049**	**0.117**	**0.081**	**8.720**	**2.131**	**5.650**	**13.328**	**8.278**	**8.000**	1.628
3200	800	0.167	400	0.999	0.02	100	0.146	0.028	0.097	0.199	0.149	11.470	2.428	7.976	16.761	10.893	8.000	3.803
3200	800	0.333	400	0.999	0.02	100	0.426	0.115	0.175	0.636	0.422	11.108	4.481	6.646	25.164	9.953	8.000	3.761
3200	800	0.667	400	0.999	0.02	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Appendix S2

Results of simulations for 90 different scenarios run on 100 datasets assuming an underlying Poisson distribution where sign-specific detection probability for an individual animal group was set at equal to 0.5. The simulations were run assuming different values for the radius of movement within which signs are deposited *R*(represented as a fraction of the length of a grid cell), the number of signs deposited by each animal cluster (N_{sign}) and the total number of animal clusters *S* in the grid. The width of the detection strip *d* is represented as the fraction of the length of a grid cell. See *Methods* for description of other parameters.

S	N_{sign}	*R*	No. of cells	r_{sign}	*d*	Data sets	Mean ($\hat{r}_{animal}$)	SD [Mean ($\hat{r}_{animal}$)]	LCL ($\hat{r}_{animal}$)	UCL ($\hat{r}_{animal}$)	Median ($\hat{r}_{animal}$)	Mean ($\hat{\lambda}$)	SD [Mean($\hat{\lambda}$)]	LCL ($\hat{\lambda}$)	UCL($\hat{\lambda}$)	Median ($\hat{\lambda}$)	True λ	RMSE ($\hat{\lambda}$)
100	50	0.042	400	0.5	0.02	100	0.038	0.020	0.004	0.088	0.037	0.395	0.479	0.102	1.976	0.230	0.250	0.707
100	50	0.083	400	0.5	0.02	100	0.052	0.017	0.019	0.086	0.051	0.334	0.143	0.160	0.788	0.290	0.250	0.388
100	50	0.167	400	0.5	0.02	100	0.062	0.018	0.031	0.102	0.059	0.444	0.138	0.263	0.779	0.419	0.250	0.419
100	50	0.333	400	0.5	0.02	100	0.087	0.019	0.051	0.123	0.086	0.410	0.096	0.290	0.625	0.391	0.250	0.348
100	50	0.667	400	0.5	0.02	100	0.056	0.014	0.031	0.087	0.055	0.717	0.199	0.413	1.152	0.665	0.250	0.646
100	200	0.042	400	0.5	0.02	100	0.048	0.021	0.015	0.093	0.046	0.237	0.113	0.108	0.529	0.195	0.250	0.337
100	200	0.083	400	0.5	0.02	100	0.076	0.018	0.046	0.104	0.078	0.263	0.070	0.177	0.385	0.243	0.250	0.265
100	200	0.167	400	0.5	0.02	100	0.123	0.016	0.092	0.155	0.123	0.346	0.042	0.282	0.440	0.340	0.250	0.227
100	200	0.333	400	0.5	0.02	100	0.211	0.017	0.177	0.241	0.211	0.410	0.033	0.353	0.478	0.407	0.250	0.242
100	200	0.667	400	0.5	0.02	100	0.186	0.018	0.152	0.223	0.185	0.701	0.067	0.581	0.841	0.695	0.250	0.520
100	800	0.042	400	0.5	0.02	100	0.053	0.021	0.016	0.098	0.050	0.254	0.354	0.096	0.604	0.201	0.250	0.595
100	800	0.083	400	0.5	0.02	100	0.085	0.018	0.054	0.120	0.086	0.252	0.062	0.173	0.390	0.236	0.250	0.249
100	800	0.167	400	0.5	0.02	100	0.153	0.015	0.131	0.184	0.152	0.320	0.028	0.270	0.369	0.321	0.250	0.182
100	800	0.333	400	0.5	0.02	100	0.298	0.021	0.259	0.343	0.299	0.429	0.031	0.375	0.484	0.426	0.250	0.250
100	800	0.667	400	0.5	0.02	100	0.379	0.018	0.350	0.420	0.376	0.877	0.053	0.779	0.970	0.885	0.250	0.667
200	50	0.042	400	0.5	0.02	100	0.034	0.017	0.006	0.065	0.033	0.715	0.616	0.237	2.530	0.493	0.500	0.814

200	50	0.083	400	0.5	0.02	100	0.049	0.016	0.018	0.079	0.049	0.726	0.355	0.386	1.659	0.638	0.500	0.637
200	50	0.167	400	0.5	0.02	100	0.061	0.015	0.031	0.090	0.061	0.890	0.282	0.554	1.512	0.836	0.500	0.659
200	50	0.333	400	0.5	0.02	100	0.086	0.014	0.062	0.113	0.087	0.800	0.131	0.614	1.102	0.777	0.500	0.471
200	50	0.667	400	0.5	0.02	100	0.059	0.014	0.034	0.088	0.059	1.337	0.319	0.870	2.151	1.297	0.500	1.010
200	200	0.042	400	0.5	0.02	100	0.043	0.017	0.016	0.075	0.041	0.527	0.267	0.232	1.234	0.439	0.500	0.517
200	200	0.083	400	0.5	0.02	100	0.075	0.018	0.043	0.108	0.077	0.536	0.134	0.339	0.863	0.510	0.500	0.368
200	200	0.167	400	0.5	0.02	100	0.125	0.014	0.097	0.150	0.124	0.687	0.073	0.575	0.867	0.679	0.500	0.329
200	200	0.333	400	0.5	0.02	100	0.213	0.016	0.182	0.244	0.214	0.817	0.060	0.710	0.961	0.819	0.500	0.401
200	200	0.667	400	0.5	0.02	100	0.188	0.017	0.151	0.216	0.187	1.403	0.138	1.205	1.783	1.378	0.500	0.977
200	800	0.042	400	0.5	0.02	100	0.046	0.017	0.017	0.075	0.047	0.507	0.238	0.256	1.236	0.444	0.500	0.488
200	800	0.083	400	0.5	0.02	100	0.085	0.017	0.050	0.117	0.084	0.502	0.105	0.340	0.738	0.488	0.500	0.324
200	800	0.167	400	0.5	0.02	100	0.154	0.013	0.129	0.179	0.155	0.634	0.051	0.548	0.745	0.631	0.500	0.263
200	800	0.333	400	0.5	0.02	100	0.298	0.016	0.266	0.332	0.297	0.858	0.053	0.765	0.976	0.856	0.500	0.426
200	800	0.667	400	0.5	0.02	100	0.389	0.016	0.359	0.418	0.389	1.704	0.094	1.535	1.878	1.698	0.500	1.242
400	50	0.042	400	0.5	0.02	100	0.033	0.014	0.006	0.059	0.032	1.505	1.863	0.552	5.944	1.122	1.000	1.455
400	50	0.083	400	0.5	0.02	100	0.049	0.012	0.027	0.073	0.048	1.329	0.354	0.817	2.204	1.209	1.000	0.679
400	50	0.167	400	0.5	0.02	100	0.056	0.012	0.034	0.076	0.056	1.864	0.417	1.294	2.803	1.803	1.000	1.079
400	50	0.333	400	0.5	0.02	100	0.086	0.013	0.060	0.111	0.085	1.601	0.269	1.182	2.108	1.604	1.000	0.794
400	50	0.667	400	0.5	0.02	100	0.061	0.014	0.035	0.084	0.061	2.635	0.671	1.760	4.173	2.471	1.000	1.829
400	200	0.042	400	0.5	0.02	100	0.043	0.014	0.016	0.070	0.043	1.014	0.494	0.548	2.219	0.897	1.000	0.703
400	200	0.083	400	0.5	0.02	100	0.075	0.013	0.052	0.101	0.076	1.042	0.170	0.745	1.409	1.025	1.000	0.415
400	200	0.167	400	0.5	0.02	100	0.120	0.014	0.093	0.147	0.119	1.425	0.173	1.173	1.779	1.420	1.000	0.595
400	200	0.333	400	0.5	0.02	100	0.213	0.014	0.187	0.241	0.212	1.630	0.118	1.420	1.851	1.636	1.000	0.718
400	200	0.667	400	0.5	0.02	100	0.197	0.018	0.163	0.228	0.198	2.677	0.276	2.265	3.318	2.660	1.000	1.757
400	800	0.042	400	0.5	0.02	100	0.046	0.014	0.020	0.074	0.046	0.964	0.363	0.540	1.886	0.897	1.000	0.604
400	800	0.083	400	0.5	0.02	100	0.084	0.012	0.063	0.104	0.085	0.994	0.145	0.776	1.302	0.982	1.000	0.381
400	800	0.167	400	0.5	0.02	100	0.151	0.013	0.123	0.175	0.151	1.304	0.125	1.121	1.593	1.280	1.000	0.466
400	800	0.333	400	0.5	0.02	100	0.302	0.016	0.270	0.329	0.303	1.692	0.097	1.509	1.896	1.685	1.000	0.759
400	800	0.667	400	0.5	0.02	100	0.424	0.022	0.389	0.470	0.422	3.086	0.189	2.716	3.443	3.075	1.000	2.131
800	50	0.042	400	0.5	0.02	100	0.031	0.013	0.010	0.057	0.031	2.769	1.577	1.209	6.126	2.204	2.000	1.473

800	50	0.083	400	0.5	0.02	100	0.047	0.013	0.025	0.073	0.048	2.809	0.863	1.790	4.649	2.562	2.000	1.232
800	50	0.167	400	0.5	0.02	100	0.055	0.012	0.036	0.079	0.054	3.794	0.809	2.555	5.528	3.720	2.000	2.007
800	50	0.333	400	0.5	0.02	100	0.088	0.014	0.061	0.113	0.090	3.162	0.536	2.406	4.402	3.050	2.000	1.373
800	50	0.667	400	0.5	0.02	100	0.061	0.012	0.036	0.081	0.061	5.130	1.151	3.720	8.233	4.823	2.000	3.308
800	200	0.042	400	0.5	0.02	100	0.040	0.012	0.021	0.065	0.039	2.099	0.633	1.205	3.477	1.999	2.000	0.802
800	200	0.083	400	0.5	0.02	100	0.071	0.013	0.045	0.099	0.070	2.223	0.443	1.606	3.432	2.149	2.000	0.702
800	200	0.167	400	0.5	0.02	100	0.114	0.016	0.083	0.139	0.113	3.031	0.437	2.373	4.061	3.006	2.000	1.224
800	200	0.333	400	0.5	0.02	100	0.219	0.018	0.184	0.250	0.221	3.182	0.298	2.742	3.860	3.123	2.000	1.302
800	200	0.667	400	0.5	0.02	100	0.210	0.019	0.174	0.244	0.210	4.965	0.504	4.182	5.997	4.883	2.000	3.049
800	800	0.042	400	0.5	0.02	100	0.044	0.012	0.023	0.067	0.044	1.990	0.604	1.229	3.521	1.847	2.000	0.777
800	800	0.083	400	0.5	0.02	100	0.080	0.013	0.056	0.108	0.079	2.096	0.362	1.564	3.046	2.030	2.000	0.609
800	800	0.167	400	0.5	0.02	100	0.141	0.017	0.110	0.174	0.140	2.817	0.356	2.250	3.531	2.781	2.000	1.012
800	800	0.333	400	0.5	0.02	100	0.313	0.020	0.279	0.354	0.314	3.262	0.231	2.842	3.718	3.230	2.000	1.350
800	800	0.667	400	0.5	0.02	100	0.492	0.031	0.429	0.548	0.493	5.126	0.379	4.495	5.811	5.118	2.000	3.186
1600	50	0.042	400	0.5	0.02	100	0.030	0.011	0.006	0.047	0.031	6.688	8.538	2.804	23.665	4.451	4.000	3.970
1600	50	0.083	400	0.5	0.02	100	0.044	0.012	0.021	0.064	0.045	6.197	2.593	3.820	11.315	5.346	4.000	2.724
1600	50	0.167	400	0.5	0.02	100	0.053	0.013	0.029	0.079	0.055	8.121	2.520	5.156	13.603	7.382	4.000	4.416
1600	50	0.333	400	0.5	0.02	100	0.086	0.016	0.051	0.113	0.087	6.523	1.491	4.775	10.675	6.208	4.000	2.803
1600	50	0.667	400	0.5	0.02	100	0.060	0.015	0.033	0.087	0.060	10.826	3.364	6.916	17.980	10.081	4.000	7.068
1600	200	0.042	400	0.5	0.02	100	0.039	0.011	0.017	0.060	0.039	4.543	2.296	2.502	9.372	3.974	4.000	1.610
1600	200	0.083	400	0.5	0.02	100	0.069	0.015	0.038	0.096	0.069	4.686	1.282	3.129	8.423	4.494	4.000	1.324
1600	200	0.167	400	0.5	0.02	100	0.107	0.018	0.073	0.140	0.107	6.560	1.172	4.879	9.202	6.349	4.000	2.780
1600	200	0.333	400	0.5	0.02	100	0.233	0.027	0.180	0.285	0.232	6.010	0.731	4.758	7.637	5.955	4.000	2.184
1600	200	0.667	400	0.5	0.02	100	0.215	0.039	0.141	0.282	0.215	9.939	2.008	7.324	14.991	9.585	4.000	6.105
1600	800	0.042	400	0.5	0.02	100	0.043	0.012	0.020	0.064	0.042	4.172	1.482	2.475	8.108	3.760	4.000	1.230
1600	800	0.083	400	0.5	0.02	100	0.076	0.015	0.044	0.102	0.075	4.517	1.050	3.191	7.601	4.384	4.000	1.148
1600	800	0.167	400	0.5	0.02	100	0.130	0.019	0.097	0.167	0.129	6.175	0.960	4.656	8.125	6.049	4.000	2.386
1600	800	0.333	400	0.5	0.02	100	0.354	0.032	0.293	0.414	0.355	5.786	0.574	4.831	6.860	5.669	4.000	1.940
1600	800	0.667	400	0.5	0.02	100	0.517	0.135	0.229	0.747	0.537	10.524	4.015	6.462	21.039	9.125	4.000	6.825
3200	50	0.042	400	0.5	0.02	100	0.031	0.012	0.013	0.053	0.031	10.623	6.386	5.164	20.937	8.687	8.000	3.642

3200	50	0.083	400	0.5	0.02	100	0.044	0.014	0.017	0.071	0.044	12.597	5.338	7.143	28.366	11.165	8.000	5.145
3200	50	0.167	400	0.5	0.02	100	0.056	0.016	0.025	0.083	0.056	16.101	7.224	9.759	33.179	14.465	8.000	8.535
3200	50	0.333	400	0.5	0.02	100	0.090	0.018	0.063	0.124	0.089	12.387	2.619	8.636	17.433	12.126	8.000	4.676
3200	50	0.667	400	0.5	0.02	100	0.066	0.020	0.033	0.109	0.065	20.161	7.990	11.124	36.717	18.816	8.000	12.486
3200	200	0.042	400	0.5	0.02	100	0.040	0.013	0.018	0.064	0.038	8.948	6.179	4.774	16.964	8.024	8.000	2.660
3200	200	0.083	400	0.5	0.02	100	0.067	0.016	0.040	0.102	0.066	9.717	2.807	5.982	16.021	9.312	8.000	2.399
3200	200	0.167	400	0.5	0.02	100	0.117	0.022	0.079	0.157	0.117	11.990	2.540	8.781	17.222	11.570	8.000	4.297
3200	200	0.333	400	0.5	0.02	100	0.265	0.048	0.189	0.365	0.263	10.735	2.105	7.449	14.925	10.527	8.000	3.096
3200	200	0.667	400	0.5	0.02	100	0.224	0.108	0.062	0.431	0.227	25.239	17.336	9.222	65.650	18.479	8.000	17.735
3200	800	0.042	400	0.5	0.02	100	0.044	0.013	0.021	0.067	0.041	8.369	4.656	4.743	15.276	7.626	8.000	2.189
3200	800	0.083	400	0.5	0.02	100	0.075	0.017	0.041	0.107	0.076	9.362	2.581	6.161	16.386	8.653	8.000	2.106
3200	800	0.167	400	0.5	0.02	100	0.147	0.026	0.098	0.191	0.145	11.001	2.061	8.164	15.706	10.897	8.000	3.326
3200	800	0.333	400	0.5	0.02	100	0.403	0.093	0.207	0.597	0.411	10.795	4.053	6.774	20.077	9.686	8.000	3.445
3200	800	0.667	400	0.5	0.02	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Appendix S3

Results of simulations for 3 scenarios run on 100 datasets assuming an underlying Poisson distribution. The radius of movement within which signs are deposited *R*, the number of signs deposited by each animal group (N_{sign}) and the cell specific detection probability are kept constant. The simulation scenarios are run by varying the number of animal groups(S) and the number of grid cells(N_{cells}).

S	N_{sign}	*R*	No. of cells	r_{sign}	*d*	Data sets	Mean ($\hat{r}_{animal}$)	SD [Mean($\hat{r}_{animal}$)]	LCL ($\hat{r}_{animal}$)	UCL ($\hat{r}_{animal}$)	Median ($\hat{r}_{animal}$)	Mean ($\hat{\lambda}$)	SD [Mean($\hat{\lambda}$)]	LCL ($\hat{\lambda}$)	UCL($\hat{\lambda}$)	Median ($\hat{\lambda}$)	True λ	RMSE ($\hat{\lambda}$)
150	800	0.083	49	0.999	0.02	100	0.080	0.044	0.009	0.171	0.076	6.240	9.990	1.393	29.992	3.367	3.061	4.483
300	800	0.083	100	0.999	0.02	100	0.079	0.030	0.023	0.136	0.077	4.494	6.922	1.858	11.301	3.390	3.000	3.026
600	800	0.083	196	0.999	0.02	100	0.079	0.018	2.197	5.110	0.078	3.441	0.860	2.197	5.110	3.347	3.061	1.002

Chapter 4

Program SPACECAP: Software for estimating animal density using spatially-explicit capture-recapture models

4.1 Citation

Gopalaswamy, A. M., Royle, J. A., Hines, J. E., Singh, P., Jathanna, D., Kumar, N. S., Karanth, K. U. (2012). Program SPACECAP: software for estimating animal density using spatially explicit capture-recapture models. *Methods in Ecology and Evolution* 3(6): 1067-1072.

4.2 Abstract

1. The advent of spatially-explicit capture-recapture models is changing the way ecologists analyse capture-recapture data. However, the advantages offered by these new models are not fully exploited because they can be difficult to implement.

2. To address this need, we developed a user-friendly software package, created within the R programming environment, called SPACECAP. This package implements Bayesian spatially explicit hierarchical models to analyze spatial capture-recapture data.

3. Given that a large number of field biologists prefer software with graphical user-interfaces for analyzing their data, SPACECAP is particularly useful as a tool to increase the adoption of Bayesian spatially explicit capture-recapture methods in practice.

4.3 Introduction

Capture-recapture(CR) sampling (Otis *et al.*, 1978; Williams *et al.*, 2002) has emerged as a powerful tool for estimating animal abundance overcoming the key problem of imperfect detection. In such studies, 'capture histories' of individually identifiable animals (tagged or naturally marked) obtained from repeated sampling across several occasions, typically presented in the '1' (capture) or '0' (no capture) format, comprise the basic data for analysis. Since, demographic closure is a critical model assumption(Otis *et al.*, 1978; Williams *et al.*, 2002), the survey period is kept short. Data are often analysed using several closed CR models (Amstrup *et al.*, 2005) that incorporate heterogeneity in capture probabilities among individuals, trap response behaviour and time-varying capture probabilities. These analyses are performed using popular user-friendly software programs like CAPTURE (White *et al.*, 1978) and MARK (White & Burnham, 1999).

Such analyses incorporate statistical tests to select the best capture-recapture model for the data on hand, estimate capture probabilities and the population size is estimated based on the assumption that the population is 'closed'. While these traditional methods provide reliable estimates of abundance, density estimation requires that one estimates the sampled area.Typically, a buffer area is added around the trap array, based on some empirical, but ad hoc, approaches developed by Wilson & Anderson (1985) and Parmenter *et al.* (2003). Disadvantages of such approaches to sampled area estimation have been well-recognised statistically (Efford, 2004; Efford *et al.*, 2009) and empirically (Soisalo & Cavalcanti, 2006; Sharma *et al.*, 2010). Conventional CR analyses suffer from some other key disadvantages and deficiencies in estimating animal density. Foremost among them is the fact that although spatial information in the form of location of a capture is available in field studies, the full potential of such critical information related to the animal's home range and movement (and thus to abundance and density) is not meaningfully exploited in conventional CR analyses. Logistical problems such as inability to adequately cover the entire array with traps without leaving any 'holes' (Karanth & Nichols, 1998), the frequent need to sample the area in successive 'blocks' rather than in a single sweep, irregularly shaped trap arrays that are subsequently treated as circles while estimating variance of density or inclusion of habitats not used by the target species, may all combine and weaken inference under conventional CR analyses.

The recent development of spatially-explicit capture-recapture (SECR) models (Borchers & Efford, 2008; Royle *et al.*, 2009a) has addressed several of these problems faced by biologists employing CR studies for animal density estimation. Such models have been developed both under a maximum likelihood-based estimation framework (Efford, 2004; Borchers & Efford, 2008) as well as under a Bayesian framework (Royle & Young, 2008; Royle *et al.*, 2009b,a; Gardner *et al.*, 2009). Efford *et al.*

(2009) provide a general review and describe various types of animal 'captures' or 'detections' that can generate data for such analyses.

In this paper, we describe a free software program SPACECAP (Gopalaswamy *et al.*, 2012a), implementing a Bayesian SECR analysis of Royle *et al.* (2009a). Another software program called DENSITY (Efford *et al.*, 2004), is already available and, implements the maximum-likelihood based SECR analysis (Borchers & Efford, 2008). The 'SECR' and the 'SCRbayes' in R (R Development Core Team, 2009) implement more sophisticated maximum-likelihood based and Bayesian estimators, respectively. However, these R packages do not have Graphical User Interfaces (GUIs).

The reader is referred to the original paper by Royle *et al.* (2009a) for technical details of the models implemented in SPACECAP. Although the models implemented in SPACECAP may suit spatial CR data originating from different types of non-invasive sources of data, such as from camera traps, hair snares or faecal DNA samples, they do not apply to some situations such as kill trapping (Efford *et al.*, 2009).

Our primary motivation here is to enable practising biologists to reliably estimate animal densities using the flexible and powerful Bayesian SECR modeling approach of Royle *et al.* (2009a) without having to decipher the underlying complexity of the original R code (Royle *et al.*, 2009a, described below), thus meeting a key need of many species conservation initiatives involving rare and elusive animals that can only be studied using non-invasive survey methods (Hatler *et al.*, 2009).

4.4 Material and methods

4.4.1 Package development

SPACECAP was developed within programming environment R (R Development Core Team, 2009), a freeware, and is available within the library of R packages in the Comprehensive R Archive Network (CRAN; http://cran.r-project.org). We used the Tool Command Language (Tcl) with the Tk graphical user interface toolkit (http://www.tcl.tk/) to develop its GUI (figure 4.1). The software package includes a detailed Help file that is automatically downloaded with SPACECAP. The software can also be downloaded from the website of US Geological Survey, Patuxent Wildlife Research Center (PWRC) software archive (at http://www.mbr-pwrc.usgs.gov/software/spacecap.html). The initial version developed is the version 1.0.6. However, this has undergone subsequent changes with improved output displays and faster performance in version 1.1.0. The original code and all the relevant documentation, including a detailed instruction manual, is available at the CRAN website: http://cran.r-project.org/web/packages/SPACECAP/index.html.

4.4.2 Model description

SPACECAP implements a hierarchical model whereby a *state* and an *observation* process are clearly defined. Here, the animal population size and their respective locations constitute the state process. Accordingly, the location of an animal i is defined by

$s_i \sim \text{Uniform}\,(\mathcal{S})$ Here, $s_i, i = 1, 2, \ldots, N$, are the individual animal activity centres described distributed randomly over some region $\mathcal{S}$.

In well-designed CR experiments, investigators maintain a careful record of the individual animal i caught at a specific trap location j on a specific sampling occasion (e.g., a night, or a week) k. Thus, in SPACECAP, binary observations (i.e., 1s and 0s) y_{ijk} are recorded for an individual animal i caught at trap location j on sample occasion k.

The model is a type of binary regression model, similar to logistic regression, in which,

$y_{ijk} \sim \text{Bernoulli}(p_{ijk})$ Here p_{ijk} is the probability of detecting an individual i, at trap j on sample occasion k. The probability p_{ijk} may be related to covariates of interest by applying a suitable transformation. In SPACECAP, we make use of the complementary log-log link transformation. Thus, the simplest possible model is expressed as

$cloglog(p_{ijk}) = b_0$

where, b_0 is a parameter to be estimated. By taking the inverse of the cloglog transformation:

$p_{ijk} = 1 - e^{-e^{b_0}}$

This particular function arises by considering the binary observations to be formally reduced from a Poisson encounter frequency model (see Royle *et al.*, 2009a, for details). That is, p_{ijk} is the Pr(*that animal i makes at least one encounter with trap j on occasion k*) under the Poisson model. The most general model that SPACECAP presently allows is a model in which

$cloglog(p_{ijk}) = \beta_0(1 - x_{ijk}) + \beta_1 x_{ijk} + \beta_2 f[dist(s_i, u_j)]$

In this expression x_{ijk} is an indicator of previous capture of individual i in trap j and f is the function that describes the mechanism of decline of detection probability as the distance between trap j and location of individual i increases, i.e., a function to make p_{ijk} have a half-normal or negative exponential form in SPACECAP. Thus, β_0 and β_1 are measures of capture probabilities (refer to Royle *et al.*, 2009a, for details) until and after the first capture, respectively. The parameter β_2 is a regression coefficient on the effect of distance between individual activity centre s_i and the location of trap j, u_j. β_2 is constrained to be < 0 implying that the probability of encountering an individual in a trap decreases as the distance between the individual and the trap increases. However, there may be situations where this is not true, for

example, when animals spend a lot of time patrolling territory borders (Borchers & Efford, 2008).

4.4.3 Models implemented

SPACECAP draws data for analysis in the form of spreadsheet files (such as from Microsoft Excel or OpenOffice) created by the user. The user also defines the specific type of spatial capture-recapture model to be run. Currently, SPACECAP can analyse the Bernoulli encounter process model with the following options (see Royle *et al.*, 2009a, for a discussion):

1. Spatially explicit OR Non-spatially explicit model
2. Behavioural OR Non-Behavioural response model
3. Half-normal OR Negative exponential detection function model

The detection function model specifies the mechanism by which the regression coefficient β_2, in the complementary log-log link function, decreases as distance between an animal and a trap increases. Further detail on how the detection function is linked to the observation process is described elsewhere (Royle *et al.*, 2009a; Borchers & Efford, 2008).

4.4.4 Parameters reported

In SPACECAP, the parameter $N(\mathcal{S})$ or N_{super} is the population size of individuals – the estimated number of activity centres located in $\mathcal{S}$ and density is calculated as $D = \frac{N_{super}}{\|\mathcal{S}\|}$, where $\|\mathcal{S}\|$ is the area of the state-space. Note that $N(\mathcal{S})$ is the population size for the prescribed state-space (see Royle *et al.*, 2009a, for a description of the Bayesian estimation procedure) and should not be interpreted as the population size in some sampled area. λ_0 is the intercept, which is the expected encounter rate of an individual i in trap location j at sampling occasion k, whose activity centre is exactly at the trap location. For a non-behavioural response model, whereby $\beta_0 = \beta_1$, $\lambda_0 = e^{\beta_0}$.

The rate of decline in detection probability as the distance between the animal centre and the trap increases is indicated by the parameter β_2 in the complementary log-log link function. SPACECAP reports the parameter $\sigma = \frac{1}{\beta_2}$. This may also be viewed as a "range parameter" of an animal if the animal movement about its activity centre has a distribution similar to the detection function used. Thus, for a highly mobile animal, this value will tend to be large (e.g.:- tigers will have a higher

σ value compared to civets). We note here that the σ estimated in SPACECAP is scaled according to the standard detection function implemented in Borchers & Efford (2008) and not according to the alternative scaling used in Royle *et al.* (2009a).

When the "behavioural response" option (see below) is chosen, SPACECAP estimates two more derived parameters:
$p_1 = 1 - e^{-\beta_0}$, is the encounter probability of an individual before initial encounter.
$p_2 = 1 - e^{-\beta_1}$, is the encounter probability of an individual after initial encounter.

All parameters are estimated using the Markov chain Monte Carlo (MCMC) estimation process using the Metropolis algorithm (Gelman *et al.*, 1996). The current version of SPACECAP only runs one MCMC chain and multiple chains can be analysed using by running SPACECAP sequentially. The user is provided with the option of setting the basic MCMC parameters before analysis. These include the number of iterations to be run, the number of iterations to be discarded (called the "burn-in period"), thinning rate of the MCMC samples and the data augmentation value (the maximum allowable number of animals possible within the state-space $\mathcal{S}$). Precautions must be taken when setting the data augmentation value. If a very low value is set, then there may be a possibility of underestimating the population size by truncating the posterior distribution of the population size estimate. Similarly, if a very high value is set, it may result in high computational time. SPACECAP reports the parameter ψ, which is the fraction of the maximum allowable number (set by the user during data augmentation), that actually represents the true population. We recommend users to set the data augmentation value at about 5-10 times the expected population size in the larger state-space.

4.4.5 Assessment of model adequacy

SPACECAP implements an assessment of model adequacy by computing the Bayesian P-value based on individual encounter frequencies (see Royle *et al.*, 2011b, for details on the implementation). In this approach, a fit-statistic,D, is constructed based on individual encounter frequencies, n_i, by using a Freeman-Tukey type statistic (Freeman & Tukey, 1950). .

$$D(data, \theta) = \sum_{i=1}^{N} (\sqrt{n_i} - \sqrt{e_i})^2$$

where we define n_i to be the encounter frequency for individual i conditional on s_i and e_i is the expected value under the model and θ are the model parameters. A Bayesian P-value estimate close to 0 or 1 would imply model *inadequacy*. We note that DENSITY utilizes the Akaike's Information Criteria as a means to facilitate model selection.

4.4.6 Summary statistics and output

The approach for parameter estimation in SPACECAP is Bayesian. Therefore, the package produces a summary of parameter estimates comprising of the posterior mean, posterior standard deviation, and the 95% highest posterior density intervals (for examples, see table 4.1). Thus, for a basic user, SPACECAP provides only all the details relating to the parameter estimates. In addition, SPACECAP also outputs files to allow the user to generate a pixellated map showing fine-scale variation in density (figure 4.2).

All the prior distributions of parameters defined in SPACECAP are improper uniform priors. So, for all positive parameters the prior distribution would be Uniform $(0, \infty)$ and for all other parameters the prior distribution would be Uniform $(-\infty, \infty)$.

Additionally, SPACECAP evaluates the Geweke's diagnostic (Geweke, 1992) on the MCMC analysis and provides the user with the z-score of the diagnostic. Values of the z-score statistic of $|z - score| > 1.6$ implies lack of convergence. Future versions of SPACECAP will permit users to define their own priors and incorporate more model variants.

4.5 Results

We report results of a re-analyses of a tiger camera trap data set, originally used in Royle *et al.* (2009a), using SPACECAP (see table 4.1). In table 4.2, we compare density estimates obtained in CR models incorporating distance-related heterogeneity in capture probabilities with software packages SPACECAP and DENSITY(Efford, 2004), with conventional CR approaches implementing the general heterogeneity CR model, M_h-jackknife (Otis *et al.*, 1978), in conjunction with ½ MMDM approach (Wilson & Anderson, 1985) for abundance and area estimation, respectively.

These comparisons show that:

1. Densities estimated from both the SECR approaches (Bayesian and maximum likelihood) were less than those obtained from conventional CR approaches.

2. Both Bayesian and maximum likelihood-based inferences from packages SPACECAP and DENSITY respectively produced similar estimates of animal density.

The overall conclusion is that as the sample size increases (a combination of number of individuals caught, and their recaptures over multiple locations), there does not appear to be any major difference in density estimates derived using programs SPACECAP or DENSITY. Conventional CR model M_h-jackknife estimates derived from program CAPTURE appear to over-estimate tiger densities if the sampled area

is defined using the ad hoc Wilson & Anderson (1985) approach using ½ MMDM as buffer. It results in underestimates of densities if the buffer is defined using the Parmenter *et al.* (2003) full MMDM approach.

4.6 Discussion

In the field of ecological investigation, the use of Bayesian inference methods (Gelman *et al.*, 1996; Bolker, 2008; Royle & Young, 2008) are becoming popular due to their inherent capacity to solve complex problems with powerful MCMC algorithms (Tierney, 1994), computational power (Venables & Ripley, 2002) and flexibility to tailor the modelling to account for particular ecological and sampling situations (e.g., Gardner *et al.*, 2010a; Sollmann *et al.*, 2011). We visualise the present and future advantages of the SECR models presented by Royle *et al.* (2009a) implemented in SPACECAP within this context. Although a typical analysis using SPACECAP now takes much longer than a comparable analysis using DENSITY (Efford, 2004), we do not view this as a serious practical disadvantage because such analyses are typically performed only once in a season or a year following, 45–65 days of field work and several weeks of organising and checking the data prior to analysis.

We envisage that the modelling framework of Royle *et al.* (2009a) used in SPACECAP for analysing animal detection data, especially from camera trap surveys, provides the opportunity to incorporate auxiliary information in the form of 'capture' data on the same individuals in the form of photographs taken by tourists, animals killed in conflicts or by poachers and even live tracking of animals tagged by radio-frequency identification tags, radio-collars or satellite-collars to improve accuracy and precision of the abundance estimates. The explicit recognition of the fact that most animals have biologically defined home ranges in the statistical modelling allows for a more realistic characterisation of the spatial ecology of the constituent individuals within a population. This realistic biological treatment, in turn, is likely to provide more reliable estimates of animal abundance and density. Further, SPACECAP also permits for inference about the locations of uncaught animals of a study, which is not possible with DENSITY, and therefore can be easily extended to modeling demographically open populations. Currently such open model analysis are available with program MARK (White & Burnham, 1999) using modelling approaches developed by Jolly-Seber (Jolly, 1965; Seber, 1965), Lebreton *et al.* (1992), Pollock *et al.* (1990) and Kendall & Pollock (2001) to derive estimates of critical vital rates such as apparent survival, temporary immigration, transience, recruitment, in addition to changes in population size and density (see Karanth *et al.*, 2006, for an example of such an analyses with tiger photo-capture data collected over a decade). Recent work by Gardner *et al.* (2010a) has demonstrated

the use of SECR approaches to analyse open population models. We hope to incorporate several such developments in SPACECAP during future releases, keeping in view our fundamental goal of making these modern statistical tools available to practitioners. For users, there is a SPACECAP user group email forum where frequently asked questions are answered. Emails to this group can be sent to spacecapforum@googlegroups.com.

4.7 Tables

Table 4.1: Posterior density estimates of parameters estimated on tiger data set (Royle *et al.*, 2009a). See *Methods and Materials - Parameters reported* for definitions of parameters.

Parameter	Posterior Mean	Posterior sd	95% Lower HPD level	95% Upper HPD level
σ	1.975	0.2166	1.5801	2.412
λ_0	0.016	0.0034	0.0096	0.023
ψ	0.4307	0.0822	0.2799	0.5936
N_{super}	210.60	38.915	137	284
Density D	13.03	2.4079	8.5388	117.6344

Table 4.2: Comparison of parameter estimates (posterior means and standard deviations) in SPACECAP, DENSITY and M_h-jackknife. See *Methods and Materials - Parameters reported* for definitions of parameters. Note: the parameter g_0 is estimated in DENSITY and is analogous to λ_0

Parameter	SPACECAP	DENSITY	M_h-jackknife
σ	1.975(0.217)	1.977 (0.226)	NA
λ_0/g_0	0.016 (0.00)	0.015 (0.00)	NA
ψ	0.43 (0.08)	NA	NA
N_{super}	210.60 (38.92)	NA	NA
N_{hat}	NA	99.69 (19.19)	100 (18.15)
$Density(/100km^2)$	13.03 (2.41)	13.23 (2.62)	20.3(3.80)

4.8 Figures

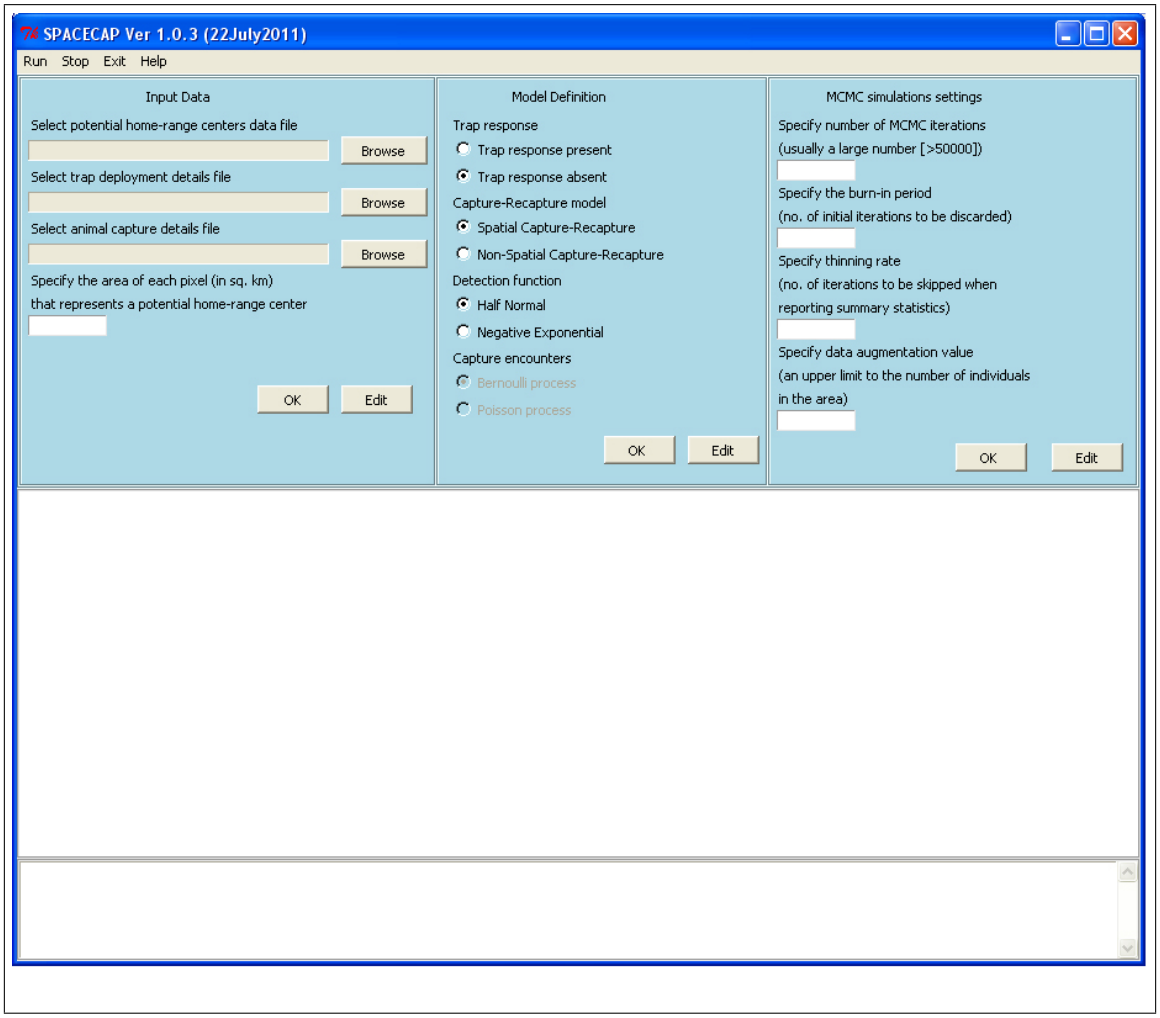


Figure 4.1: The Graphic User Interface(GUI) of SPACECAP

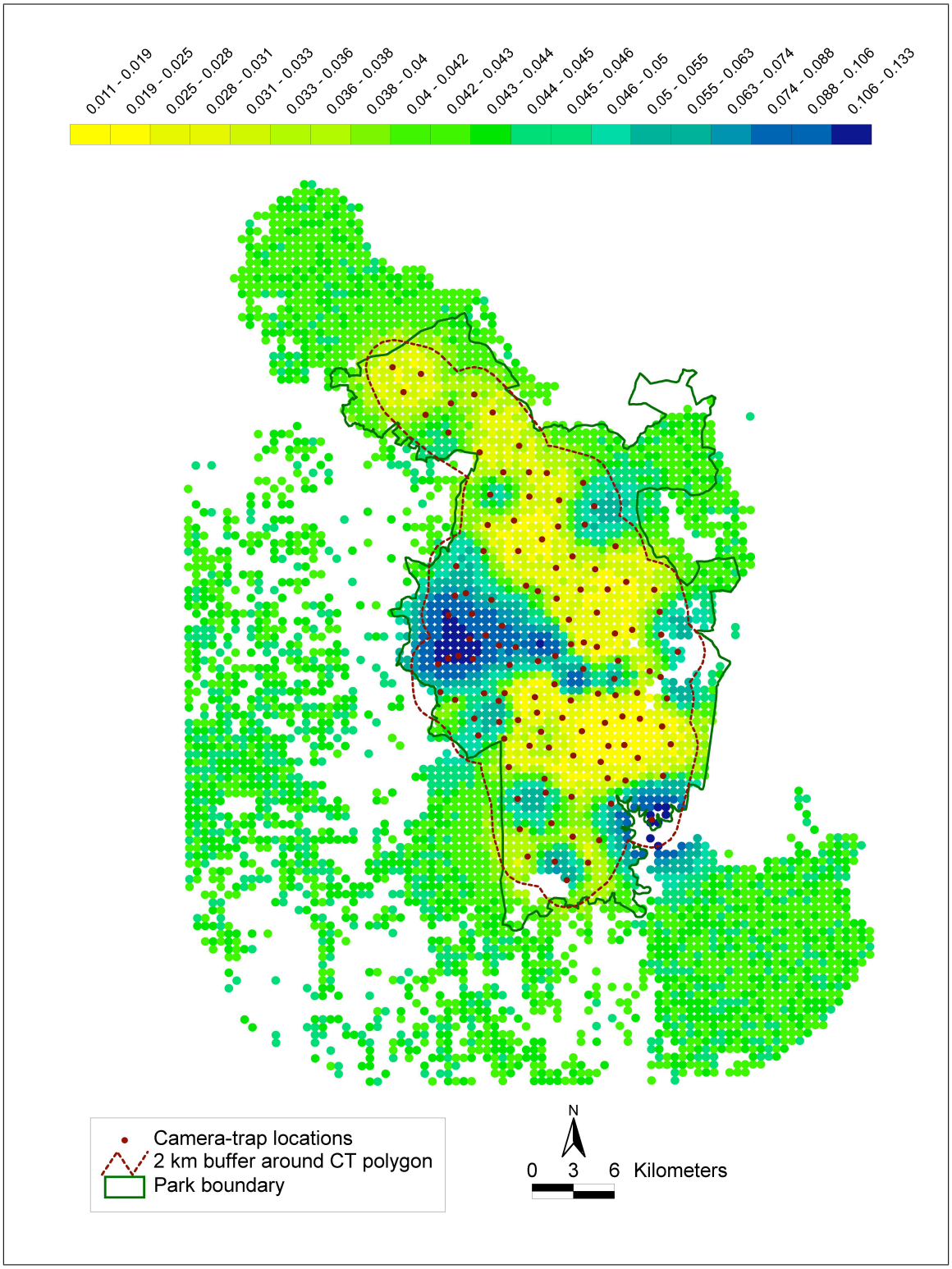


Figure 4.2: A pixellated density map produced in SPACECAP showing relative animal densities. This map reports estimated tiger densities per pixel of size 0.336 km^2. White regions represent unsuitable habitat.

Chapter 5

Density estimation in tiger populations: combining information for strong inference on tiger density

5.1 Citation

Gopalaswamy, A. M., Royle, J. A., Delampady, M., Nichols, J. D., Karanth, K. U., Macdonald, D. W. (2012). Density estimation in tiger populations: combining information for strong inference on tiger density. *Ecology* 93(7) 1741-1751.

5.2 Abstract

A productive way forward in studies of animal populations, is to efficiently make use of all the information available, either as raw data or as published sources, on critical parameters of interest. In this study, we demonstrate two approaches to use of multiple sources of information on a parameter of fundamental interest to ecologists: animal density. The first approach produces estimates simultaneously from two different sources of data. The second approach was developed for situations in which initial data collection and analysis are followed up by subsequent data collection, and updates prior knowledge with new data using a step-wise process. Both approaches are used to estimate density of a rare and elusive predator, the tiger, by combining photographic and fecal DNA spatial capture-recapture data. The model, which combined information, provided the most precise estimate of density [8.5(1.95) tigers/100 km^2], relative to a model that utilized only one data source [photographic: 12.02(3.02) and fecal DNA: 6.65(2.37) tigers/100 km^2]. Our study demonstrates that by accounting for multiple sources of available information, estimates of animal density can be significantly improved.

5.3 Introduction

Animal abundance or density is a state variable fundamental to answering important questions in ecology and conservation biology (May, 1999). But estimating this parameter accurately, especially for rare, elusive and wide-ranging species, such as large carnivores, has proven to be particularly challenging (Thompson, 2004). In most landscapes, such species occur at densities that can be too low to permit collection of sufficient data for accurate density estimation; therefore, there is a premium on developing approaches that make full use of data from multiple sources and using an inferential framework that permits the use of ecological constraints and prior knowledge in density estimation.

For instance, a recent study (Walston *et al.*, 2010) identified 42 global tiger *Panthera tigris* source populations which occupied areas adding up to about 88,000 km^2, each requiring rigorous annual monitoring. However, even in these sites, tigers survive at relatively low densities of 1–10 tigers per 100 km^2 or lower. Current robust tiger estimation methodologies, such as photographic capture-recapture (Karanth & Nichols, 1998) or fecal DNA capture-recapture (Mondol *et al.*, 2009), at such low tiger densities, may not yield precise estimates when used alone. It is important, therefore, to develop a sampling and analytical methodology that can effectively combine information from multiple survey methods to estimate, more reliably, tiger abundance.

In this study, we develop such an approach to estimate abundance and density of a tiger *Panthera tigris* population in the Western Ghats, India, utilizing a recently-developed spatially explicit capture-recapture model (Royle *et al.*, 2009a, 2011a) to combine information about from two pre-designed but distinct sources of tiger capture-recapture survey data: (1) photographic captures of tigers from camera-traps and (2) fecal samples of tigers containing DNA collected during the same season at a single site. We believe that because of its rarity, elusiveness, role as an important apex predator (Estes *et al.*, 2011) and global conservation priority, the tiger serves as an ideal model to demonstrate the general utility of our approach.

Advances in spatially explicit capture-recapture (SECR) modeling have facilitated improved estimation of animal density by explicitly modelling a major source of heterogeneity in capture probabilities among individuals in a population, viz. the distance of an animal from the detection device (Borchers & Efford, 2008; Royle & Young, 2008; Royle *et al.*, 2009a). Bayesian SECR models (Royle *et al.*, 2009a) have been extended to estimating critical vital rates in open populations (Gardner *et al.*, 2010a). Further, SECR approaches appropriately describe the reality that animal detection locations typically occur within individual home ranges, a crucial biological fact that was essentially ignored in earlier conventional capture-recapture modeling approaches (Otis *et al.*, 1978; Amstrup *et al.*, 2005). We emphasize that

this biological knowledge is particularly relevant to carnivore populations, and especially felids, which are strongly territorial (Sunquist & Sunquist, 2002; Macdonald & Loveridge, 2010). Consequently, SECR approaches have been successfully applied to accurately determine large carnivore density, for example, with tigers using camera trap data (Royle *et al.*, 2009a,b), with bears using DNA samples from hair snare data (Gardner *et al.*, 2010b)and with various other species using unstructured survey data based on samples of DNA(Kéry *et al.*, 2010).

A potential benefit of Bayesian inference to ecologists (Ellison, 2004) is the ability to combine information on important ecological parameters from multiple sources of data to strengthen conclusions. This benefit has recently found application in the development of integrated population models in ecology (see Schaub *et al.*, 2007; Abadi *et al.*, 2010a,b), especially when data sets are sparse and limited, to estimate population dynamic parameters. Our primary objective in this study was to extend the use of this idea to spatially-explicit capture-recapture approaches in order to improve the estimation of tiger density by incorporating information from two distinct but rigorous surveys, viz., a photographic survey via camera traps and a scat survey for fecal DNA samples. Our secondary objective was to assess competing models of the combined data, each producing independent estimates of tiger density, and to identify those ecological and sampling processes that produce the most reliable estimate of density.

5.4 Methods

5.4.1 Field surveys

5.4.1.1 Photographic surveys

Our study was carried out in the 871 km^2 Bandipur Tiger Reserve ($76\,°120 - 76\,°460$ E and $11\,°370 - 11\,°570$ N) in Karnataka state, India. This rugged landscape is drained by several rivers and varies in altitude ($680 - 1454$ m), annual rainfall ($625 - 1250$ mm) and mean monthly temperature ($18\,°C - 29\,°C$). Its deciduous forests support high densities of ungulate prey species (approximately 35.2 animals/km^2), and consequently, a large tiger population (Karanth *et al.*, 2004).

The study area contained 118 camera trap locations set along well-used dirt roads and trails frequently used by tigers (figure 5.1). The study included 1246 camera trap nights of effort across 12 sampling occasions, over a 39-day period in 3 trapping blocks according to the design recommendations in Karanth & Nichols (1998). The survey resulted in 35 photographic 'captures' of 29 distinct individuals (Mondol *et al.*, 2009), from the same tiger population that we subsequently sampled using genetic captures. Our camera-trap survey was conducted from November 2005 to January 2006. Date, time and location of captures of each individual tiger were

the additional data used in our analysis with spatially-explicit capture recapture models.

5.4.1.2 Fecal DNA surveys

The genetic survey contained 18 'search routes' (figure 5.2) spatially distributed along a representative set of camera trap trails, to ensure that scats of each individual tiger within the study area had a nonzero probability of being detected. Each route was sampled on foot by a team of two trained trackers at seven-day intervals. Thus, the entire study area was surveyed over six successive days (followed by a day of rest) by three teams, which covered a distance of 237 km, collecting 'fresh' tiger scats judged to be less than 7-days-old based on physical appearance and moisture content (Mondol *et al.*, 2009; Andheria, 2006). The survey was repeated over six consecutive weeks in the same order to ensure that each 'search route' was visited regularly once a week. The survey yielded a total of 63 samples, of 24 individually identified tigers, collected over 42 days. Based on previous studies of tigers, the duration was adequately short for us to assume geographic and demographic closure (Karanth & Nichols, 1998). In our approach, it is not necessary for the population to be closed over the entire season (i.e., during photographic and genetic sampling together). We assume, however, that the population size and density is the same during each survey. We considered this to be a reasonable assumption considering that the fecal sampling was carried out only 6 weeks after the photographic survey.

5.4.2 Analytical methods

We demonstrate two approaches to use of multiple sources of information for estimating animal density. We recommend the first approach for situations in which both sets of raw data are available for analysis. The second two-step approach may be required when data collection efforts are sequential, and when density estimates, but not the raw data, are available from an initial survey.

5.4.2.1 The direct approach

This approach is developed for the situation in which raw data from two independent surveys are available. This will be the case, for example, when the two data collection efforts occur simultaneously, frequently by the same investigator. Suppose D_P and D_G are data from the two different surveys (photographic and genetic surveys), which are expected to be informative about 3 different sets of parameters, θ_P, θ_b and θ_G, where
θ_P and θ_G are parameters which are *exclusively* informed by data sets D_P and D_G, respectively and θ_b is the set of parameters informed by *both* data sets. Note that these parameters, θ_b for example, can vary among models considered in a single

analysis, reflecting different hypotheses about the degree of overlap between the processes generating the 2 data sets.

Then the joint model $f(D_P, D_G|\theta_P, \theta_b, \theta_G)$ is expressed as

$$f(D_P, D_G|\theta_P, \theta_b, \theta_G) = f_P(D_P|\theta_P, \theta_b) f_G(D_G|D_P, \theta_P, \theta_b, \theta_G).$$

When data are directly available from the first phase (in our case, D_P, the photographic survey), the joint model can be directly analyzed by maximizing the likelihood or by Bayesian analysis using Markov Chain Monte Carlo (MCMC) approaches.

5.4.2.2 The step-wise approach

However, it is common for ecologists to face a situation where concurrent data are not readily available but they wish to utilize information on parameters from other published sources. Information may be only available in the form of estimates of parameters with their associated standard errors. In such cases, we propose the application of the 'step-wise updating process' (Bernardo & Smith, 1994). Suppose, the prior model for θ_P, θ_b and θ_G is given by

$$\pi(\theta_P, \theta_b, \theta_G) = \pi_{Pb}(\theta_P, \theta_b)\pi_G(\theta_G).$$

We can now express the joint posterior density of θ_P, θ_b and θ_G as

$$\begin{aligned}
\pi(\theta_P, \theta_b, \theta_G|D_P, D_G) &= \frac{f_P(D_P|\theta_P, \theta_b)\pi_{Pb}(\theta_P, \theta_b) f_G(D_G|D_P, \theta_P, \theta_b, \theta_G)\pi_G(\theta_G)}{m(D_P, D_G)} \\
&= \frac{f_P(D_P|\theta_P, \theta_b)\pi_{Pb}(\theta_P, \theta_b)}{m_1(D_P)} \frac{m_1(D_P)}{m(D_P, D_G)} \times \\
&\qquad f_G(D_G|D_P, \theta_P, \theta_b, \theta_G)\pi_G(\theta_G) \\
&\propto \pi(\theta_P, \theta_b|D_P)\pi_G(\theta_G) f_G(D_G|D_P, \theta_P, \theta_b, \theta_G)
\end{aligned}$$

where m_1 and m above are the respective predictive densities.

From the above development, we see that the product $\pi(\theta_P, \theta_b|D_P)\pi_G(\theta_G)$ represents the state of our knowledge about all the model parameters at the end of the first phase (i.e., the photographic survey). And when confronted with data D_G from the second phase (i.e., the genetic survey) we obtain our accumulated knowledge of model parameters from both the phases. We recognize that the extent of updating that occurs on a set of model parameters at each phase will depend entirely upon the model being used. So a good model will yield precise posterior distributions on the set of model parameters.

Gain in precision from multiple sources of data. We may formally ask how the precision about the set of parameters θ_b will improve at each phase in the stepwise process (Bernardo & Smith, 1994). Let us consider the distribution of a single parameter θ, which is expected to be informed by both, the photographic and the genetic phases.

Suppose the posterior distribution of θ after being informed by the first phase (the photographic phase) can be approximated as a normal distribution,

$$\pi(\theta|D_P) \sim \text{Normal}\left[\mu(D_P), \tau^2(D_P)\right]$$

where, $\mu(D_P)$ is the posterior mean estimate of θ and $\tau^2(D_P)$ is the posterior variance of θ after data are observed from the first phase. Assuming that data from the second phase, D_G, can be summarized in terms of a statistic $\nu(D_G)$ (which is statistically sufficient for θ with regard to information content, such as the sample mean), the sampling model for D_G can be considered to be normal,

$$\nu(D_G)|\theta \sim \text{Normal}(\theta, \delta_G^2)$$

where δ_G^2 is the sampling variance of $\nu(D_G)$, so that when the sample size of D_G becomes large, δ_G^2 becomes small.

So the final posterior distribution for θ from both the phases is also given by a normal distribution

$$\pi\left[\theta|D_P, D_G\right] \sim \text{Normal}\left[\frac{\tau^2(D_P)}{\tau^2(D_P)+\delta_G^2}\nu(D_G) + \frac{\delta_G^2}{\tau^2(D_P)+\delta_G^2}\mu(D_P), \frac{\tau^2(D_P)\delta_G^2}{\tau^2(D_P)+\delta_G^2}\right]$$

The above expression provides the posterior density for the parameter θ, and represents a weighted average of the contributions from each phase, weighted by the inverse of the variance of the estimate, separately from each phase. We note that the final posterior variance, which denotes the final uncertainty satisfies

$$\frac{\tau^2(D_P)\delta_G^2}{\tau^2(D_P)+\delta_G^2} = \tau^2(D_P)\left(1 - \frac{\tau^2(D_P)}{\tau^2(D_P)+\delta_G^2}\right) = \tau^2(D_P) - \frac{\tau^4(D_P)}{\tau^2(D_P)+\delta_G^2} < \tau^2(D_P)$$

This equation demonstrates that the relative gain in precision (or the reduction in variance) brought about by the data from the second phase (i.e. the genetic phase) is the proportion

$$\frac{\tau^4(D_P)/(\tau^2(D_P)+\delta_G^2)}{\tau^2(D_P)} = \frac{\tau^2(D_P)}{\tau^2(D_P)+\delta_G^2}$$

5.4.2.3 The spatially explicit capture-recapture model

In this study, we make full use of the spatial locations of tiger captures from the photographic and fecal DNA surveys –information that was ignored in Mondol *et al.* (2009) – in our modeling and density estimation. We describe here the modeling

structure and define the parameters being estimated. Our approach utilizes one of the versions of the Bayesian spatially-explicit capture-recapture models described in Royle *et al.* (2009a).

State process. The distribution of tigers over the region of interest is first described by the state process and modeled as follows:

$$\mathbf{s}_i \sim \text{ Uniform}(\mathcal{S})$$

where, $\mathbf{s}_i$ represents the coordinates of the activity center of individual tiger i on a Cartesian plane composed of its respective x and y-coordinates $(\mathbf{s}_{1i}, \mathbf{s}_{2i})$. This model regards the realized activity centers as the outcome of a point process, where $\mathcal{S}$ is the state-space of the point process.

In our approach, the photographic and the fecal DNA surveys represent the two observational processes. Through each of these observation processes, only a fraction of the actual number of tigers available in the population, is caught. However, these fractions may vary between the two survey types. But our central problem is that population size itself is unknown. To account for this situation we use the approach of data augmentation (Royle *et al.*, 2007), whereby, a large number of 'all-zero' encounter histories are added to 'augment' to the observed data set of detected individuals. These augmented encounter histories represent a mixture of: (1) members of the population that were simply not captured and (2) fixed zeros that are not members of the population.

If the size of the augmented data set is M, then the population size of tigers N is binomial:

$$[N|M, \psi] \sim \text{Binomial}[M, \psi]$$

where, ψ is the probability that a member of M (the total augmented population) is an actual tiger (a member of N). With a uniform(0, 1) prior on ψ, the implied marginal (unconditional on ψ) prior distribution for N is uniform on the integers 0 to M (Royle *et al.*, 2007).

In our situation, M is a fixed value, leaving only the parameter ψ to be estimated. We note that the value chosen for M can have important implications for the estimation of ψ. In optimum circumstances, we would wish to set M so that the estimate of ψ is in the range of 0.2-0.8, as recommended in site occupancy studies (MacKenzie *et al.*, 2002). At extreme values of ψ, estimation of N becomes more difficult.

Observation process. From the development in Royle *et al.* (2009a) one version of the observation process takes the form

$$[y_{ij}|\lambda_0, \sigma, \mathbf{s_i}] \sim Poisson[\lambda_0 e^{-(d_{ij}^2/\sigma)}]$$

for the Gaussian form of the detection function or

$$[y_{ij}|\lambda_0, \sigma, \mathbf{s_i}] \sim Poisson[\lambda_0 e^{-(d_{ij}/\sigma)}]$$

for the negative exponential form. Here,
y_{ij} is the observed number of encounters of individual tiger i in trapping device j during the sampling period,
λ_0 is the basal encounter rate in the observation process; or the expected encounter rate of a tiger whose activity center is located at the trapping device j,
d_{ij} is the distance between the location of the activity center of tiger i, $\mathbf{s}_i$ and the trapping device j,
σ represents the rate of decline in detection probability with increasing distance between the animal's activity center and the trap. This parameter maybe related to the tiger movement in our sampling situation.

The two survey types used represent two observation processes from a common state process (see Royle & Dorazio, 2008). However, from our sampling process, it is not possible to associate the photograph of an individual tiger with its corresponding fecal DNA sample. Furthermore, there was a time gap of about 6 weeks between the two surveys. Hence, we expect that even while the densities may remain unchanged between the two survey types, there may be shifts in the activity centers of individual tigers (Karanth *et al.*, 2006).

The performance of this Bayesian spatially explicit capture-recapture model with relatively low sample sizes, such as ours, has already been discussed elsewhere (see Royle & Young, 2008; Royle *et al.*, 2009a, 2011b, for details) based on some simulation exercises. Low sample sizes tend to produce skewed posterior distributions, and the posterior mean tends to be biased. However, as is logically implicit in our development of the direct and step-wise approaches, since we are essentially increasing the size of our data set, the estimates will be more precise than when each data set is analyzed separately. In our implementation for tiger survey data we will, in fact, assess for model adequacy under various cases.

5.4.3 Implementation for tiger survey data in Bandipur

The area under consideration for analysis was defined by framing a rectangular polygon along the northernmost, westernmost, southernmost and easternmost of the camera trapping devices and a 15 km buffer set around this frame. We considered this large spatial extent to represent an area that was large enough to ensure that

no tiger outside this frame, based on previous camera trapping studies of tigers (Karanth *et al.*, 2004; Karanth & Nichols, 1998), had even the slightest probability of being exposed to the trapping area during sampling.

The set of 18 routes in the fecal DNA survey that involved 237 kilometers of walk effort (per sweep) was divided into 1 km segments (figure 5.2). Every 1 kilometer segment was considered to be a 'trapping device' that accumulates DNA evidence and is represented by the mid-point of each segment on the xy-plane. The choice of a 1 km segment length was based on accounting for a trade-off between the specificity of information on the spatial positions of tigers against the likelihood of violating of the assumption of independence of fecal DNA samples collected along trails. A similar approach of approximation was employed in Royle *et al.* (2011b) while assessing densities of a bird population. We retained the same state-space for the camera trapping analysis (figure 5.1).

Our photographic and fecal DNA sampling for tigers yielded 29 and 24 individual tigers captured, respectively. We refer to these values as n_P and n_G, respectively. The parameter N, which is common to both the survey processes, must lie in between n_P and the data augmented value M (because $n_P > n_G$). Since N is the parameter we are estimating, we must ensure that the posterior distribution of N is not truncated by setting a conservative value of M. For our data set, we set a value for $M = 800$. The density D is estimated by
$D = \frac{N}{A}$. Here, A is the area of the state-space (a rectangular polygon defined by the trapping devices + 15 km buffer set around this frame).

5.4.3.1 A priori models

We implement our Bayesian model using WinBUGS (Lunn *et al.*, 2000) via the package R2WinBUGS available in R (R Development Core Team, 2009). We confront three a priori models (table 5.1) in our analysis:

Model 1 In this model we assume that the population size, described by the parameter, ψ, remains constant between the two surveys. Our hypothesis is that 6 weeks (the time between the two surveys) is too short a time period for ecological processes, such as demographic stochasticity (Roughgarden, 1975), to produce substantive differences in density. We also assumed that the detection function parameter, σ, remains the same between the two surveys. Since this parameter is related to animal movement, we did not expect for any biological reason, that the parameter σ should be different between the two surveys. We did, however, assume that the basal-encounter rate parameter, λ_0, will be survey-specific, because factors governing detection rates via camera traps and via sampling 1 km trail segments for fecal DNA samples may be completely different and not really comparable. Thus,

this model is a 4-parameter model (table 5.1) and contains 2 additional derived parameters (population size N and density D). BUGS code is provided in Appendix 5.8.1.

Model 2 In this alternative to Model 1, we allowed for the detection function parameter, σ, to be survey-specific. Tigers are known to increase their marking near the boundaries of their territories (Smith *et al.*, 1989). Although, Smith *et al.* (1989) observe that feces and anal gland deposits form only 17% of marking, they may play a vital role in increasing the value of the parameter, σ, in the fecal DNA survey. We would expect that the parameter σ, in such circumstances would be survey-specific and is expected to be larger in fecal DNA surveys. Further, there may also be an effect of reducing 1 km segments to point locations during analysis, causing a difference in σ between the two surveys. Assumptions about all other parameters are as described for Model 1. This is a 5-parameter model (table 5.1) and contains 2 additional derived parameters (population size N and density D). BUGS code is provided in Appendix 5.8.2.

Model 3 All parameters (including ψ) are assumed to be survey-specific in this model. We considered this alternative only to investigate the impact of not using information from multiple sources in our inference. However, we do not, *a priori*, consider this model to be an ecologically or scientifically reasonable alternative, because tiger densities are not expected to change substantially within a short time frame of 6 weeks. Assumptions about all other parameters are as described for Model 2. This is a 6-parameter model (table 5.1) and contains 2 additional derived parameters (population size N and density D). BUGS code is provided in Appendix 5.8.3.

The first two models are confronted using the *direct* and *step-wise* approaches described earlier. The step-wise method involves the construction of prior distributions from the available posterior distributions at the end of the first (photographic) survey. In our example, we construct the prior distributions for the parameters common to both surveys, viz., ψ and σ. We use the beta and gamma prior distributions to model for the ψ and σ, respectively, when analyzing faecal DNA capture data. The prior model for ψ is developed as follows:

$$\psi \sim beta(a, b)$$

where, a and b are the shape parameters of the β distribution. a represents the number of successes + 1 and b represents the number of failures + 1. By the properties of a β distribution, the mean m_ψ is given by $m_\psi = \frac{a}{a+b}$ and the variance v_ψ is given by $v_\psi = \frac{ab}{(a+b)^2(a+b+1)}$. Solving the 2 equations simultaneously for a and b we get

$a = m_\psi^2 \left[\frac{1-m_\psi}{v_\psi} - \frac{1-m_\psi}{m_\psi}\right]$ and $b = \frac{1-m_\psi}{m_\psi}a$

Similarly, we model the prior of σ using

$$\sigma \sim Gamma(r, \mu)$$

where, r and μ are the shape and scale parameters of the gamma distribution. If m_σ and v_σ represent the mean and variance of the gamma distribution, then $m_\sigma = r\mu$ and $v_\sigma = r\mu^2$, by the properties of the gamma distribution (Casella & Berger, 1990). Solving for r and μ, we get
$r = \frac{m_\sigma^2}{v_\sigma}$ and $\mu = \frac{v_\sigma}{m_\sigma}$

In our step-wise approach, we equate the mean and variance of the posterior distribution of ψ and σ from the photographic survey, to the mean and variance of the beta and gamma priors, respectively.

5.4.3.2 MCMC implementation and convergence diagnostics

For each data set we implemented the MCMC runs with the following settings: Number of iterations = 10000; burn-in period = 2000 iterations; number of chains = 3 and thinning rate = 1. We assessed the convergence of the MCMC chains using a combination of visual assessment of trace plots in WinBUGS and by the Gelman-Rubin convergence diagnostic (Gelman & Rubin, 1992) implemented in package 'CODA' in R. Additionally, since our problem involved multiple parameters, the multivariate version of the Gelman-Rubin diagnostic described by Brooks & Gelman (1998) was also used in our assessment of convergence.

5.4.4 Assessment of model adequacy

A recent discussion of putting Bayesian statistics into practice (Gelman & Shalizi, 2010) emphasizes the importance of 'model checking' and recommends the use of the Bayesian p-value as the appropriate measure to assess the adequacy of the model. While Link & Barker (2010) indicate that this measure may be slightly biased towards the defined model, they advocate its use for discarding unacceptable models.

The application involves the generation of replicate data sets from the resulting posterior distributions of parameters, and comparing these data sets with the empirically obtained data set. The comparison usually involves an investigation of the visual fit of the model with the aid of graphs and a more formal statistical comparison using a fit statistic, called the discrepancy measure. The problem, however, is that there are no standard omnibus discrepancy measures available for multiple applications, and often these statistics have to be derived based on the features of the data

that need exploration (Gelman & Shalizi, 2010). Royle *et al.* (2009a, 2011b) have suggested a couple of discrepancy measures for spatially-explicit capture-recapture models. They identify two components of this model, viz. a component describing individual encounters and a component that assesses the spatial clustering of animals. For the individual encounters component they use the Freeman & Tukey (1950) statistic based on individual total encounter frequencies. In a separate analysis, Royle *et al.* (2011a) use the Bayesian-p value to test whether the assumption of complete spatial randomness (CSR) serves as a good approximation to sparse spatial capture-recapture datasets. Rather than separately investigate the above two hypothesis, we preferred to directly investigate if the observed trap-specific encounter rates were comparable to the replicated trap-specific encounters obtained from posteriors, or not. We derive the discrepancy measure for this statistic as follows:

$$T_{obs} = \sum_{j=1}^{J} \frac{[x_j - E(x_j)]^2}{E(x_j)}$$

where,
x_j is the observed number of encounters in trapping device,j, over the entire sampling duration. $E(x_j)$ is the expected number of encounters in trapping device, j, over the entire sampling duration.

For a given iteration t in the MCMC analysis, $E(x_j)$ is given by:

$$E(x_j)^{(t)} = \sum_{i=1}^{N^{(t)}} \lambda_0(t) e^{-\left[\frac{d_{ij}^2(t)}{\sigma(t)}\right]}$$

where,
$N^{(t)}$ is the number of individuals in the population at iteration t
$\lambda_0(t)$ is the basal encounter rate of individuals during sampling
at iteration t
$d_{ij}(t)$ is the distance between the range center of individual i and trapping device j at iteration t
$\sigma(t)$ is the standard deviation of the movement range of a tiger at iteration t
Similarly, for the replicate data sets the discrepancy measure will be:

$$T_{rep} = \sum_{j=1}^{J} \frac{\left[x_j^{rep} - E(x_j)\right]^2}{E(x_j)}$$

where,
$\left[x_j^{rep} | E(x_j)\right] \sim Poisson\left[E(x_j)\right]$ represents a particular replicate data set.

We note here that although the use of Deviance Information Criterion (DIC) for model selection in Bayesian analysis (Spiegelhalter *et al.*, 2002) has been widespread,

the criterion has been known to perform poorly with hierarchical models (Millar, 2009). We hence did not use it in our study.

5.5 Results

The posterior densities of all the parameters estimated in the spatially-explicit capture-recapture model are summarized in tables 5.2, 5.3 and 5.4. Models 1 and 2, which combined information from the two surveys (tables 5.2, 5.3), produced a different and more precise estimate of density (posterior means=8.7(2.14) and 8.5(1.95) tigers/100 km^2 for models 1 and 2, respectively), compared to model 3 (table 5.4), which did not share any information (posterior means=12.02(3.02) and 6.65(2.37) tigers/100 km^2 for the photographic and genetic surveys, respectively). Interestingly, when these two density estimates of Model 3 are combined using the weighted estimate formula (see *Methods*), the combined density was found to be 8.73(1.86) tigers/100 km^2, which is similar to the estimate from the more formal analysis of Models 1 and 2.

Direct and step-wise approaches produced similar estimates of tiger density (posterior mean estimates of 8.7(2.14) and 8.25(2.7) tigers/100 km^2 from model 1 for the direct and step-wise methods, respectively; posterior mean estimates of 8.5(1.95) and 8.69(2.61) tigers/100 km^2 from model 2 for the direct and step-wise methods, respectively). However, estimates of other parameters in the model differed between the two approaches to a greater degree. For example, the posterior mean estimate of σ from the direct approach using model 1 = 0.38(0.13) as compared to an estimate of 0.89(0.53) using the step-wise approach.

Model assessment indicated that no model in our selection set could be defined as unacceptable ($|p_{bayesian} - 0.5| << 0.4$ for all the models). This is expected considering that the spatially explicit capture-recapture data set is relatively sparse. However, we obtained the most precise estimate of tiger density (and of other parameters), by measure of having the smallest coefficient of variation, when our data was confronted with model 2, where we assumed that only population size was the same between the two surveys. The direct and step-wise approaches produced nearly identical estimates of all parameters when inference was based on Model 2.

For all the MCMC runs (tables 5.2, 5.3 and 5.4), the Brooks & Gelman (1998) multivariate potential scale reduction factor was < 1.04, and all the potential scale reduction factors on individual parameters using the Gelman & Rubin (1992) statistic provided values between $1.00 - 1.06$. The combination of these values, along with the visual inspection of trace plots of parameters, indicates a satisfactory convergence of the 3 MCMC chains on each run.

5.6 Discussion

We confronted the problem of estimating tiger density using data from (1) a photographic capture-recapture survey and (2) a fecal DNA capture-recapture survey by proposing three plausible models and describe two ways of updating our knowledge on the central parameter (tiger density). As is evident from our results, selecting the best model is not a straightforward or readily automated exercise (Link & Barker, 2010), especially when dealing with relatively sparse data sets. At the time of conducting the fecal DNA survey, we had, readily available to us, information on tiger density from a photographic survey. Had we ignored our knowledge of some parameters from the photographic survey, we would have ended up drawing up conclusions based on results from table 5.4. Interestingly, and contrary to our original prediction, the mean posterior tiger densities, expected to be the same because of the short interval, appeared to be substantially different between the two survey methods (density of 12.18 as compared to 6.65 tigers tigers/100 km^2). Ecologically, it is not reasonable to believe that tiger densities would decline so drastically in the span of merely 6 weeks in a prey-rich habitat (Karanth *et al.*, 2006). Further, the high CV on the density parameter obtained with this assumption (table 5.4) provides evidence against the use of completely uninformative priors for the density estimation question on hand.

Applying the same ecological rationale, we favor the selection of Model 2, which combines information only on ψ, to estimate tiger density in Bandipur (table 5.3). Under this model, and consistent with our a priori prediction, we do observe that the values of the posterior mean estimates of the detection function parameter σ are survey-specific (from tables 5.3 and 5.4). The tiger density estimate from this model is also partially consistent with our prior knowledge acquired from the photographic survey in that the posterior mean estimate of tiger density lies within the 95 % credible interval of the density estimate obtained from the photographic survey. Further, the CVs for all the parameter estimates are reasonably low relative to the other models, and the Bayesian p-value estimate indicates that the model is adequate (p-value of 0.62) to describe these data. In addition, the consistency in the estimates produced by the direct and step-wise approaches with Model 2 favors the selection of this model for inference. Our results suggest some evidence that tigers deposit feces more frequently at their territory boundaries, although our data set is very sparse to come to this conclusion. Our results demonstrate that information from posterior distributions from previous surveys (in our case the photographic survey) when used as priors using our step-wise approach, also provides good estimates of density. Further, our result emphasizes the point that it is worth all the effort to incorporate prior knowledge, where available, in the design and conduct of ecological studies.

Our study demonstrates a solution to a practical dilemma faced by ecologists in being able to combine information from multiple data sources or in utilizing information from other studies or other sources about inference on a critical parameter of interest. While there is much debate among philosophers and statisticians on the issue of selecting priors or not for inference (see Lindley & Smith, 1972, for a discussion), we demonstrate with our example (using the step-wise approach) that this choice must be made after critically considering ecological and sampling rationale. Indeed, in our case, by selecting uninformative prior distributions on all the parameters, our inference about tiger density after the genetic survey in Bandipur (table 5.4) would have been inconclusive and possibly flawed. Considering that most ecological studies simply ignore prior ecological knowledge about parameters, this result serves as a timely warning. We expect that such statistical approaches which favor combining multiple sources of data, such as the integrated population models (Abadi *et al.*, 2010b,a) and our approaches, present the way forward in density estimation of rare and elusive carnivores.

Spatially-explicit capture-recapture models are now being increasingly used in density estimation for various taxa and survey methods (Borchers & Efford, 2008; Royle *et al.*, 2009a; Gardner *et al.*, 2009, 2010a; Royle *et al.*, 2011b). Our study is the first example where data sets from multiple sources have been brought together to bear on the density estimation question at hand using SECR models. Our approach of combining information on common parameters from multiple sources, without having to assume that the spatial arrangement of animals will remain the same between the two survey processes, has broad ecological applications. For instance, during the same year but in different seasons, overall animal density and abundance could still be the same, but resource-patches could vary seasonally. In our example with tigers, and more generally for carnivores with strong territoriality, it might be worth accounting for the fact that carnivore spatial arrangements are likely to be stable over long time periods. We see potential for future research in developing models, more case-specific to carnivores, where the centers of activity remain unchanged during multiple surveys. These types of models may potentially yield more precise estimates of density, although their use will require careful consideration of the consequences of individual misidentifications (Link *et al.*, 2010).

5.7 Tables

Table 5.1: A priori models for analysis and references to their results. θ_P, θ_b and θ_G are as defined in *Methods*

Model	θ_P	θ_b	θ_G	Results
Model 1	λ_0^P	ψ, σ, D, N	λ_0^G	Table 5.2
Model 2	λ_0^P, σ^P	ψ, D, N	λ_0^G, σ^G	Table 5.3
Model 3	λ_0^P, σ^P, ψ^P, D^P, N^P	None	λ_0^G, σ^G, ψ^G, D^G, N^G	Table 5.4

Table 5.2: Posterior density estimates of parameters from using Model 1 – Information on parameters accumulated ψ and σ from both surveys. Brooks & Gelman (1998) multivariate potential scale reduction factor = 1.01(step-wise) and 1.03 (direct)

Estm meth	Node	Post. mean	Post. sd	MC error	2.5 %	Median	97.5%	CV
Direct	D	8.7	2.136	0.113	5.271	8.456	13.77	0.246
	N	379.4	93.19	4.946	230	369	601	0.246
	λ_0^P	0.129	0.042	0.001	0.057	0.122	0.241	0.325
	λ_0^G	0.064	0.025	0.001	0.028	0.061	0.123	0.382
	ψ	0.476	0.116	0.006	0.290	0.464	0.752	0.244
	σ	0.384	0.128	0.004	0.211	0.358	0.703	0.332
St-wise	D	8.25	2.701	0.142	4.079	7.792	14.71	0.327
	N	360.0	117.9	6.194	178.0	340.0	642.0	0.327
	λ_0^P	0.196	0.097	0.003	0.37	0.654	0.975	0.249
	λ_0^G	0.035	0.018	0.001	0.010	0.031	0.079	0.524
	ψ	0.452	0.147	0.008	0.225	0.428	0.804	0.325
	σ	0.888	0.529	0.019	0.342	0.747	2.346	0.595

Table 5.3: Posterior density estimates of parameters from using Model 2 – Information on parameter ψ accumulated from both surveys. Brooks & Gelman (1998) multivariate potential scale reduction factor = 1.01(step-wise) and 1.04 (direct)

Estm. meth	Node	Post. mean	Post. sd	MC error	2.5 %	Median	97.5%	CV
Direct	D	8.5	1.947	0.1	5.386	8.25	13.02	0.229
	N	371	84.96	4.378	235	360	568	0.229
	λ_0^P	0.225	0.112	0.004	0.066	0.204	0.498	0.498
	λ_0^G	0.047	0.021	0.001	0.017	0.044	0.099	0.447
	ψ	0.466	0.105	0.005	0.297	0.451	0.711	0.226
	σ^G	0.562	0.251	0.008	0.264	0.507	1.172	0.447
	σ^P	0.254	0.153	0.006	0.093	0.214	0.649	0.602
St-wise	D	8.69	2.607	0.132	4.538	8.434	14.78	0.3
	N	379.2	113.8	5.749	198.0	368.0	645.0	0.3
	λ_0^P	0.196	0.097	0.003	0.37	0.654	0.975	0.249
	λ_0^G	0.046	0.021	0.001	0.016	0.043	0.098	0.46
	ψ	0.476	0.142	0.007	0.250	0.462	0.806	0.298
	σ^G	0.575	0.248	0.007	0.268	0.518	1.213	0.430
	σ^P	0.212	0.120	0.004	0.083	0.182	0.513	0.568

Table 5.4: Posterior density estimates of parameters from using Model 3 – All parameters are survey-specific. Brooks & Gelman (1998) multivariate potential scale reduction factor = 1.01

Node	Post. mean	Post. sd	MC error	2.5 %	Median	97.5%	CV
D^P	12.18	3.021	0.165	6.806	12.01	17.88	0.248
D^G	6.65	2.365	0.121	3.323	6.188	12.79	0.355
N^P	531.4	131.8	7.192	297.0	524.0	780.0	0.248
N^G	290.2	103.2	5.276	145.0	270.0	558.0	0.355
λ_0^P	0.196	0.097	0.003	0.37	0.654	0.975	0.249
λ_0^G	0.055	0.025	0.001	0.02	0.051	0.113	0.449
σ^P	0.212	0.120	0.004	0.083	0.182	0.513	0.568
σ^G	0.634	0.309	0.01	0.28	0.559	1.43	0.487
ψ^P	0.664	0.165	0.009	0.370	0.654	0.975	0.248
ψ^G	0.363	0.13	0.007	0.18	0.338	0.697	0.357

5.8 Figures

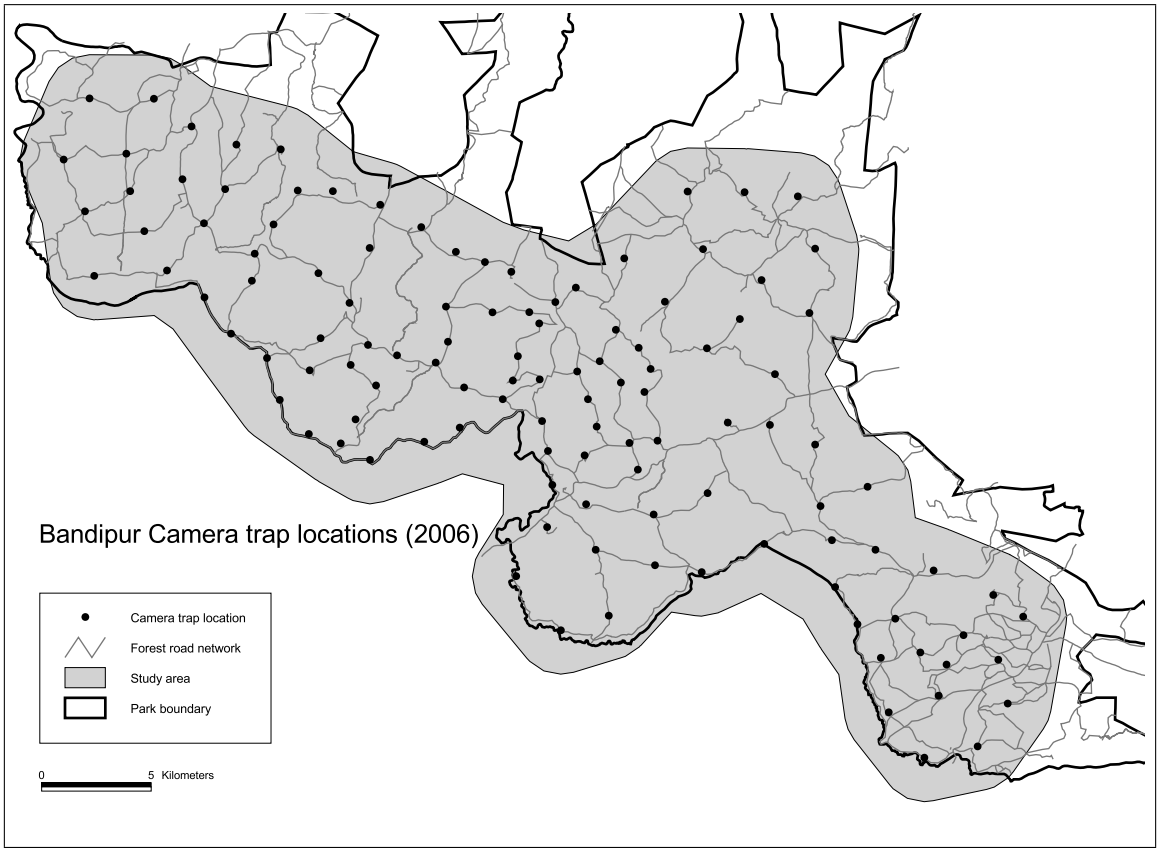


Figure 5.1: Camera trap locations used in Bandipur Tiger Reserve during 2006. The area shaded grey is the approximate area of sampling that includes a 2 km buffer around the convex polygon.

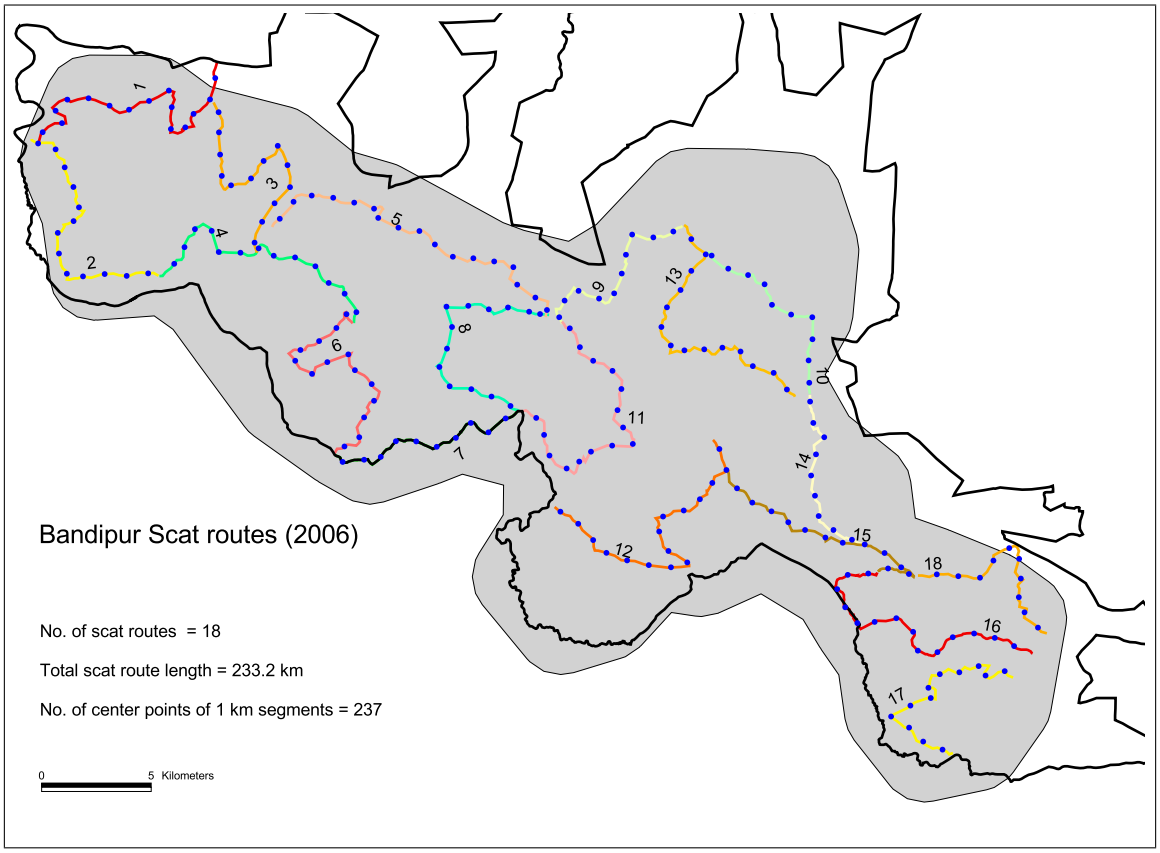


Figure 5.2: Fecal DNA sampling routes used in Bandipur Tiger Reserve in 2006. Blue dots represent mid-points of the 1 km segments along the sampling routes. These blue dots serve as the fecal DNA trapping devices in this study.The area shaded grey represents the area that was sampled using photographic sampling to represent the extent of overlap between the two sampling procedures.

5.9 Appendices

APPENDIX 5.8.1: BUGS code for Model 1 – Information about parameters psi and sigma informed by both photographic and fecal DNA surveys

```
model{
        sigma ~ dunif(0, 20)

        lam0P ~ dgamma(0.1, 0.1)
        lam0G ~ dgamma(0.1, 0.1)

        psi ~ dunif(0, 1)

        for (i in 1:M) {
                wP[i] ~ dbern(psi)
                sxP[i] ~ dunif(Xl, Xu)
                syP[i] ~ dunif(Yl, Yu)

                for (j in 1:ntrapsP) {
                        dist2P[i, j]<-(pow(sxP[i]-traplocsP[
                            j,1],2)+pow(syP[i]-traplocsP[j
                            ,2],2))
                        tmpP[i,j]<- lam0P*exp(-dist2P[i,j]/(
                            sigma))
                        muyP[i,j]<- tmpP[i,j]*wP[i]
                        yaugP[i,j] ~ dpois(muyP[i,j])
                }
        }
        for (i in 1:M) {
                wG[i] ~ dbern(psi)
                sxG[i] ~ dunif(Xl, Xu)
                syG[i] ~ dunif(Yl, Yu)

                for (m in 1:ntrapsG) {
                        dist2G[i, m]<-(pow(sxG[i]-traplocsG[
                            m,1],2)+pow(syG[i]-traplocsG[m
                            ,2],2))
                        tmpG[i,m]<- lam0G*exp(-dist2G[i,m]/(
                            sigma))
                        muyG[i,m]<-tmpG[i,m]*wG[i]
                        yaugG[i,m] ~ dpois(muyG[i,m])
                }
        }

NP<-sum(wP[1:M])
DP<-(NP/((Xu-Xl)*(Yu-Yl)))*4

NG<-sum(wG[1:M])
DG<-(NG/((Xu-Xl)*(Yu-Yl)))*4
}
```

APPENDIX 5.8.2: BUGS code for Model 2 – Information about parameter psi informed by both photographic and fecal DNA surveys

```
model{
        sigmaP ~ dunif(0, 20)
        sigmaG ~ dunif(0, 20)

        lam0P ~ dgamma(0.1, 0.1)
        lam0G ~ dgamma(0.1, 0.1)

        psi ~ dunif(0, 1)

        for (i in 1:M) {
                wP[i] ~ dbern(psi)
                sxP[i] ~ dunif(Xl, Xu)
                syP[i] ~ dunif(Yl, Yu)

                for (j in 1:ntrapsP) {
                        dist2P[i, j]<-(pow(sxP[i]-traplocsP[
                            j,1],2)+pow(syP[i]-traplocsP[j
                            ,2],2))
                        tmpP[i,j]<- lam0P*exp(-dist2P[i,j]/(
                            sigmaP))
                        muyP[i,j]<- tmpP[i,j]*wP[i]
                        yaugP[i,j] ~ dpois(muyP[i,j])
                 }
        }

        for (i in 1:M) {
                wG[i] ~ dbern(psi)
                sxG[i] ~ dunif(Xl, Xu)
                syG[i] ~ dunif(Yl, Yu)

                for (m in 1:ntrapsG) {
                        dist2G[i, m]<-(pow(sxG[i]-traplocsG[
                            m,1],2)+pow(syG[i]-traplocsG[m
                            ,2],2))
                        tmpG[i,m]<- lam0G*exp(-dist2G[i,m]/(
                            sigmaG))
                        muyG[i,m]<-tmpG[i,m]*wG[i]
                        yaugG[i,m] ~ dpois(muyG[i,m])
                }
        }

NP<-sum(wP[1:M])
DP<-(NP/((Xu-Xl)*(Yu-Yl)))*4

NG<-sum(wG[1:M])
```

```
DG<-(NG/((Xu-Xl)*(Yu-Yl)))*4
}
```

APPENDIX 5.8.3: BUGS code for Model 3 – All parameters are completely survey–specific

```
model{
        sigmaP ~ dunif(0, 20)
        sigmaG ~ dunif(0, 20)

        lam0P ~ dgamma(0.1, 0.1)
        lam0G ~ dgamma(0.1, 0.1)

        psiP ~ dunif(0, 1)
        psiG ~ dunif(0, 1)

        for (i in 1:M) {
                wP[i] ~ dbern(psiP)
                sxP[i] ~ dunif(Xl, Xu)
                syP[i] ~ dunif(Yl, Yu)

                for (j in 1:ntrapsP) {
                        dist2P[i, j]<-(pow(sxP[i]-traplocsP[
                            j,1],2)+pow(syP[i]-traplocsP[j
                            ,2],2))
                        tmpP[i,j]<- lam0P*exp(-dist2P[i,j]/(
                            sigmaP))
                        muyP[i,j]<- tmpP[i,j]*wP[i]
                        yaugP[i,j] ~ dpois(muyP[i,j])
                 }
        }

        for (i in 1:M) {
                wG[i] ~ dbern(psiG)
                sxG[i] ~ dunif(Xl, Xu)
                syG[i] ~ dunif(Yl, Yu)

                for (m in 1:ntrapsG) {
                        dist2[i, m]<-(pow(sxG[i]-traplocsG[m
                            ,1],2)+pow(sy[i]-traplocsG[m
                            ,2],2))
                        tmpG[i,m]<- lam0G*exp(-dist2[i,m]/(
                            sigmaG))
                        muyG[i,m]<-tmpG[i,m]*wG[i]
                        yaugG[i,m] ~ dpois(muyG[i,m])
                }
        }

NP<-sum(wP[1:M])
DP<-(NP/((Xu-Xl)*(Yu-Yl)))*4

NG<-sum(wG[1:M])
```

```
DG<-(NG/((Xu-Xl)*(Yu-Yl)))*4
}
```

Chapter 6

An examination of index-calibration experiments: counting tigers at macroecological scales

6.1 Citation

Gopalaswamy, A. M., Delampady, M., Karanth, K. U., Kumar, N. S., Macdonald, D. W. (2015). An examination of index-calibration experiments: counting tigers at macroecological scales. *Methods in Ecology and Evolution*

6.2 Abstract

1. An index-calibration experiment involves rigorous estimation of animal abundance at a small scale to calibrate a, less rigorously derived, index of abundance. The efficacy of such index-calibration experiments has been a matter of much controversy. In this study, we develop theoretical models and confront them with empirical data on large scale index-calibration experiments on tigers *Panthera tigris* to advance our understanding of this controversy.

2. We propose two models that describe typical sampling processes involved in typical index-calibration experiments. Using analytical derivations and some simulations, we evaluate the relative roles of these sampling parameters on the R^2 (coefficient of determination) statistic - a common inferential tool in such calibration experiments.

3. We make predictions about the R^2 statistic using our theoretical derivations and estimates from large-scale occupancy surveys of tigers in India. We then compare our predictions with empirical estimates of two tiger sign index-calibration experiments (IC-Karanth and IC-Jhala).

4. Our theoretical models shows that the R^2 statistic increases when individual-specific detection probability is high and is constant, increases when the variance-to-mean ratio of abundance increases, increases when precision of abundance estimates improves and declines when mean abundance increases. All predictions about the two index-calibration experiments showed a poor performance of the R^2 statistic (R^2 ¡ 0.40). Inference from IC-Karanth was extremely poor ($R^2 = 0.0004$; Pvalue < 0.00001), and comparable to model predictions. Anomalously, inference from IC-Jhala was exceedingly high ($R^2 = 0.95$; Pvalue < 0.0001). 5. Our study show that such direct index-calibration experiments using the R^2 statistic yield poor inferences unless all the sampling process parameters lie within a limited range. Ignoring the consequence of the effect of these parameters during survey design could result in vast amount of resources being consumed for virtually little gain in ecological inference. Analysis using joint likelihood models, with appropriate survey designs, may be more fruitful than clinging on to such composite R^2-based models.

6.3 Introduction

A prerequisite to tackling any problem in macroecology is a reliable assessment of state variables such as abundance or relative abundance at large geographic and temporal scales (May, 1999; Clark *et al.*, 2001; Hastings *et al.*, 2005). Whether such patterns are sought to help understand erosion of biodiversity, speciation in organisms or the spread of infectious diseases, we need accurate assessments of these state variables. However, logistical challenges and resource constraints typically preclude application of rigorous field surveys to estimate abundance at large scales. Consequently, field ecologists utilize a variety of indices in an attempt to overcome these challenges (Johnson, 2008). In this study, we aim to uncover some of the theoretical underpinnings of a very basic and long-standing controversy in ecology, about the utility of abundance indices as surrogates for true estimates of animal abundance at macroecological scales. We then confront these theoretical findings with empirical analysis from some recent macroecological scale surveys of tiger *Panthera tigris*, an endangered species of high conservation importance, distributions and abundance in India (see Karanth & Nichols, 2000; Karanth *et al.*, 2004; Jhala *et al.*, 2011b,a; Karanth *et al.*, 2011b). In all these studies, huge resources were invested across large spatial scales, thereby accentuating the importance and also providing the motivation for this study.

To overcome logistical constraints in assessing animal abundance at large scales, a commonly adopted approach among field ecologists is to conduct surveys for animal abundance at a small scale using rigorous methods, such as capture-recapture

sampling (Williams *et al.*, 2002) or distance sampling (Buckland *et al.*, 2001), and, try to use these estimates to calibrate *directly* an index of abundance, derived from more easily conducted surveys (such as a raw count of animal signs), using a simple linear regression model. The calibrated index is then used to draw inference on animal abundance at larger spatial or temporal scales. Intuitively, this approach seems ecologically reasonable, since animal abundance may be monotonically, and also perhaps linearly, associated with signs deposited. The approach was first proposed by Cochran (1937) and later formalized as 'double sampling' by Eberhardt & Simmons (1987).

A tide of criticisms followed in the literature, peaking during the period 1995-2005 (see Johnson, 2008, for a mention of over 20 studies), on the general utility of abundance indices. The thrust of these criticisms was that predictions about population size from such indices are obscured by the variation in detection rates over time or space. Despite strong criticisms, attempts to calibrate indices directly against estimates of absolute densities have continued persistently since 2005 (see Gusset & Burgener, 2005; Stephens *et al.*, 2006; Barea-Azcón *et al.*, 2007; Gruber *et al.*, 2008; Johnson, 2008; Funston *et al.*, 2010; Price & Rachlow, 2011; Bengsen *et al.*, 2011; Keane *et al.*, 2011), based on the assumption that detection probabilities may be constant or can be controlled for in some manner.

It appears to us that this persistence rests largely on claims of apparent successes (e.g., Kindberg *et al.*, 2009; Jhala *et al.*, 2011b)($R^2 = 0.6 - 0.88$, $R^2 = 0.95$, respectively) or at least moderate successes (e.g., Bonesi & Macdonald, 2004), even while improved alternatives to 'double-sampling' have been suggested (Pollock *et al.*, 2002) in which the problem of imperfect detection is explicitly estimated. These studies have generated hope that, at least for specific ecological systems, indices when calibrated directly using simple linear regression models may work reasonably well. It therefore becomes imperative to ask what processes can lead to a successful calibration experiment? And are answers from such calibrations reliable? The answer to these questions would lie in uncovering all the parameters (and not limited to detection probability) that determine the values of the coefficient of determination (R^2 statistic) in such index-calibration experiments, because a high R^2 value usually implies strong inference in these experiments, if the assumed relationship between and index and true abundance is linear (Kvålseth, 1985).

We thus aim to address the following objectives in this study:

1. Propose models to describe the observation process in index-calibration experiments. Consequently, derive formulae for the R^2 statistic based on these models.

2. Make predictions of the R^2 statistics from these theoretical models utilizing estimates of detection probability from studies of tiger occupancy across India (Jhala *et al.*, 2011a; Karanth *et al.*, 2011b)

3. Compare these predictions with the R^2 statistics derived from two empirical, large-scale tiger sign abundance index-calibration experiments, namely, "IC-Jhala" (Jhala *et al.*, 2011b) and "IC-Karanth" (Karanth & Nichols, 2000, 2002; Karanth *et al.*, 2004; Karanth & Kumar, 2005; Andheria, 2006).

For many reasons, this pair of index-calibration experiments (IC-Jhala and IC-Karanth) aimed to assess tiger abundance from indices at large geographic scales, serves as an ideal case for investigation. First, we have two examples of very similar large-scale index-calibration experiments where an index, such as tiger sign encounter rates, is calibrated against rigorous estimates of tiger density by photographic capture-recapture sampling in a representative set of sites using a simple linear regression type of approach (Eberhardt & Simmons, 1987). Second, these two studies report two extreme degrees of calibration success ($R^2 = 0.95$ in IC-Jhala in comparison to R^2=0.0004 in IC-Karanth), making our investigation more purposeful. Third, we have two similar large-scale assessments of the tiger spatial distribution from tiger sign surveys, wherein, the problem of imperfect detection is explicitly accounted for (Karanth *et al.*, 2011b; Jhala *et al.*, 2011a) and enabling us to use estimates of detection probability in our analysis. Fourth, making direct inferences from some of these studies, without critical examination, leads to some unreal ecological implications - for example, an unrealistic increase in tiger density at a country-wide scale by 49% in a span of only 4 years (Karanth *et al.*, 2011a). Fifth, and most importantly, huge resources have been spent on these experiments on a species which is critically endangered, and thus addresses an important conservation need. For example, IC-Jhala study alone, report an approximate expenditure of USD 380,000 and a manpower investment of approximately 19,740 man-days of effort (for 21 sites), while IC-Karanth report an approximate expenditure of USD 40,000 and a manpower investment of 3000 man-days (for 8 sites).

6.4 Material and Methods

6.4.1 Sampling situation

We envisage a straightforward sampling situation. A set of study sites are selected for a species with presumably varying densities. Density or abundance in these sites are measured using rigorous methods, such as capture-recapture sampling, or distance sampling, or double-observer survey methods or even where true abundances might be known, for example, in re-introduced populations (Diefenbach *et al.*, 1994).

It could also involve measuring density or abundance of a species at a particular site over a period of time, as in most assessments of population dynamics (de Valpine & Hastings, 2002). At the time of each measuring densities, the investigator will also measure an index of abundance or density. These could be measures such as encounter rates of an animal from specific hides, camera trapping rates, sign encounter rates, scent-station markings or even dung counts (Long & MacKay, 2008). We are now interested in assessing a possibly direct linear relationship between estimated densities or abundances of a species with its index of abundance measured concurrently.

6.4.2 Objective 1: Proposing models to uncover the parameters of the R^2 statistic in index-calibration experiments

The population size N of a species in a site can be defined as a random variable with a mean μ and variance σ^2. This variance is indicative of the natural variability in abundance over multiple points in time or space. Consider a situation where these N animals are exposed to a sampling array on k occasions. The array, for example, could be a spatial arrangement of automated camera traps triggered by animals to generate a sample of photographs that generate a count of a particular species. The sampling array could also be a spatial complex of natural trail segments, which are searched by observers to record and count animal signs such as tracks or scats which cannot be assigned to particular individuals (Royle *et al.*, 2011b). In both cases, k can be thought of as some discrete time period (e.g., a day, a week etc.) so that encounters of species signs between any two periods may be considered as statistically binary independent events. We can then define a random variable C_i as the raw count of encounters of an individual animal i in a sampling array after being exposed for k occasions, where $i = 1, 2, 3, \ldots N$.

Model 1: When p^* is a constant

In the first instance, we assume that every animal i has a constant and equal probability p^* of being sampled on a given sampling occasion. If p^* extends over the dimension of comparison (time or space), then this is the ideal situation for an index - a perfect association with the quantity of interest (the abundance). This situation is analogous to the M_o model in standard capture-recapture literature (Otis *et al.*, 1978). The encounters of each individual are modeled as below:

$$
\begin{aligned}
[C_1|k, p^*] &\sim \text{Binomial}\,[k, p^*] \\
[C_2|k, p^*] &\sim \text{Binomial}\,[k, p^*] \\
\vdots \qquad &\qquad \vdots \qquad\qquad \vdots \\
[C_N|k, p^*] &\sim \text{Binomial}\,[k, p^*]
\end{aligned}
$$

where, p^* is the probability of detecting an individual animal in the sampling array on any given sampling occasion.

The total number of encounters S after time period k, is therefore, simply a sum of the individual binomial counts.

$$
\begin{aligned}
S &= C_1 + C_2 + C_3 + \cdots + C_N \\
[S|N, k, p^*] &\sim \text{Binomial}\,[kN, p^*]
\end{aligned}
$$

Our aim is to investigate the correlation between actual animal abundance N and the total encounters S, i.e. $\text{Cor}(N, S)$. We note that $R^2 = [\text{Cor}(N, S)]^2$. But, in reality, the true animal abundance N is rarely known, and one has to rely on an estimate of N, say $\hat{N}$. We would then be interested in the relationship $\text{Cor}(\hat{N}, S)$. This relationship forms the basis for inference about the strength of the relationship between an index of abundance and the estimated animal abundance.

$$
\text{Cor}\left(\hat{N}, S\right) = \frac{\text{Cov}\left(\hat{N}, S\right)}{\sqrt{\text{Var}\left(\hat{N}\right)\text{Var}(S)}}
$$

We define an additive model for $\hat{N}$, i.e. $\hat{N} = N + \epsilon$, where ϵ can be defined as a sampling error term that is independent of N. While this assumption may seem counter-intuitive, because we may expect ϵ to increase when N increases, often in practice, this may not be the case (see Karanth *et al.*, 2004, to observe the independent nature of the error term in density estimates). This is because unpredictability in ϵ is largely influenced by the variation of detection probabilities during the estimation of N. Consequently, ϵ is assumed to be independent of S.

$$\begin{aligned}\text{Cor}\left(\hat{N}, S\right) &= \frac{\text{Cov}\left(\hat{N}, S\right)}{\sqrt{\text{Var}\left(\hat{N}\right)\text{Var}(S)}} \\ &= \frac{\text{Cov}\left(N+\epsilon, S\right)}{\sqrt{\left[\text{Var}\left(N\right)+\text{Var}\left(\epsilon\right)\right]\text{Var}(S)}} \\ &= \frac{\text{Cov}\left(N, S\right)}{\sqrt{\left[\text{Var}\left(N\right)+\text{Var}\left(\epsilon\right)\right]\text{Var}(S)}}\end{aligned}$$

The presence of $\text{Var}(\epsilon)$ in the denominator indicates that $\text{Cor}\left(\hat{N}, S\right) \leq \text{Cor}\left(N, S\right)$. The term $\text{Var}(N)$ is the true variance (spatial/temporal). And the term $\text{Var}(\epsilon)$ is the sampling variance associated with the measurement of N. If $\text{Var}(\epsilon) = \tau^2$, it can be shown that (see Appendix 6.8.1 for a complete derivation)

$$R^2 = \frac{\sigma^2 k p^*}{Q\left(\mu - p^*\mu + k p^* \sigma^2\right)}$$

Here, Q is the ratio $\frac{\sigma^2+\tau^2}{\sigma^2}$, which we term as the 'sampling factor' and $Q \geq 1$.

We may consider factors such as deposition rate of animal sign in track, dung and scat surveys, arrival rate of animals in camera trapping experiments and substrate quality or choice of trails in track detection experiments to influence the parameter p^*. Similarly, we may consider factors such as decay rate of animal sign, search or walk effort in sign detection experiments or the number of trap nights in camera trapping experiments, to influence the parameter k (Long & MacKay, 2008).

In practice, the number of sampling occasions k may not be known to us. For example, a survey that involves detection of field signs deposited by an animal, will not necessarily contain information about the number of days prior to the sampling over which these signs may have been deposited. Such variation in k will likely cause a variation in R^2. We therefore determine the upper and lower bounds for R^2 to incorporate uncertainty in k. Also, p^* may not be directly estimable in our problem - although, there are other approaches that permit inference about p^*, for example, distance-sampling (Buckland *et al.*, 2001) and double-observer survey methods (Nichols *et al.*, 2000). Similarly, from modern occupancy modeling studies (MacKenzie *et al.*, 2002; Hines *et al.*, 2010) it is possible to estimate a related probability p, the probability of detecting an individual animal at least once during a survey, so we define
$p_k^* = 1 - (1-p)^{1/k}, 0 < p < 1, k = 1, 2, 3, \ldots$
and p_k^* can be defined as the value of p^*, conditional upon the value of k.

It can then be shown that (see Appendix 6.8.2 for a complete derivation) for any given value for k, the following relation must hold true.

$$\frac{V\ln\left(\frac{1}{1-p}\right)}{Q\left[1+V\ln\left(\frac{1}{1-p}\right)\right]} < R^2 \leq \frac{Vp}{Q\left[Vp+(1-p)\right]}$$

Here, $V = \frac{\sigma^2}{\mu}$ is the variance-to-mean ratio of N. Interestingly, when V=1(i.e., for the Poisson model), the above relation reduces to

$$\frac{\ln\left(\frac{1}{1-p}\right)}{Q\left[1+\ln\left(\frac{1}{1-p}\right)\right]} < R^2 \leq \frac{p}{Q}$$

From the above formulae, we plot graphs (see Figures 6.1 and 6.2) to demonstrate the relative effects of V, Q and p on the upper and lowers bounds of the R^2 statistic in index-calibration experiments under the assumptions of Model 1, where p^* is a constant. See Appendix 6.8.3 for the R code to enable users to generate graphs of R^2 bounds for user-specified parameter values under this model.

6.4.2.1 Model 2: When p^* is a variable

We establish another relationship where p^* is a variable, and perhaps, more aptly represents the ground situations due to inexplicable heterogeneity in capture probabilities across individuals and sites (Karanth *et al.*, 2004, see). Such unpredictable heterogeneity has been modeled using the beta-binomial model (Dorazio & Royle, 2003; Royle, 2006) in capture-recapture analysis. We employ this formulation here to derive the correlation between N and S as earlier. So,

$$p_k^* \sim \text{Beta}\left[\alpha_k, \beta_k\right]$$

where α_k and β_k are shape parameters of the beta distribution which are influenced by the value of k, the number of sampling occasions. The count of the animal encounters S can then be modeled as:

$$[S|N, k, \alpha_k, \beta_k] \sim \text{beta} - \text{binomial}\left[kN, \alpha_k, \beta_k\right]$$

It can be shown that (see Appendix 6.8.4 for a detailed derivation):

$$R^2 = \frac{\sigma^2 k\alpha_k(\alpha_k+\beta_k+1)}{Q\left[\sigma^2 k\alpha_k(\alpha_k+\beta_k+1)+\alpha_k\beta_k\mu+\beta_k^2\mu+k\beta_k\mu^2+k\beta_k\sigma^2\right]}$$

Often, k is not known to us. But we know,

$$p_k^* = 1 - (1-p)^{1/k} \sim \text{beta}\left[\alpha_k, \beta_k\right]$$

When $k = 1$, $p^* = p$ and we can assume,

$$[p|a, b] \sim \text{beta}\left[a, b\right]$$

Here, p is the probability of detecting an animal at least once and a and b are the respective shape parameters describing the distribution of p. It can then be shown that (see Appendix 6.8.5 for a detailed derivation):

$$\frac{V\left[\ln\left(\frac{a+b}{b}\right)\right]^2}{Q\left(V\left[\ln\left(\frac{a+b}{b}\right)\right]^2 + \ln\left(\frac{a+b}{b}\right) + (\mu + V)\frac{a}{b(a+b+1)}\right)} < R^2$$

$$R^2 < \frac{aV(a+b+1)}{Q\left[aV(a+b+1) + ab + b^2 + (\mu+V)b\right]}$$

From the above formulae, we plot graphs (see Figures 6.3 and 6.4) to demonstrate the relative effects of V, Q, μ, a and b on the R^2 statistic in index-calibration experiments under the assumptions of this model (p^* is a variable). See Appendix 6.8.6 for the R code to enable users to generate graphs of R^2 bounds for user-specified parameter values.

6.4.3 Objective 2: Application to large scale, macro-level assessments of tiger distribution and abundance in India

We use our theoretical derivations of the R^2 statistic and estimates of detection probabilities from recent studies of tiger occupancy in India (Jhala *et al.*, 2011a; Karanth *et al.*, 2011b) to examine recent large-scale tiger index-calibration experiments (IC-Jhala and IC-Karanth).

Our analysis of these tiger studies at macroecological scales involves four critical steps:

1. To establish the relationship between the detection probability parameters of two occupancy models, namely, MacKenzie *et al.* (2002) and Hines *et al.* (2010) used in Jhala *et al.* (2011a) and Karanth *et al.* (2011b) respectively.

2. To standardize the scale of detection probability estimates across different studies to enable a valid comparison.

3. To predict the R^2 statistics for index-calibration experiments, using the detection probability estimates from the tiger occupancy studies, with the two theoretical models (Model 1 and Model 2) proposed in this study.

4. To compare the theoretical predictions of R^2 statistics from the tiger distribution surveys against the actual estimates of R^2 from the tiger index-calibration experiments (IC-Jhala and IC-Karanth).

6.4.3.1 Establishing the relationship between the detection probability parameters of two occupancy models

In order to perform a successful comparison, it is necessary, as a first step, to resolve methodological uncertainties (Stewart, 2010) arising from the constituent individual studies. Thus, in our case it is relevant to establish the relationship between two related detection probability parameters in the studies of MacKenzie *et al.* (2002) and Hines *et al.* (2010), since these are the two different methods employed in tiger occupancy studies (Jhala *et al.*, 2011a; Karanth *et al.*, 2011b) examined here.

The basic occupancy model (MacKenzie *et al.*, 2002) estimates the parameter p, the probability of detecting a species, given that the sampling unit is occupied. The sampling situation in this model involves the use of temporal replicates (repeated sampling within a unit) to estimate p. An alternative, and more advanced model (Hines *et al.*, 2010), permits for spatial replication (sampling smaller parts within a unit) and temporal replication (that allows for correlation) during sampling. Here, the parameter p is further broken down into, essentially, two parameters: a) the local occupancy parameter θ_0, which is the probability that a randomly selected sample sub-unit is occupied given that the larger sampling unit is occupied and b) the local detection probability parameter, p_z, which is the probability of detecting a species on sampling sub-unit z, given that the sampling sub-unit is occupied and the larger sampling unit is occupied. It follows, therefore, that $p = \theta_0 \times p_z$.

6.4.3.2 Standardizing the scale of detection probability estimates across multiple studies to enable a valid comparison

When attempting to compare detection probability estimates across multiple studies, it is often necessary to ensure that the per occasion sampling effort is standardized across these studies. Per occasion sampling effort could be, for example, defined as the number of camera trap nights per sampling occasion in typical camera trap studies (involving temporal replication) (Karanth & Nichols, 1998) or the number of kilometres walked per sampling occasion (or per spatial replicate) in trail-based occupancy surveys (Hines *et al.*, 2010).

So, if $p_{(m)}$ represents the detection probability for m units of effort, then we have:

$$p_{(m)} = 1 - \left(1 - p_{(1)}\right)^m$$
$$p_{(1)} = 1 - \left(1 - p_{(m)}\right)^{1/m}$$

Here, $p_{(1)}$ represents detection probability per sampling occasion.

We are also interested to investigate the effect of heterogeneity in $p_{(1)}$ on our predictions. As per our model description (Model 2), when number of sampling occasions is one $p^* = p = p_{(1)}$. Hence, we assume that $p_{(1)}$ varies according to a beta distribution with shape parameters a and b.

We employ the method of moments in our case to obtain $\hat{a}$ and $\hat{b}$, as below:

$$\hat{a} = \hat{\bar{p}}_{(1)} \left(\frac{\hat{\bar{p}}_{(1)}(1 - \hat{\bar{p}}_{(1)})}{\hat{v}} - 1 \right)$$
$$\hat{b} = (1 - \hat{\bar{p}}_{(1)}) \left(\frac{\hat{\bar{p}}_{(1)}(1 - \hat{\bar{p}}_{(1)})}{\hat{v}} - 1 \right)$$

where, $\hat{\bar{p}}_{(1)}$ is the sample mean of the detection probabilities, $\hat{v}$ is estimated sample variance of the detection probabilities. We recognize that although the method of moments (MOM) estimator, in general, is not the best estimator (Casella & Berger, 1990), it is more functional in this analysis since we are essentially involving only two moments and the MOM performs reasonably well at lower orders.

6.4.3.3 Predicting R^2 statistics using the detection probability estimates from the tiger occupancy studies

From our proposed models it is necessary to also estimate the variance-to-mean ratio V and the sampling factor Q. A common approach used to establish direct relationships between an index of abundance and actual abundance is to treat a set of populations as a superpopulation and each selection of a set of populations as a sample emerging from this superpopulation. In such situations, it is necessary to employ a variance decomposition approach (Link & Nichols, 1994) to accurately estimate $\hat{V}$ from the sampled population abundances. In our example with the tiger studies, the variability among a set of estimates for n tiger populations each of size N_i, is given by:

$$\chi^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(\hat{N}_i - \hat{\bar{N}} \right)^2$$

where

$$\hat{\bar{N}} = \frac{1}{n} \sum_{i=1}^{n} \hat{N}_i$$

And $\hat{\bar{N}}$ is the unbiased estimate of μ, the mean abundance among the set of populations.

Here χ^2 measures the composite variability due to both the natural variation of abundance among the set of populations σ^2 as well as the sampling variance τ^2 occurring due to sampling and data analytic procedures. So, $\mathrm{E}\{\chi^2\} = \sigma^2 + \tau^2$. If t^2 represents the unbiased estimate of τ^2, we have,

$$t^2 = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left\{\mathrm{Var}\left(\hat{N}_i | N_i\right)\right\}$$

If , then the true spatial variation σ^2 can be estimated using:

$$\hat{\sigma^2} = \chi^2 - t^2$$

And the estimate of the variance-to-mean ratio is thus given by $\hat{V} = \frac{\chi^2 - t^2}{\hat{\bar{N}}}$. Similarly, the sampling factor Q can be estimated by:

$$\hat{Q} = \frac{\chi^2}{\chi^2 - t^2}$$

Thus, it is now possible to estimate the upper and lower bounds of the R^2 statistic, based on the empirical estimates of $\hat{V}$, $\hat{Q}$, $\hat{p_{(1)}}$ (for Model 1, where p^* is a constant). The bounds may also be estimated with estimates of $\hat{V}$, $\hat{Q}$, $\hat{\bar{N}}$, $\hat{a}$ and $\hat{b}$ (for Model 2, where p^* is a variable).

6.4.3.4 Empirical estimation of R^2 from tiger index-calibration experiments and assessments against predictions of R^2 from detection probability estimates.

In their large-scale field index-calibration experiment, IC-Jhala, attempt to calibrate encounter rates of tiger signs (tracks and scat deposits) with actual tiger densities (estimated using camera trap surveys within a capture-recapture framework) using a simple linear regression model and assessed the fit of the R^2 statistic. Similarly, we fit a simple linear regression model using tiger sign encounter rate and tiger density estimates from IC-Karanth (Karanth & Nichols, 2000, 2002; Karanth *et al.*, 2004; Karanth & Kumar, 2005; Andheria, 2006), and assessed the R^2 statistic.

6.5 Results

6.5.1 Objective 1: Uncovering the sampling process parameters of the R^2 statistic in index-calibration experiments

6.5.1.1 The R^2 statistic when p^* is a constant

Figures 6.1 and 6.2 show the performance of the R^2 statistic, when p, the probability of detecting an animal at least once during the survey, V, the variance-to-mean ratio and Q, the sampling factor, are set at different values in the simulation. As indicated by the figure, the performance of the R^2 statistic increases as the value of p increases. However, an increase in the value of p also introduces uncertainty in the predictability of R^2 owing to the uncertainty in the value of k, the number of sampling occasions used in the study. We notice this effect in the widening of the gap between the upper and lower bounds of R^2 as p becomes larger. Although the extent of this uncertainty is not noticeable for this model. Similarly, inference using R^2 statistic becomes stronger if the natural variation (spatial or temporal) in the population becomes larger, as indicated by the increase in R^2 value when V, the variance-to-mean ratio increases. Hence, selection of sites for conducting the index-calibration experiment becomes important, since good inference using the R^2 is possible only with a good spread in the density estimates of the species in concern at the selected sites (Hanski *et al.*, 1994). Interestingly, when $\mu = \sigma^2$, i.e., $V = 1$, which implies a simple random fluctuation in population size, R^2 is independent of μ or σ^2. Finally, the value of R^2 statistic drops when the value of Q increases. This implies that much of the calibration success, will depend on the precision of density estimates at the sampled sites.

6.5.1.2 The R^2 when p^* is a variable

Figures 6.3 and 6.4 shows the performance of the R^2 statistic, when p^* is a variable. In these simulations, the shape parameters a and b are set so that $a = b$. However, we have explored the effect of increase in the magnitude of these parameters on the R^2 statistic. We observe, that the increase in the magnitude of a and b causes an increase in the value of the R^2 statistic. We observe that the increase in the magnitude of a and b causes an increase in the value of the bounds of the R^2 statistic. This is expected because increases in the magnitude of a and b results in more concentrated accumulation of probability mass at the centre of the distribution, thus decreasing the uncertainty in p. However, the uncertainty in the value of R^2, due to uncertainty in the value of k, the number of sampling occasions utilized in the study, drastically increases in this model. The effects of V and Q on the value of the R^2 statistic is quite similar to their effects in the model where p^* is a constant. In

addition, an interesting feature of this model is that the value of the R^2 statistic is inversely density-dependent. This is likely because a high average abundance leads to a higher variation in expected sign counts as a result of a fluctuating p. However, the extent of this effect apparently seems low, relative to the other effects.

6.5.2 Objective 2: Empirical assessments of tiger index-calibration experiments in India

6.5.2.1 Relationship between the detection probability parameters of two occupancy models

We assess the estimates of detection probabilities associated with surveys for signs of tigers along well-used trails in these two large scale occupancy studies: Karanth *et al.* (2011b) and Jhala *et al.* (2011a). The estimate of the mean replicate-level detection probability (i.e. on a 1-km long trail segment), $\hat{\bar{p}}$, was found to be 0.17(0.17) in the Malenad-Mysore Tiger Landscape of Karnataka, India (Karanth *et al.*, 2011b). The local level occupancy on a randomly selected trail segment of length 1-km, $\hat{\theta_0}$ was found to be 0.418(0.079).

However, we recognize that the sampling unit size was 100 km^2 and 188 km^2 in the occupancy studies of Jhala *et al.* (2011a) and Karanth *et al.* (2011b), respectively. So in order to standardize this area effect, we employ the identity, $\theta_{0adj} = 1 - (1 - \theta_0)^{1.88}$ to make the detection probability estimates from Karanth *et al.* (2011b) comparable with those from Jhala *et al.* (2011a), where θ_{0adj} is the adjusted value of θ_0 in Karanth *et al.* (2011b).

To arrive at the probability of detecting a tiger sign on a segment of length 1-km given presence in the sample unit, say $p_{(1)}$, as defined in MacKenzie *et al.* (2002), we multiply the constant $\hat{\theta_0}$, with the random variable $\hat{\bar{p}}$, resulting in $\hat{\bar{p}}_{(1)} = \theta_{0adj} \times 0.17(0.17) = 0.134(0.134)$ from the Karanth *et al.* (2011b) study. We use parametric bootstrapping (Efron & Tibshirani, 1986) to account for the sampling uncertainty around $\hat{\theta_0}$. We generated 10000 bootstrap values using a $Normal(mean = 0.418, sd = 0.079)$ distribution and to compute $\hat{\theta}_{0adj}$ and consequently $\hat{SE}$ of $\hat{\bar{p}}_{(1)}$. Appendix 6.8.7 contains the R code.

6.5.2.2 Standardizing scale of detection probability estimates across multiple studies to enable a valid analysis

Jhala *et al.* (2011a) apply the single-season occupancy model (MacKenzie *et al.*, 2002) to derive estimates of occupancy over three important tiger landscapes across India. They have used spatial replicates of length 5-km in their design, so standardizing the detection probability estimates per sample of length 5-km to a 1-km segment, implies that $m = 5$, giving us:

$$p_{(1)} = 1 - \left(1 - p_{(5)}\right)^{1/5}$$

Applying the above identity to estimates of $p_{(5)}$ obtained from Jhala *et al.* (2011a) provides us with equivalent estimates of $p(1)$ from the same study (see Table 7.1). Similarly, we generated 10000 bootstrap values using a Normal(mean=$p_{(5)}$, sd=0.001) distribution and computed $\hat{SE}$ of $\hat{\hat{p}}_{(1)}$.

The only source of information on the variation of $p_{(1)}$, in our case, comes from the estimate of $p_{(1)}$ from the Karanth *et al.* (2011b) study. From this estimate, $\hat{p_{(1)}^{\hat{}}} \left[\hat{\text{SE}}\left(\hat{p_{(1)}^{\hat{}}}\right)\right] = 0.134(0.134)$, we observe that the sample coefficient of variation (standard error/mean), $c_v = 1$. We apply this extent of variation, i.e., $c_v = 1$, to estimate the shape parameters a and b using the above estimator for the studies where the natural variation in $p_{(1)}$ is unknown. See Table 7.1 for estimates of $p_{(1)}$, a and b.

6.5.2.3 Predicting R^2 statistics for index-calibration experiments, from detection probability estimates from large-scale tiger distribution studies

We compute the estimates of average detection probability (Average $p_{(1)}$, see Table 7.1) using the variance decomposition approach (Link & Nichols, 1994) as used to obtain the true variance of abundance earlier. We arrive at the $\hat{V}$, $\hat{Q}$, $\hat{\hat{N}}$ (from Table 6.2) from IC-Jhala and IC-Karanth, to predict the R^2 statistics based on the two theoretical models we have proposed. Table 6.3 reports results of these predictions.

These predictions demonstrate the futility (as evidenced by extremely low R^2) of such index-calibration exercises as a means to predict densities at large spatial scales.

6.5.2.4 Assessing predictions of R^2 statistics from our proposed models against empirical estimates of R^2 from known tiger index-calibration exercises

When a simple linear regression was conducted to assess the predictive power of encounter rates of tiger signs with mean estimates of tiger densities (without accounting for sampling variation) in IC-Karanth, we obtained an extremely poor fit ($\hat{R^2} = 0.0004$) with an insignificant regression slope of 0.091(SE=1.98). Figure 6.5 illustrates the very poor quality of this fit, suggesting a large extent of heterogeneity in $p_{(1)}$. The empirical estimate of R^2 in IC-Karanth was anomalously low (Z score = -14.15, P value ¡ 0.00001), suggesting that there exists a pronounced variation in $p_{(1)}$, much larger than the extent of $c_v = 1$ based on our predictions using information from Karanth *et al.* (2011b). This is not surprising, because the variation we did consider is only indicative of the extent of variation of $p_{(1)}$ within a single

landscape. However, the heterogeneity in $p_{(1)}$ across more than a single landscape is potentially much greater. Table 6.3 provides the comparison between the estimated R^2 from our theoretical predictions and the actual fit. Figure 6.5 shows the best fit regression line from this index-calibration experiment. In stark contrast to this result, the empirical fit of R^2 in the Jhala *et al.* (2011b) index-calibration experiment is anomalously high ($\hat{R^2} = 0.95$) relative to our theoretical predictions.

6.6 Discussion

It is a basic goal in macroecology and global conservation, to assess changes in abundance of a species over large spatial and geographic scales. Knape & de Valpine (2012) demonstrate how uncertainty in abundance estimates changes, drastically, conclusions about fundamental ecological processes such as density-dependence in long term population dynamics studies. Therefore, it is undesirable to add another level of sampling uncertainty, with abundance indices, unless they contribute valuable information on spatial and temporal variation in abundance.

In this study, we have statistically revealed, for the first time, the exact roles of key sampling processes (imperfect detection p, spatial heterogeneity in sampling V and sampling uncertainties of population estimates Q) on the coefficient of determination R^2 generated by simple index-calibration experiments (Eberhardt & Simmons, 1987). Because we have now made clear that any combination among these three uncontrollable sampling processes can lead to high or low R^2 values, we provide a major step forward to address the question of how field studies can exhibit such a wide variation in degree of inferential success in such index-calibration experiments. In addition, we have also demonstrated through Model 2, that a fourth factor, the mean abundance μ of the species in concern, also plays an important role in determining the R^2.

Optimistic advocates of this simple index-calibration approach rely on the argument that under restrictive circumstances, p can be controlled for. However, even if this were true, they have little or no control against the impact of $\hat{V}$, $\hat{Q}$ and even μ (as in Model 2). This is because, in reality, much of the contribution to the parameter $\hat{V}$ comes from the natural variation in animal abundance across various sites or within species. Such natural variation is highly species- and system-specific and also related to spatial sampling (Thompson, 2002).

We have demonstrated a practical application of this theoretical advance to draw inference on recent large-scale assessments of tiger distribution and abundance across India. Our theoretical predictions of R^2 based on our models developed, combined with estimates of detection probability from two recent large-scale occupancy surveys of tigers (Karanth *et al.*, 2011b; Jhala *et al.*, 2011a) in India, clearly indicate that

the explanatory power of the relationship between tiger sign encounter rates and independent estimates of tiger density, is poor ($R^2 << 0.5$;Table 6.3). The empirical assessment of the R^2 statistic in IC-Karanth, conforms to the poor predictive power of the R^2 statistic based on Model 2 ($c_v = 1$) and suggests the basic futility of such an index-calibration exercise. In fact, results of the IC-Karanth experiment indicate that the heterogeneity around p is much larger than our assumption of $c_v = 1$ derived from Karanth *et al.* (2011b) study. This is not surprising considering that we only considered the variation within a single landscape and the actual variation could be much larger across various landscapes.

In stark contrast, however, the unusually high estimate of R^2 ($\hat{R^2} = 0.95$) in IC-Jhala (Jhala *et al.*, 2011b), stands out as an anomaly and contradicts the predictions derived from both the occupancy-based results of Karanth *et al.* (2011b) as well as Jhala *et al.* (2011a), since predictions indicate that even in the most ideal circumstance (when p^* is a constant), our model predicts a maximum R^2 value of 0.37(0.05), which is $<< 0.95$. In view of our study, such a result is possible only if reported estimates of V was much larger, p is much higher and unwarying, Q is very close to 1 and μ is also much lower. Given the similarity in the study species, study system and sampling methods, we expected a closer degree of conformance between the predicted and empirical estimates of the R^2 in the IC-Jhala study.

Our study, however, is limited in scope to the assumptions based on the two models developed for the R^2 statistic. It is also only relevant to direct index-calibration experiments that are described by simple general linear models. The study provides a framework to describe a goodness-of-fit statistic such as the R^2 in such a context and enables valid comparisons across studies adopting different methodologies (Stewart, 2010). It should also be noted here that although we generate the upper and lower bounds of R^2 based on minimum and maximum values of k in our theoretical derivations, we have not taken into account the heterogeneity of k that can exist. Further, we have not tested the efficacy of our models describing the R^2 statistic when subjected to violations of linearity in the general abundance index vs actual abundance relationships which could take non-linear forms. And, it is not straightforward to conceive of non-linear mechanistic models to describe this relationship easily. But there may be latent sampling covariances that could produce such unusual non-linear effects. For instance, in our case study, tiger landscapes in India typically consist of a system of protected areas set within larger patches of multiple-use forests with high human population densities. Since tigers extensively use man-made dirt roads and trails, in areas where human population densities are high and tiger densities are low, tiger tracks are often less detectable because of extensive movement of vehicles, people and cattle along the same track (Karanth *et al.*, 2011b). Such a situation may yield a positive covariation between p^* and N.

If this is the case, then we encounter two problems: 1) there is a lack of identifiability between p^* and N and 2) the abundance index versus actual abundance relationship is no more linear. This is because $\mathrm{E}(S|N) = kNp^*$ becomes now a function of N^2, if p^* covaries with N due to sampling covariance. Such a chance phenomenon may have occurred due to the selection of sites in IC-Jhala study.

It then leads to a more basic statistical question on how appropriate it is to use the R^2 statistic as a measure of goodness-of-fit for such non-linear models given that its unsuitability is too well-documented in statistical literature(Kvålseth, 1985; Willett & Singer, 1988; Nagelkerke, 1991; Juliano & Williams, 1987). Spiess & Neumeyer (2010) also go on to demonstrate that the R^2 statistic can also oddly produce high values for very bad or very good models in the presence of non-linearity. And, transforming non-linear data to linearize them is generally not recommended (Bolker, 2008; O'Hara & Kotze, 2010) because the two goals of simultaneously having to linearize a non-linear relationship and also produce an error structure of a defined statistical form (usually normal distribution with equal variance) is often incompatible. We surmise that the anomaly in the empirically derived R^2 estimate in the IC-Jhala study could possibly have occurred due to the above reasons.

In such contexts, further research is needed to investigate the performance of these R^2 statistics for various non-linear forms of the abundance index-abundance relationships. But given the poor performance of the R^2 statistics in non-linear systems in other disciplines (Spiess & Neumeyer, 2010), we would not be very optimistic of the result. For instance, different types of abundance indices could have different types of functional relationships with actual abundance. Lambin *et al.* (2000) show the difference between vole sign indices (VSI) and vole trapping indices. Hence, simply accumulating putative 'indices of abundance' or 'raw counts' of animals, or even trying to calibrate such indices directly, are poor methodological practices that will not be informative. Even in those rare situations where the exact value of abundance is known, thus eliminating the effect of the sampling factor, predictions from R^2 are shown to be unreliable (Diefenbach *et al.*, 1994).

Our study shows that such direct index-calibration exercises are worth considering only if p^* is very high and non-varying, V is very high, Q is very low and μ is also low. Meeting these assumptions are highly unrealistic in the field. We therefore strongly recommend against wasting limited conservation and scientific resources in such direct index-calibration exercises. We recommend only those models which explicitly distinguishes the observation process parameter (especially p) from the state process, for example, see Conroy *et al.* (2008) and more recently, Chandler & Royle (2013) which propose alternative ways of estimating abundance at large geographic scales by explicitly estimating detection probability within a joint model.

More recently, Efford & Fewster (2013), in their study of the skink *Oligosoma infrapunctatum*, demonstrate a possible way of estimating abundance at large spatial scales by using suitable covariates within a spatial capture-recapture framework. It is still relevant, however, that even with these newer techniques the sampling design should consist of a large V, a Q very close to 1 (using rigorous methods of estimating densities), and applying accurate covariates to describe the detection probability parameter to reduce its unaccounted for heterogeneity.

Although, this development was largely motivated by the tiger studies discussed here, the parameters that determine the R^2 in simple linear regression problems is relevant to a wide variety of macroecological studies from examinations of speciation or diversity patterns and processes (Signor & Lipps, 1982; Tarver *et al.*, 2007) to identifying disease hotspots (McClintock *et al.*, 2010; Lachish *et al.*, 2012), and more readily, to questions on long-term population dynamics (de Valpine & Hastings, 2002; De Valpine, 2003). Further, simple regression models are used widely across a range of disciplines (covering the entire spectrum from the sciences to social sciences) to assess linear correlations between two quantities of interest. The implications of our study may find relevance to similar sampling situations in other disciplines too.

6.7 Figures

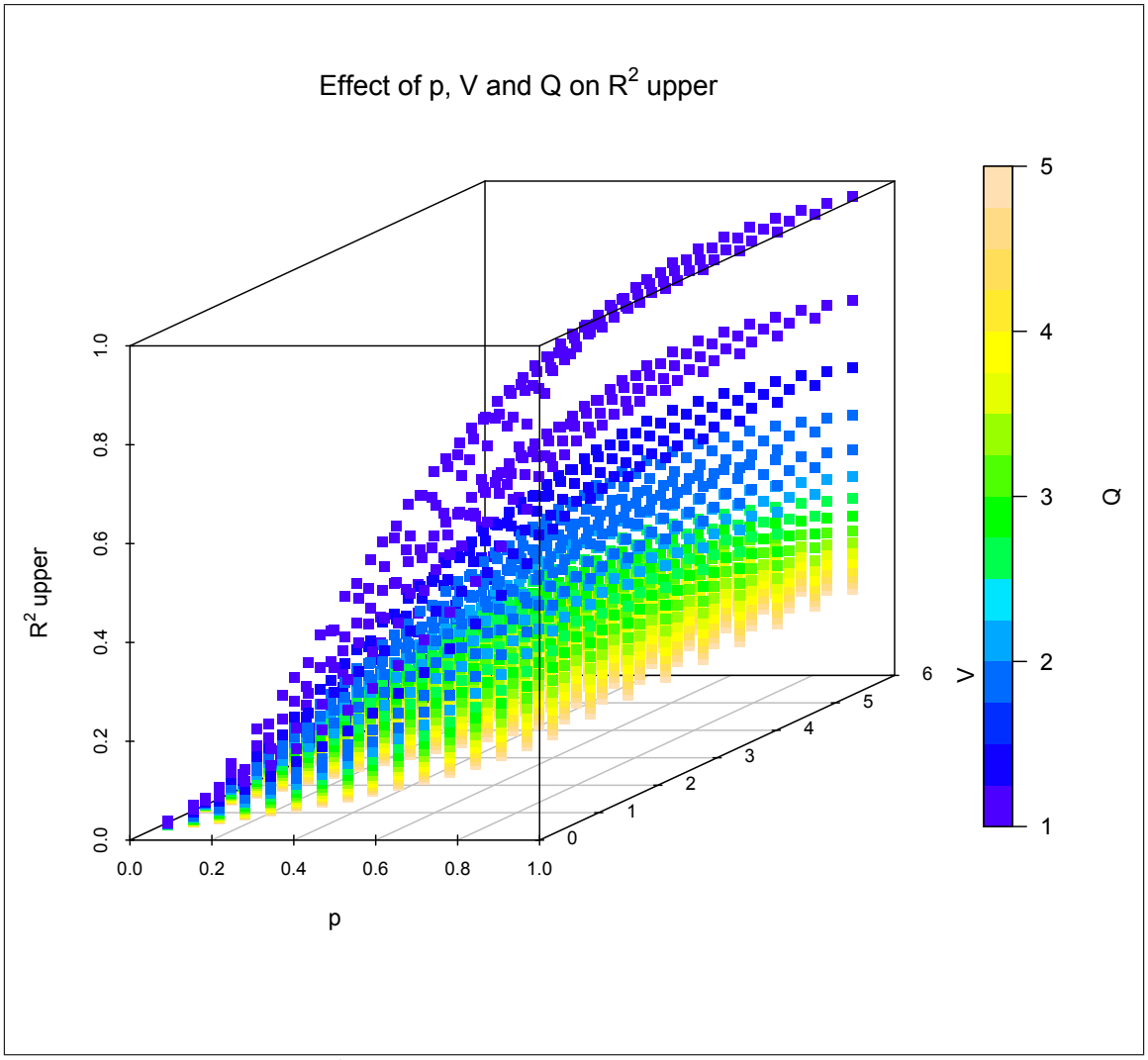


Figure 6.1: Plot of R^2 upper bound values when, p, V and Q are varied using Model 1: p^* = constant.

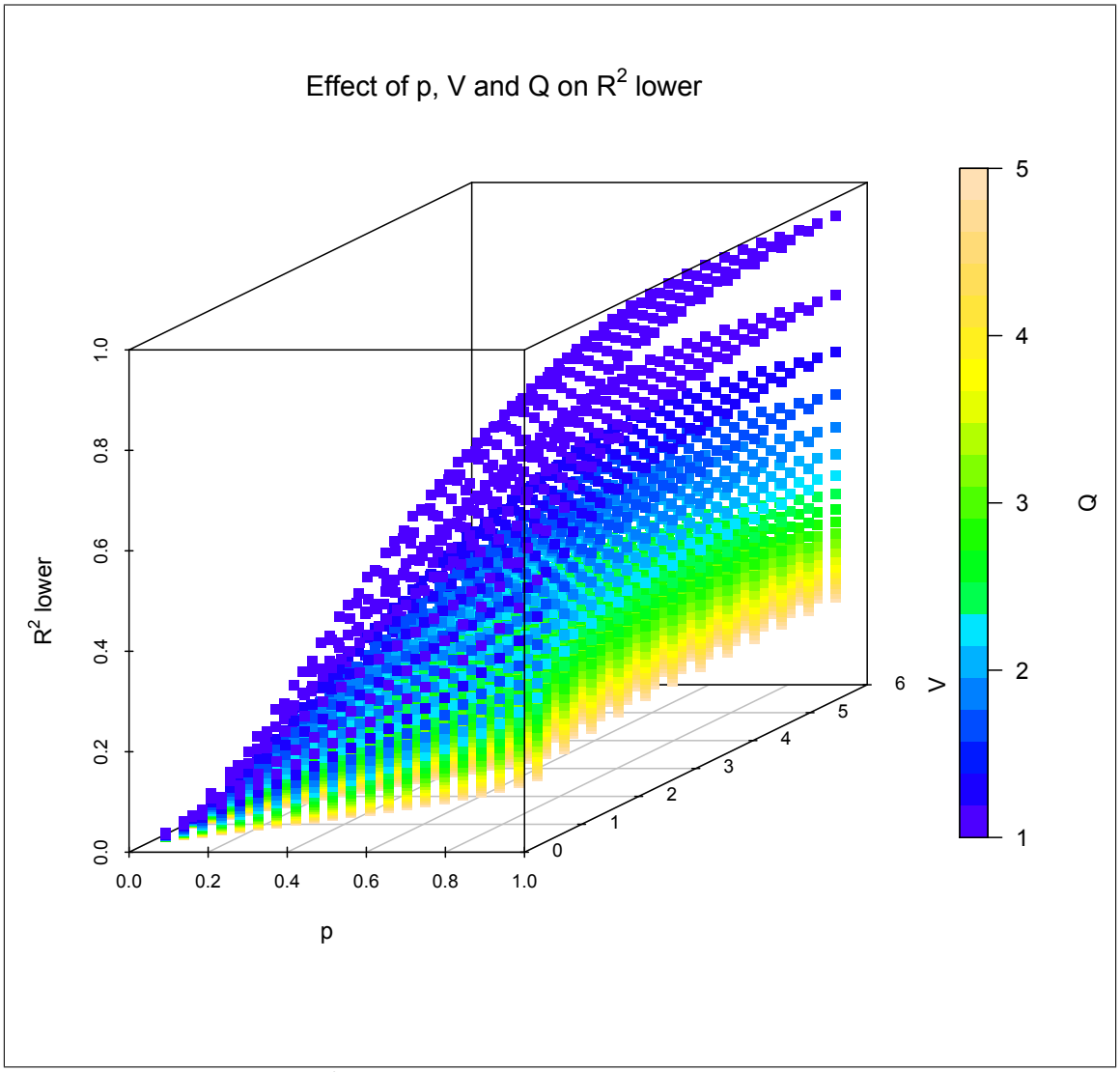


Figure 6.2: Plot of R^2 lower bound values when, p, V and Q are varied using Model 1: $p^* =$ constant.

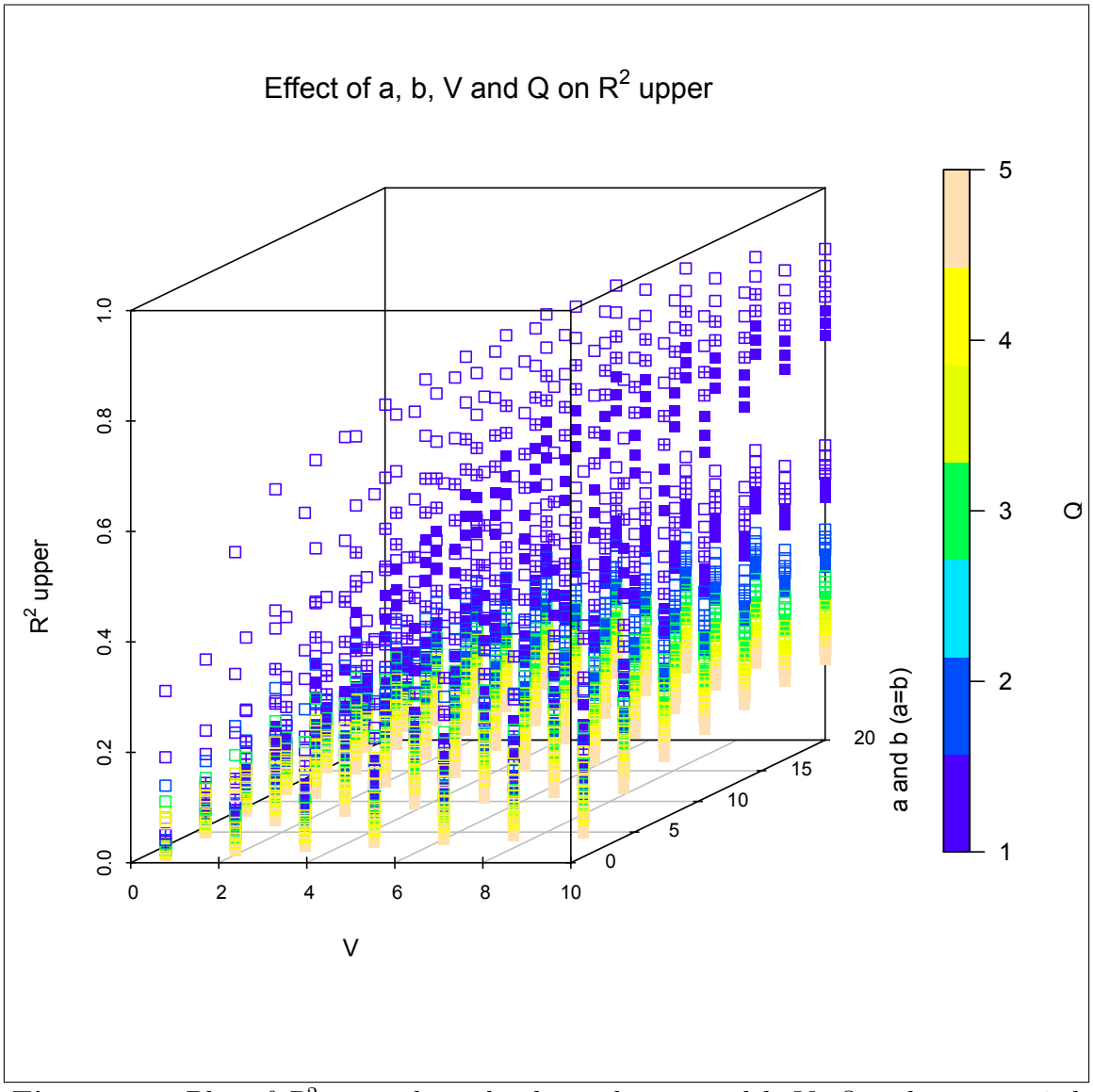


Figure 6.3: Plot of R^2 upper bound values when, a and b, V, Q and μ are varied using Model 2: $p^* =$ variable. Note, here μ is represented as three categories - $\mu < 33$ is represented by an empty square, $33 < \mu < 66$ is represented by a crossed-out square and $66 < \mu < 100$ is represented by a dark square

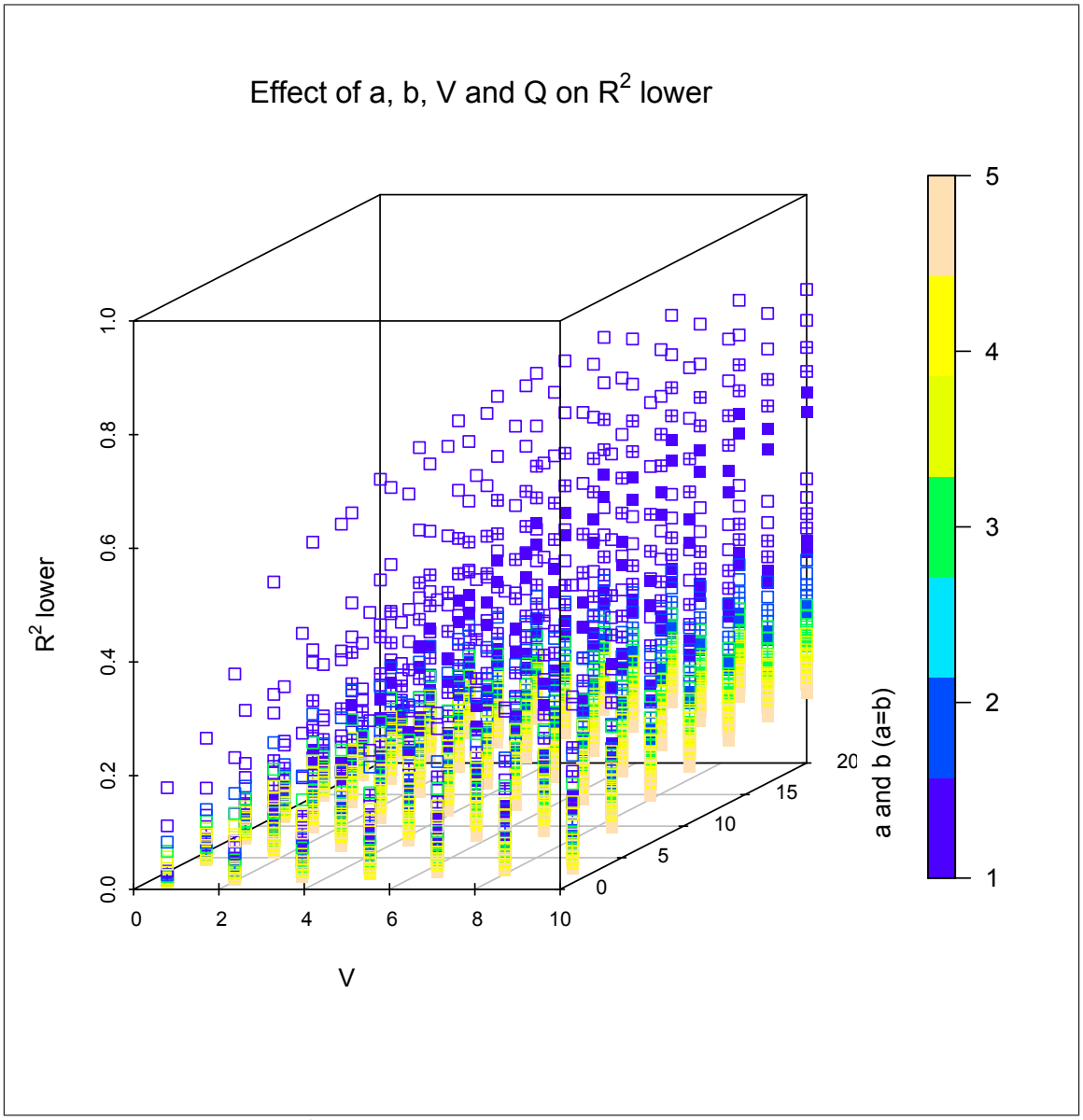


Figure 6.4: Plot of R^2 lower bound values when, a and b, V, Q and μ are varied using Model 2: $p^* =$ variable. Note, here μ is represented as three categories - $\mu < 33$ is represented by an empty square, $33 < \mu < 66$ is represented by a crossed-out square and $66 < \mu < 100$ is represented by a dark square.

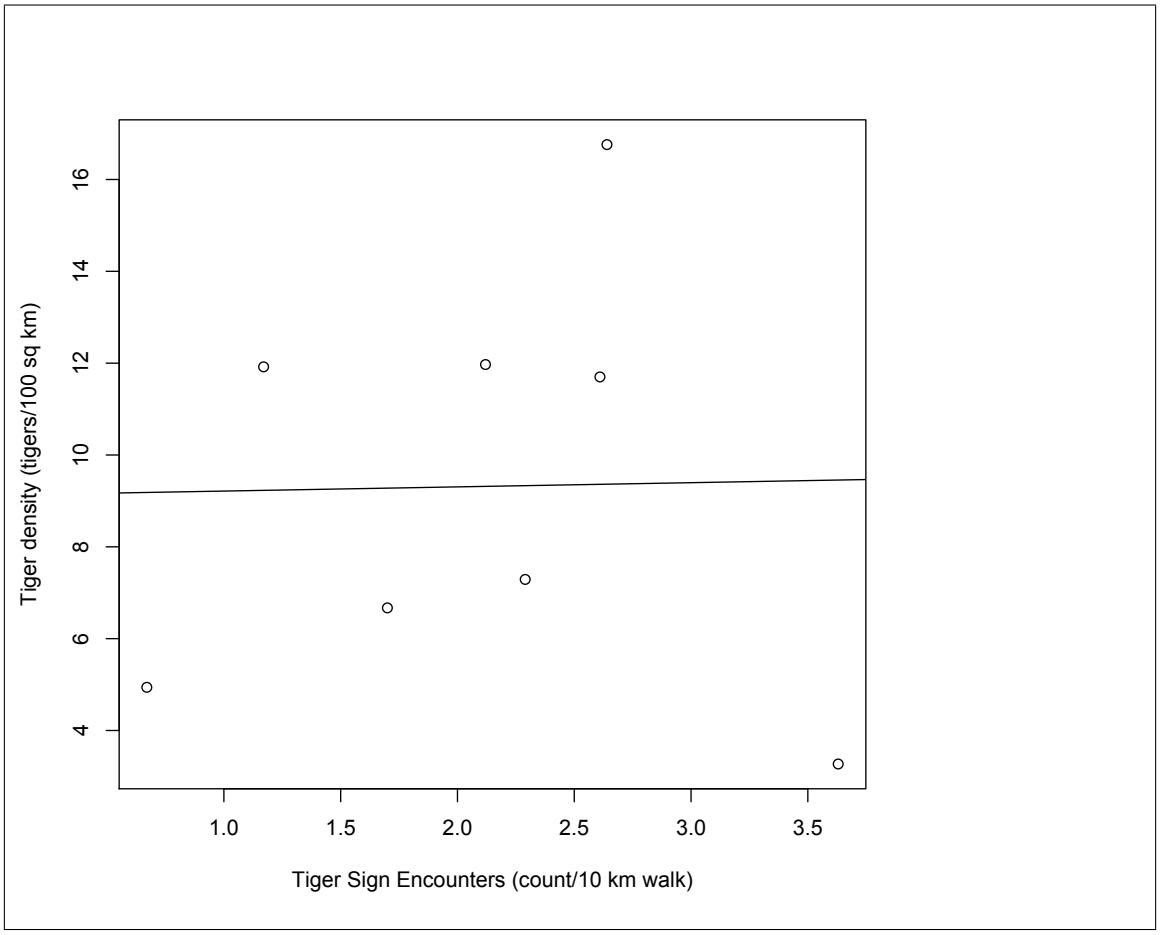


Figure 6.5: Plot of the relationship between tiger sign encounter rates and estimated tiger densities from 8 different sites in India in IC-Karanth. The line represents the estimated regression line.

6.8 Tables

Table 6.1: Estimates of $p_{(1)}$ and a and b in four tiger landscapes of India. Note: the '*' indicates that the estimated standard error for this parameter was assumed, since data were not available.

Landscape	Estimate of $p_{(1)}$ (p^* constant)	Estimates of a and b (p^* variable) $c_v = 1$	Source
Shivalik-Gangetic Plains	0.146 (0.0003)	0.708 (0.0007), 4.137 (0.015)	Jhala *et al.* (2011a)
Central India	0.08 (0.0001)*	0.84 (0.0003), 9.70 (0.022)	Jhala *et al.* (2011a)
South India	0.14 (0.0009)	0.717 (0.0017), 4.362 (0.0414)	Jhala *et al.* (2011a)
Malenad-Mysore Landscape	0.134 (0.031)	0.731 (0.0614), 4.714 (2.558)	Karanth *et al.* (2011b)
Average $p_{(1)}$	0.125 (0.0261)	0.75 (0.052), 5.25 (2.089)	

Table 6.2: Estimates of variance-to-mean ratio V, sampling factor Q and mean abundance from the two large-scale index-calibration experiments in India

Index-calibration study	Estimate of V	Estimate of Q	Estimate of μ
IC-Jhala	4.33	1.033	5.45
IC-Karanth	1.64	1.32	9.32

Table 6.3: Predicted and Empirical estimates of R^2 from two large-scale index-calibration experiments. We use Average $p_{(1)}$ from Table 7.1 for arriving at predicted R^2 values

Index-calibration study	Predicted R^2 estimate (p^* is constant)	Predicted R^2 estimate (p^* is variable)	Empirical estimate of R^2
IC-Jhala	0.355 (0.051)–0.37 (0.0553)	0.182 (0.0124)–0.208 (0.0186)	0.95
IC-Karanth	0.136 (0.0249)–0.144 (0.0277)	0.057 (0.004)–0.0668 (0.0064)	0.0004

6.9 Appendices

6.9.1 Derivation of R^2 when p^* is constant

$$
\begin{aligned}
\mathrm{Cor}(\hat{N}, S) &= \frac{\mathrm{Cov}(N, S)}{\sqrt{[\mathrm{Var}(N) + \mathrm{Var}(\epsilon)]\,\mathrm{Var}(S)}} \\
\mathrm{Cov}(N, S) &= \mathrm{E}(NS) - \mathrm{E}(N)\mathrm{E}(S) \\
&= \mathrm{E}\left[\mathrm{E}(NS|N)\right] - \mathrm{E}(N)\mathrm{E}\left[\mathrm{E}(S|N)\right] \\
&= \mathrm{E}\left[N\mathrm{E}(S|N)\right] - \mu\mathrm{E}\left[\mathrm{E}(S|N)\right] \\
&= \mathrm{E}(NkNp^*) - \mu\mathrm{E}(kNp^*) \\
&= kp^*\mathrm{E}(N^2) - kp^*\mu\mathrm{E}(N) \\
&= kp^*(\mu^2 + \sigma^2) - kp^*\mu^2 \\
\mathrm{Cov}(N, S) &= \sigma^2 kp^* \\
\mathrm{Var}(S) &= \mathrm{E}\left[\mathrm{Var}(S|N)\right] + \mathrm{Var}\left[\mathrm{E}(S|N)\right] \\
&= \mathrm{E}\left[kNp^*(1-p^*)\right] + \mathrm{Var}\left[kNp^*\right] \\
&= kp^*(1-p^*)\mathrm{E}(N) + k^2p^{*2}\mathrm{Var}(N) \\
&= kp^*(1-p^*)\mu + k^2p^{*2}\sigma^2 \\
&= kp^*(\mu - p^*\mu + kp^*\sigma^2) \\
\mathrm{Cor}(\hat{N}, S) &= \frac{\sigma^2 kp^*}{\sqrt{(\sigma^2 + \tau^2)\,kp^*(\mu - p^*\mu + kp^*\sigma^2)}}
\end{aligned}
$$

The term $\mathrm{Var}\left(\hat{N}\right) = \sigma^2 + \tau^2$ can be represented by the term $Q\sigma^2$, where Q is the ratio $\frac{\sigma^2+\tau^2}{\sigma^2}$ and $Q \geq 1$, which we refer to as the 'sampling factor'.

$$
\begin{aligned}
\mathrm{Cor}(\hat{N}, S) &= \frac{\sigma\sqrt{kp^*}}{\sqrt{Q\,(\mu - p^*\mu + kp^*\sigma^2)}} \\
\left[\mathrm{Cor}(\hat{N}, S)\right]^2 &= \frac{\sigma^2 kp^*}{Q\,(\mu - p^*\mu + kp^*\sigma^2)}
\end{aligned}
$$

6.9.2 Derivation for lower and upper bounds of R^2 when p^* is constant

From modern occupancy modeling studies (MacKenzie *et al.*, 2002; Hines *et al.*, 2010) it is possible to estimate p, the probability of detecting any individual animal *at least* once during a survey. So we have the relationship,
$p_k^* = 1 - (1-p)^{1/k}, 0 < p < 1, k = 1, 2, 3, \ldots$
and p_k^* can be defined as the value of p^* conditional on the number of occasions k.

Therefore,

$$\lim_{k\to\infty} p_k^* = 0.$$

Now

$$\begin{aligned}
kp_k^* &= k\left[1-(1-p)^{1/k}\right] \\
&= k\left[1-e^{\ln(1-p)/k}\right] \\
&= k\left[1-\left\{1+\frac{\ln(1-p)}{k}+\frac{1}{2}\frac{(\ln(1-p))^2}{k^2}+\frac{1}{3!}\frac{(\ln(1-p))^3}{k^3}+\dots\right\}\right] \\
&= -\ln(1-p)-\frac{\ln(1-p)}{k}\left\{\frac{1}{2}\frac{\ln(1-p)}{k}+\frac{1}{3!}\frac{(\ln(1-p))^2}{k}+\dots\right\}
\end{aligned}$$

$$\lim_{k\to\infty} kp_k^* = -\ln(1-p)$$

$$\begin{aligned}
R_\infty^2 &= \lim_{k\to\infty} R_k^2 \\
&= \lim_{k\to\infty}\left[\frac{\sigma^2 kp_k^*}{Q\left\{\mu(1-p_k^*)+\sigma^2 kp_k^*\right\}}\right] \\
&= \frac{-\sigma^2\ln(1-p)}{Q\left[\mu-\sigma^2\ln(1-p)\right]}
\end{aligned}$$

If $\sigma^2 = V\mu$, where V is the variance-to-mean ratio, then

$$
\begin{aligned}
R_\infty^2 &= \frac{-V\ln(1-p)}{Q\left[1-V\ln(1-p)\right]} \\
&= \frac{V\ln\left(\frac{1}{1-p}\right)}{Q\left[1+V\ln\left(\frac{1}{1-p}\right)\right]} \\
\frac{1}{R_\infty^2} &= \frac{Q\left[1+V\ln\left(\frac{1}{1-p}\right)\right]}{V\ln\left(\frac{1}{1-p}\right)} \\
\frac{1}{R_\infty^2} - Q &= \frac{Q}{V\ln\left(\frac{1}{1-p}\right)} = \frac{1-QR_\infty^2}{R_\infty^2} \\
V\ln\left(\frac{1}{1-p}\right) &= \frac{QR_\infty^2}{1-QR_\infty^2} \\
\ln\left(\frac{1}{1-p}\right) &= \frac{QR_\infty^2}{V\left(1-QR_\infty^2\right)} \\
\frac{1}{1-p} &= \exp\left[\frac{QR_\infty^2}{V\left(1-QR_\infty^2\right)}\right] \\
&= 1+\left[\frac{QR_\infty^2}{V\left(1-QR_\infty^2\right)}\right]+\frac{1}{2}\left[\frac{QR_\infty^2}{V\left(1-QR_\infty^2\right)}\right]^2+\frac{1}{6}\left[\frac{QR_\infty^2}{V\left(1-QR_\infty^2\right)}\right]^3+\dots \\
&> 1+\frac{QR_\infty^2}{V\left(1-QR_\infty^2\right)} \\
\frac{QR_\infty^2}{V\left(1-QR_\infty^2\right)} &< \frac{p}{1-p} \\
\frac{QR_\infty^2}{1-QR_\infty^2} &< \frac{Vp}{1-p} \\
\frac{1-QR_\infty^2}{QR_\infty^2} &> \frac{1-p}{Vp} \\
\frac{1}{R_\infty^2} &> \frac{Q(1-p)}{Vp}+Q \\
\frac{1}{R_\infty^2} &> \frac{Q(1-p)+QVp}{Vp} \\
R_\infty^2 &< \frac{Vp}{QVp+Q(1-p)} \\
R_\infty^2 &< \frac{Vp}{Q\left[Vp+(1-p)\right]}
\end{aligned}
$$

So, for any given value for k, p, V and Q the following relation must hold true:

$$
\frac{V\ln\left(\frac{1}{1-p}\right)}{Q\left[1+V\ln\left(\frac{1}{1-p}\right)\right]} < R^2 \leq \frac{Vp}{Q\left[Vp+(1-p)\right]}
$$

6.9.3 R code for predicting R^2 when p^* is a constant

```
####### Appendix 6.8.3: R code to generate upper and lower
    bounds of R^2 when p* is a constant #######

library(rgl)
library(fields)
###### Set values for variables ######
### Create divisions for the graph ###

divs<-20

p<-c(seq(from = 0.02, to=0.96, by=(0.96-0.02)/divs))
V<-c(seq(from = 0.5, to=10, by=(10-0.5)/divs))
Q<-c(seq(from = 1, to=5, by=(5-1)/divs))

# Create a matrix of combinations of p, V and Q

arr.trues <- array(TRUE, c(length(p), length(V), length(Q)))
idx <- which(arr.trues, arr.ind=TRUE)

pVQmatrix<-matrix(data=NA, nrow=length(p)*length(Q)*length(V
    ), ncol=5)

pVQmatrix[,1] <- p[idx[,1]]
pVQmatrix[,2] <- V[idx[,2]]
pVQmatrix[,3] <- Q[idx[,3]]

### Calculation of R2 Upper bounds ###

pVQmatrix[,4] <- (pVQmatrix[,2]*pVQmatrix[,1])/(pVQmatrix
    [,3]*((pVQmatrix[,2]*pVQmatrix[,1])+(1-pVQmatrix[,1])))

### Calculation of R2 Lower bounds ###

pVQmatrix[,5] <- ((pVQmatrix[,2])*(log((1)/(1-pVQmatrix[,1])
    )))/((pVQmatrix[,3])*(1+((pVQmatrix[,2])*(log((1)/(1-
    pVQmatrix[,1])))))))

### Creating colors for 4th dimension ###

# library(plotrix)
library(scatterplot3d)

## Plot for p, V, and R2 upper and lower. Here, Q is
    represented by colors. ##
# colours <- color.scale(pVQmatrix[,3], c(0,1,1), c(1,1,0)
    ,0)
# plot3d(pVQmatrix[,1], pVQmatrix[,2], pVQmatrix[,4], col=
```

```
    colours)
# old.par <- par(mfrow=c(1, 2))

nbcol <- topo.colors(divs+1)
# standardize pVQmatrix[,3] to be from 1 to divs+1

zcol <- ((pVQmatrix[,3]-min(pVQmatrix[,3]))/(max(pVQmatrix
    [,3])-min(pVQmatrix[,3])))*divs+1

###### R2 upper graph #######
R2uppergraph <- scatterplot3d(pVQmatrix[,1], pVQmatrix[,2],
    pVQmatrix[,4], color=nbcol[zcol], main=expression(paste("
    Effect of p, V and Q on ", R^2, " upper")), xlab=
    expression("p"), ylab= expression("V"),zlab= expression(
    paste(R^2," upper")), type="p", angle=25, mar=c(5, 3, 5,
    7)+0.1)
par(mar=c(5, 4, 4, 2) + 0.1)
image.plot( legend.only=TRUE, zlim= c(min(pVQmatrix[,3]),
    max(pVQmatrix[,3])), nlevel=(divs+1), col=topo.colors(
    divs+1))

###### R2 lower graph #######
R2uppergraph <- scatterplot3d(pVQmatrix[,1], pVQmatrix[,2],
    pVQmatrix[,5], color=nbcol[zcol], main=expression(paste("
    Effect of p, V and Q on ", R^2, " lower")), xlab=
    expression("p"), ylab= expression("V"),zlab= expression(
    paste(R^2," lower")), type="p", angle=25, mar=c(5, 3, 5,
    7)+0.1)
par(mar=c(5, 4, 4, 2) + 0.1)
image.plot( legend.only=TRUE, zlim= c(min(pVQmatrix[,3]),
    max(pVQmatrix[,3])), nlevel=(divs+1), col=topo.colors(
    divs+1))


# R2M1Upexplore<-(V[2]*p)/(Q[3]*((V[2]*p)+(1-p)))
# p<-c(0.02, 0.05, 0.1, 0.2, 0.4, 0.6, 0.7, 0.8, 0.9)
# V<-c(0.5, 0.75, 1, 2, 4, 8, 10)
# Q<-c(1, 1.25, 1.5, 2, 3, 6, 10)

# R^2 values for upper and lower bounds of p keeping V and Q
    constant
# R2M1Upexplore<-(V[2]*p)/(Q[3]*((V[2]*p)+(1-p)))
# R2M1Lpexplore<-((V[2])*(log((1)/(1-p))))/((Q[3])*(1+((V
    [2])*(log((1)/(1-p))))))

# R^2 values for upper and lower bounds of V keeping p and Q
    constant
# R2M1UVexplore<-(V*p[5])/(Q[3]*((V*p[5])+(1-p[5])))
#R2M1LVexplore<-((V)*(log((1)/(1-p[5]))))/((Q[3])*(1+((V)*(
```

```
    log((1)/(1-p[5])))))

# R^2 values for upper and lower bounds of Q keeping p and V
    constant
# R2M1UQexplore<-(V[2]*p[5])/(Q*((V[2]*p[5])+(1-p[5])))
#R2M1LQexplore<-((V[2])*(log((1)/(1-p[5]))))/((Q)*(1+((V[2])
    *(log((1)/(1-p[5]))))))

# Generating graphs for Model 1: p* is constant
# old.par <- par(mfrow=c(1, 3))

# Exploring p keeping V and Q constant
# plot(0,0,xlim=c(0,1),ylim=c(0,1),type="n", xlab=expression
    (p), ylab=expression(R^{2}~value))
# cl<-rainbow(2)
# lines(p,R2M1Upexplore,col=cl[1],type='b')
#lines(p,R2M1Lpexplore,col=cl[2],type='b')
# legend("topleft", legend = c(expression(R^{2}~upper~bound)
    ,expression(R^{2}~lower~bound)), col=cl[1:2], pch=1)

# Exploring V keeping p and Q constant
# plot(0,0,xlim=c(0,10),ylim=c(0,1),type="n", xlab=
    expression(V), ylab=expression(R^{2}~value))
# cl<-rainbow(2)
# lines(V,R2M1UVexplore,col=cl[1],type='b')
# lines(V,R2M1LVexplore,col=cl[2],type='b')
#legend("topleft", legend = c(expression(R^{2}~upper~bound),
    expression(R^{2}~lower~bound)), col=cl[1:2], pch=1)

# Exploring Q keeping p and V constant
# plot(0,0,xlim=c(1,10),ylim=c(0,1),type="n", xlab=
    expression(Q), ylab=expression(R^{2}~value))
# cl<-rainbow(2)
# lines(Q,R2M1UQexplore,col=cl[1],type='b')
# lines(Q,R2M1LQexplore,col=cl[2],type='b')
# legend("topleft", legend = c(expression(R^{2}~upper~bound)
    ,expression(R^{2}~lower~bound)), col=cl[1:2], pch=1)

# par(old.par)
```

6.9.4 Derivation for R^2 when p^* is variable

S represents the binomial count of animal encounters in a trapping device that is conditional on a *fixed* p^*. However, in most realistic ecological settings, p^* is likely to be variable. If,

$$p_k^* \sim \text{Beta}\left[\alpha_k, \beta_k\right]$$

then, the animal encounters S can be modeled as:

$$[S|N,k,\alpha_k,\beta_k] \sim \text{Beta} - \text{binomial}\left[kN,\alpha_k,\beta_k\right]$$

We can then go about determining $\text{Cor}\left(\hat{N},S\right)$ as follows:

$$\begin{aligned}
\text{Cov}(N,S) &= \text{E}(NS) - \text{E}(N)\text{E}(S) \\
&= \text{E}\left[\text{E}(NS|N)\right] - \text{E}(N)\text{E}\left[\text{E}(S|N)\right] \\
&= \text{E}\left[N\text{E}(S|N)\right] - \text{E}(N)\text{E}\left[\text{E}(S|N)\right] \\
&= \text{E}\left[NkN\left(\frac{\alpha_k}{\alpha_k+\beta_k}\right)\right] - \text{E}(N)\text{E}\left[kN\left(\frac{\alpha_k}{\alpha_k+\beta_k}\right)\right] \\
&= \frac{k\alpha_k}{\alpha_k+\beta_k}\text{E}\left(N^2\right) - \text{E}(N)\frac{k\alpha_k}{\alpha_k+\beta_k}\text{E}(N) \\
&= \frac{k\alpha_k}{\alpha_k+\beta_k}\left(\mu^2+\sigma^2\right) - \frac{k\alpha_k}{\alpha_k+\beta_k}\mu^2 \\
\text{Cov}(N,S) &= \sigma^2\frac{k\alpha_k}{\alpha_k+\beta_k}
\end{aligned}$$

$$\begin{aligned}
\text{Var}(S) &= \text{Var}\left[\text{E}(S|N)\right] + \text{E}\left[\text{Var}(S|N)\right] \\
&= \text{Var}\left[\frac{kN\alpha_k}{\alpha_k+\beta_k}\right] + \text{E}\left[\frac{kN\alpha_k\beta_k(\alpha_k+\beta_k+kN)}{(\alpha_k+\beta_k)^2(\alpha_k+\beta_k+1)}\right] \\
&= \frac{k^2\alpha_k^2}{(\alpha_k+\beta_k)^2}\text{Var}(N) + \text{E}\left[\frac{kN\alpha_k^2\beta_k + kN\alpha_k\beta_k^2 + k^2N^2\alpha_k\beta_k}{(\alpha_k+\beta_k)^2(\alpha_k+\beta_k+1)}\right] \\
&= \frac{k^2\alpha_k^2}{(\alpha_k+\beta_k)^2}\text{Var}(N) + \\
&\frac{1}{(\alpha_k+\beta_k)^2(\alpha_k+\beta_k+1)}\left[k\alpha_k^2\beta_k\text{E}(N) + k\alpha_k\beta_k^2\text{E}(N) + k^2\alpha_k\beta_k\text{E}\left(N^2\right)\right] \\
&= \frac{k^2\alpha_k^2}{(\alpha_k+\beta_k)^2}\sigma^2 + \frac{k\alpha_k^2\beta_k\mu + k\alpha_k\beta_k^2\mu + k^2\alpha_k\beta_k\left(\mu^2+\sigma^2\right)}{(\alpha_k+\beta_k)^2(\alpha_k+\beta_k+1)} \\
\text{Var}(S) &= \frac{\sigma^2k^2\alpha_k^2(\alpha_k+\beta_k+1) + k\alpha_k^2\beta_k\mu + k\alpha_k\beta_k^2\mu + k^2\alpha_k\beta_k\mu^2 + k^2\alpha_k\beta_k\sigma^2}{(\alpha_k+\beta_k)^2(\alpha_k+\beta_k+1)}
\end{aligned}$$

$$\begin{aligned}
\text{Cor}(\hat{N},S) &= \frac{\sigma^2k\left(\frac{\alpha_k}{\alpha_k+\beta_k}\right)}{\sqrt{Q\sigma^2\left[\frac{\sigma^2k^2\alpha_k^2(\alpha_k+\beta_k+1)+k\alpha_k^2\beta_k\mu+k\alpha_k\beta_k^2\mu+k^2\alpha_k\beta_k\mu^2+k^2\alpha_k\beta_k\sigma^2}{(\alpha_k+\beta_k)^2(\alpha_k+\beta_k+1)}\right]}} \\
&= \frac{\sigma k\alpha_k\sqrt{\alpha_k+\beta_k+1}}{\sqrt{Q\left[\sigma^2k^2\alpha_k^2(\alpha_k+\beta_k+1) + k\alpha_k^2\beta_k\mu + k\alpha_k\beta_k^2\mu + k^2\alpha_k\beta_k\mu^2 + k^2\alpha_k\beta_k\sigma^2\right]}} \\
\text{Cor}(\hat{N},S) &= \frac{\sigma\sqrt{k\alpha_k}\sqrt{\alpha_k+\beta_k+1}}{\sqrt{Q\left[\sigma^2k\alpha_k(\alpha_k+\beta_k+1) + \alpha_k\beta_k\mu + \beta_k^2\mu + k\beta_k\mu^2 + k\beta_k\sigma^2\right]}} \\
R^2 &= \frac{\sigma^2k\alpha_k(\alpha_k+\beta_k+1)}{Q\left[\sigma^2k\alpha_k(\alpha_k+\beta_k+1) + \alpha_k\beta_k\mu + \beta_k^2\mu + k\beta_k\mu^2 + k\beta_k\sigma^2\right]}
\end{aligned}$$

6.9.5 Derivation for lower and upper bounds for R^2 when p^* is variable

We do have the relation:

$$p_k^* = 1 - (1-p)^{1/k} \sim \text{Beta}\left[\alpha_k, \beta_k\right]$$

Suppose,

$$[p|a,b] \sim \text{Beta}\left[a,b\right]$$

It is obvious that when $k=1$, $p_k^* = p$, so $\alpha_k = a$ and $\beta_k = b$. So

$$\begin{aligned}
\text{E}(kp_k^*) &= \text{E}(p) = \frac{a}{a+b} \\
\text{Var}(kp_k^*) &= \text{Var}(p) = \frac{ab}{(a+b)^2\,(a+b+1)}
\end{aligned}$$

However, when k takes a very large value, we would like:

$$\begin{aligned}
\frac{k\alpha_k}{\alpha_k+\beta_k} &= \text{E}\left(kp_k^*\right) \to \text{E}\left[-\ln(1-p)\right] = V_1 \\
\frac{k^2\alpha_k\beta_k}{(\alpha_k+\beta_k)^2\,(\alpha_k+\beta_k+1)} &= \text{Var}\left(kp_k^*\right) \to \text{Var}\left[-\ln(1-p)\right] = V_2
\end{aligned}$$

We shall set $\alpha_k = \alpha$ and $\beta_k = k(\alpha+\beta)$, for some $\alpha > 0$ and $\beta > 0$. Then

$$\begin{aligned}
\lim_{k\to\infty} \frac{k\alpha_k}{\alpha_k+\beta_k} &= \lim_{k\to\infty} \frac{k\alpha}{\alpha+k(\alpha+\beta)} = \frac{\alpha}{\alpha+\beta} = V_1 \\
\lim_{k\to\infty} \frac{k^2\alpha_k\beta_k}{(\alpha_k+\beta_k)^2\,(\alpha_k+\beta_k+1)} &= \lim_{k\to\infty} \frac{k^3\alpha(\alpha+\beta)}{[\alpha+k(\alpha+\beta)]^2\,[\alpha+k(\alpha+\beta)+1]} = \frac{\alpha}{(\alpha+\beta)^2} = V_2
\end{aligned}$$

Solving for α and β in terms of V_1 and V_2 gives:

$$\begin{aligned}
\alpha &= \frac{V_1^2}{V_2} > 0 \\
\beta &= \frac{V_1(1-V_1)}{V_2}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\lim_{k\to\infty} R_k^2 &= \lim_{k\to\infty} \frac{\sigma^2 k\alpha_k(\alpha_k+\beta_k+1)}{Q\left[\sigma^2 k\alpha_k(\alpha_k+\beta_k+1)+\alpha_k\beta_k\mu+\beta_k^2\mu+k\beta_k\mu^2+k\beta_k\sigma^2\right]} \\
&= \lim_{k\to\infty} \frac{\sigma^2 k\alpha\left[\alpha+k(\alpha+\beta)+1\right]}{Q\left\{\sigma^2 k\alpha\left[\alpha+k(\alpha+\beta)+1\right]+k\alpha(\alpha+\beta)\mu+k^2(\alpha+\beta)^2\mu+k^2(\alpha+\beta)(\mu^2+\sigma^2)\right\}} \\
&= \frac{\sigma^2\alpha(\alpha+\beta)}{Q\left[\sigma^2\alpha(\alpha+\beta)+(\alpha+\beta)^2\mu+(\alpha+\beta)(\mu^2+\sigma^2)\right]} \\
&= \frac{\sigma^2\alpha}{Q\left[\sigma^2\alpha+(\alpha+\beta)\mu+\mu^2+\sigma^2\right]} \\
R_\infty^2 &= \frac{\sigma^2 V_1}{Q\left[\sigma^2 V_1+\mu+(\mu^2+\sigma^2)\frac{V_2}{V_1}\right]}
\end{aligned}$$

If $\sigma^2 = V\mu$, where V is the variance-to-mean ratio,

$$R_\infty^2 = \frac{VV_1}{Q\left[VV_1+1+(\mu+V)\frac{V_2}{V_1}\right]} < \frac{VV_1}{Q\left(VV_1+1\right)} < \frac{VV_1}{Q}$$

And if $V=1$, then $R_\infty^2 < \frac{V_1}{Q}$. More generally, we have the relation:

$$R_\infty^2 < R^2 \leq R_1^2$$

where, R_1^2 is the R^2 when $k=1$.

Determining the quantities V_1 and V_2 in practice Now consider the distribution of p:

$$[p|a,b] \sim \text{Beta}\left[a,b\right]$$

To determine V_1 and V_2 we can apply the Taylor "first-order" approximation:

$$\begin{aligned}
y = g(p) &= -\ln(1-p) \\
V_1 = \mathrm{E}(y) = g\left[\mathrm{E}\left(p\right)\right] &= \mathrm{E}\left[-\ln(1-p)\right] = -\ln\left(1-\frac{a}{a+b}\right) = \ln\left(\frac{a+b}{b}\right) \\
g'(p) &= \frac{1}{1-p} \\
g'\left[\mathrm{E}(p)\right] &= \frac{a+b}{b} \\
V_2 = \mathrm{Var}(y) = \left(g'\left[\mathrm{E}(p)\right]\right)^2\mathrm{Var}(p) &= \left(\frac{a+b}{b}\right)^2\frac{ab}{(a+b)^2(a+b+1)} = \frac{a}{b(a+b+1)}
\end{aligned}$$

So we have the lower and upper limits for R^2 defined as:

$$\frac{V\left[\ln\left(\frac{a+b}{b}\right)\right]^2}{Q\left(V\left[\ln\left(\frac{a+b}{b}\right)\right]^2+\ln\left(\frac{a+b}{b}\right)+(\mu+V)\frac{a}{b(a+b+1)}\right)}<R^2$$

$$R^2<\frac{aV(a+b+1)}{Q\left[aV(a+b+1)+ab+b^2+(\mu+V)b\right]}$$

6.9.6 R code for predicting R^2 when p^* is variable

```
% \verbatiminput{Appendix1-6R2statisticsPstarVariable.R}
####### Appendix 6.8.6: R code to generate upper and lower
    bounds of R^2 when p* is variable #######

### Call the required libraries ###

library(rgl)
library(fields)
library(scatterplot3d)

# Create divisions for the graphs #

divs <- 6

## Set parameter ranges for the graphs ##

a <- c(seq(from = 1, to = 20, by = (20-1)/divs))
b <- c(seq(from = 1, to = 20, by = (20-1)/divs))
V <- c(seq(from = 0.5, to=10, by=(10-0.5)/divs))
Q <- c(seq(from = 1, to=5, by=(5-1)/divs))
mu <- c(seq(from = 1, to = 100, by = (100-1)/divs))

# a<-c(20, 15, 10, 5, 4, 3, 2, 1)
# b<-c(20, 15, 10, 5, 4, 3, 2, 1)
# V<-c(0.5, 0.75, 1, 2, 4, 8, 10)
# Q<-c(1, 1.25, 1.5, 2, 3, 6, 10)
# mu<-c(1, 2, 3, 5, 10, 15, 30, 50, 100, 200, 500, 1000)

## Create a matrix of combinations of a, b, V, Q and mu
arr.trues <- array(TRUE, c(length(a), length(V), length(Q),
    length(mu)))
idx <- which(arr.trues, arr.ind=TRUE)

abVQmumatrix <- matrix(data=NA, nrow = length(a)*length(V)*
    length(Q)*length(mu), ncol=8)

abVQmumatrix[,1] <- a[idx[,1]]
abVQmumatrix[,2] <- b[idx[,1]]
abVQmumatrix[,3] <- V[idx[,2]]
abVQmumatrix[,4] <- Q[idx[,3]]
abVQmumatrix[,5] <- mu[idx[,4]]

## Categorizing mu for graphing in 4 categories ##
for (i in 1:nrow(abVQmumatrix)){
        if(abVQmumatrix[i,5] < 33){
        abVQmumatrix[i,8] <- 0
}else{
```

```
        if(abVQmumatrix[i,5] < 66){
        abVQmumatrix[i,8] <- 12
}else{
    if(abVQmumatrix[i,5] < 100){
                        abVQmumatrix[i,8] <- 15
}
}
}
}
## Calculation of R^2 upper bounds ##

abVQmumatrix[,6] <- (abVQmumatrix[,1]*abVQmumatrix[,3]*(
    abVQmumatrix[,1]+abVQmumatrix[,2]+1))/(abVQmumatrix[,4]*
    ((abVQmumatrix[,1]*abVQmumatrix[,3]*(abVQmumatrix[,1]+
    abVQmumatrix[,2]+1))+(abVQmumatrix[,1]*abVQmumatrix[,2])
    +(abVQmumatrix[,2]^2)+((abVQmumatrix[,5]+abVQmumatrix
    [,3])*abVQmumatrix[,2]))))

## Calculation of R^2 lower bounds ##

abVQmumatrix[,7] <- (abVQmumatrix[,3]*((log((abVQmumatrix
    [,1]+abVQmumatrix[,2])/(abVQmumatrix[,2])))^2))/(
    abVQmumatrix[,4]*((abVQmumatrix[,3]*((log((abVQmumatrix
    [,1]+abVQmumatrix[,2])/(abVQmumatrix[,2])))^2))+(log((
    abVQmumatrix[,1]+abVQmumatrix[,2])/(abVQmumatrix[,2])))
    +((abVQmumatrix[,5]+abVQmumatrix[,3])*((abVQmumatrix[,1])
    /(abVQmumatrix[,2]*(abVQmumatrix[,1]+abVQmumatrix[,2]+1))
    ))))

### Creating colors for 4th dimension ###

nbcol <- topo.colors(divs+1)
# standardize pVQmatrix[,3] to be from 1 to divs+1

zcol <- ((abVQmumatrix[,4]-min(abVQmumatrix[,4]))/(max(
    abVQmumatrix[,4])-min(abVQmumatrix[,4])))*divs+1

###### R2 upper graph #######
R2uppergraph <- scatterplot3d(abVQmumatrix[,3], abVQmumatrix
    [,1], abVQmumatrix[,6], color=nbcol[zcol], main=
    expression(paste("Effect of a, b, V and Q on ", R^2, "
    upper")), xlab= expression("V"), ylab= expression("a and
    b (a=b)"),zlab= expression(paste(R^2," upper")), pch=
    abVQmumatrix[,8], type="p", angle=25, mar=c(5, 3, 5, 7)
    +0.1)
par(mar=c(5, 4, 4, 2) + 0.1)
image.plot( legend.only=TRUE, zlim= c(min(abVQmumatrix[,4]),
     max(abVQmumatrix[,4])), nlevel=(divs+1), col=topo.colors
    (divs+1))
```

```
###### R2 lower graph #######
R2lowergraph <- scatterplot3d(abVQmumatrix[,3], abVQmumatrix
    [,1], abVQmumatrix[,7], color=nbcol[zcol], main=
    expression(paste("Effect of a, b, V and Q on ", R^2, "
    lower")), xlab= expression("V"), ylab= expression("a and
    b (a=b)"),zlab= expression(paste(R^2," lower")), pch=
    abVQmumatrix[,8], type="p", angle=25, mar=c(5, 3, 5, 7)
    +0.1)
par(mar=c(5, 4, 4, 2) + 0.1)
image.plot( legend.only=TRUE, zlim= c(min(abVQmumatrix[,4]),
     max(abVQmumatrix[,4])), nlevel=(divs+1), col=topo.colors
    (divs+1))
```

6.9.7 R code for parametric bootstrapping to obtain standard errors for p(1) and calculations of predicted R^2 values

```
### Appendix 6.8.7 Parametric Bootstrapping to obtain
    standard errors for p(1) and calculations of predicted R
    ^2 values.

#### Generating bootstraps to get standard errors of p(1)
    for Malenad-Mysore Tiger Landscape ####
A <- 1.89 ## Is the area adjustment for cell size difference
theta0 <- 0.418
theta0adj.area <- theta0*1.89
theta0SE <- 0.079
ptmean <- 0.17
p1mean <- theta0adj.area*ptmean
amean <- (p1mean)*((((p1mean)*(1-p1mean))/(p1mean^2))-1)
bmean <- (1-p1mean)*((((p1mean)*(1-p1mean))/(p1mean^2))-1)

## Generate bootstraps
BS <- 10000
theta0bs <- c(rep(0,BS))
theta0adj.areabs <- c(rep(0,BS))
p1 <- c(rep(0,BS))
a <- c(rep(0,BS))
b <- c(rep(0,BS))

for (i in 1:BS){
        while (theta0bs[i]<=0 | theta0bs[i]>=1){
                theta0bs[i] <- rnorm(1, mean=theta0, sd=
                    theta0SE)
        }
        theta0adj.areabs[i] <- 1-((1-theta0bs[i])^A)
        p1[i] <- theta0adj.areabs[i]*ptmean
        a[i] <- (p1[i])*((((p1[i])*(1-p1[i]))/(p1[i]^2))-1)
        b[i] <- (1-p1[i])*((((p1[i])*(1-p1[i]))/(p1[i]^2))
            -1)

}

p1var <- sum((p1-p1mean)^2)/(BS-1)
p1SE <- sqrt(p1var)

avar <- sum((a-amean)^2)/(BS-1)
aSE <- sqrt(avar)

bvar <- sum((b-bmean)^2)/(BS-1)
```

```
bSE <- sqrt(bvar)

cat("p1 mean =", p1mean, "p1 SE =", p1SE)
cat("a mean =", amean, "a SE =", aSE)
cat("b mean =", bmean, "b SE =", bSE)

#
   ##########################################################################


### Obtain bootstrap estimates for other sets from Jhala et
    al (2010) ###


#### Shivalik Gangetic plains ###

p5mean <- 0.546
p5SE <- 0.001
m <- 5
p1mean <- 1-((1-p5mean)^(1/m))
amean <- (p1mean)*((((p1mean)*(1-p1mean))/(p1mean^2))-1)
bmean <- (1-p1mean)*((((p1mean)*(1-p1mean))/(p1mean^2))-1)

## Generate bootstraps
BS <- 10000
p5bs <- c(rep(0,BS))
p1 <- c(rep(0,BS))
a <- c(rep(0,BS))
b <- c(rep(0,BS))

for (i in 1:BS){
        while (p5bs[i]<=0 | p5bs[i]>=1){
                p5bs[i] <- rnorm(1, mean=p5mean, sd=p5SE)
        }
        p1[i] <- 1-((1-p5bs)^(1/m))
        a[i] <- (p1[i])*((((p1[i])*(1-p1[i]))/(p1[i]^2))-1)
        b[i] <- (1-p1[i])*((((p1[i])*(1-p1[i]))/(p1[i]^2))
            -1)

}

p1var <- sum((p1-p1mean)^2)/(BS-1)
p1SE <- sqrt(p1var)

avar <- sum((a-amean)^2)/(BS-1)
aSE <- sqrt(avar)

bvar <- sum((b-bmean)^2)/(BS-1)
bSE <- sqrt(bvar)
```

```
cat("p1 mean =", p1mean, "p1 SE =", p1SE)
cat("a mean =", amean, "a SE =", aSE)
cat("b mean =", bmean, "b SE =", bSE)

#### Central India ###

p5mean <- 0.34
p5SE <- 0.001 ## Assumed the SE, it is not reported
m <- 5
p1mean <- 1-((1-p5mean)^(1/m))
amean <- (p1mean)*((((p1mean)*(1-p1mean))/(p1mean^2))-1)
bmean <- (1-p1mean)*((((p1mean)*(1-p1mean))/(p1mean^2))-1)

## Generate bootstraps
BS <- 10000
p5bs <- c(rep(0,BS))
p1 <- c(rep(0,BS))
a <- c(rep(0,BS))
b <- c(rep(0,BS))

for (i in 1:BS){
        while (p5bs[i]<=0 | p5bs[i]>=1){
                p5bs[i] <- rnorm(1, mean=p5mean, sd=p5SE)
        }
        p1[i] <- 1-((1-p5bs)^(1/m))
        a[i] <- (p1[i])*((((p1[i])*(1-p1[i]))/(p1[i]^2))-1)
        b[i] <- (1-p1[i])*((((p1[i])*(1-p1[i]))/(p1[i]^2))
           -1)

}

p1var <- sum((p1-p1mean)^2)/(BS-1)
p1SE <- sqrt(p1var)

avar <- sum((a-amean)^2)/(BS-1)
aSE <- sqrt(avar)

bvar <- sum((b-bmean)^2)/(BS-1)
bSE <- sqrt(bvar)

cat("p1 mean =", p1mean, "p1 SE =", p1SE)
cat("a mean =", amean, "a SE =", aSE)
cat("b mean =", bmean, "b SE =", bSE)

#### South India ###

p5mean <- 0.533
p5SE <- 0.001
```

```
m <- 5
p1mean <- 1-((1-p5mean)^(1/m))
amean <- (p1mean)*((((p1mean)*(1-p1mean))/(p1mean^2))-1)
bmean <- (1-p1mean)*((((p1mean)*(1-p1mean))/(p1mean^2))-1)

## Generate bootstraps
BS <- 10000
p5bs <- c(rep(0,BS))
p1 <- c(rep(0,BS))
a <- c(rep(0,BS))
b <- c(rep(0,BS))

for (i in 1:BS){
        while (p5bs[i]<=0 | p5bs[i]>=1){
                p5bs[i] <- rnorm(1, mean=p5mean, sd=p5SE)
        }
        p1[i] <- 1-((1-p5bs)^(1/m))
        a[i] <- (p1[i])*((((p1[i])*(1-p1[i]))/(p1[i]^2))-1)
        b[i] <- (1-p1[i])*((((p1[i])*(1-p1[i]))/(p1[i]^2))
           -1)

}

p1var <- sum((p1-p1mean)^2)/(BS-1)
p1SE <- sqrt(p1var)

avar <- sum((a-amean)^2)/(BS-1)
aSE <- sqrt(avar)

bvar <- sum((b-bmean)^2)/(BS-1)
bSE <- sqrt(bvar)

cat("p1 mean =", p1mean, "p1 SE =", p1SE)
cat("a mean =", amean, "a SE =", aSE)
cat("b mean =", bmean, "b SE =", bSE)


########################## CALCULATION OF AVERAGE p1s
   ###################

p1hat.shivalik <- 0.146
p1hat.central <- 0.08
p1hat.south <- 0.14
p1hat.MMTL <- 0.134

p1SE.shivalik <- 0.0003
p1SE.central <- 0.0001
p1SE.south <- 0.0009
p1SE.MMTL <- 0.031
```

```
## Employ the variance decomposition method ##

p1hatmean <- (1/4)*(p1hat.shivalik+p1hat.central+p1hat.south
   +p1hat.MMTL)

TotVarp1 <- (1/(4-1))*(((p1hat.shivalik-p1hatmean)^2)+((
   p1hat.central-p1hatmean)^2)+((p1hat.south-p1hatmean)^2)
   +((p1hat.MMTL-p1hatmean)^2))

p1s2 <- (1/4)*((p1SE.shivalik^2)+(p1SE.central^2)+(p1SE.
   south^2)+(p1SE.MMTL^2))

TrueVarp1 <- TotVarp1 - p1s2
TrueSEp1 <- sqrt(TrueVarp1)

## Generate bootstraps to estimate other parameters
   including R2 ##

BS <- 10000
p1bs <- c(rep(0,BS))
a <- c(rep(0,BS))
b <- c(rep(0,BS))

R2LowerPconstJ <- c(rep(0,BS))
R2UpperPconstJ <- c(rep(0,BS))
R2LowerPconstK <- c(rep(0,BS))
R2UpperPconstK <- c(rep(0,BS))

R2LowerPvarJ <- c(rep(0,BS))
R2UpperPvarJ <- c(rep(0,BS))
R2LowerPvarK <- c(rep(0,BS))
R2UpperPvarK <- c(rep(0,BS))

a.avg <- (p1hatmean)*((((p1hatmean)*(1-p1hatmean))/(
   p1hatmean^2))-1)
b.avg <- (1-p1hatmean)*((((p1hatmean)*(1-p1hatmean))/(
   p1hatmean^2))-1)

### Abundance variances from Jhala et al
Vj <- 4.33
Qj <- 1.033
muj <- 5.45

### Abundance variances from Karanth et al
Vk <- 1.64
Qk <- 1.32
muk <- 9.32
```

```
### Predicted R-squared from IC - Jhala (p constant) ###

R2LowerPconstantJmean <- ((Vj)*(log((1)/(1-p1hatmean))))/((
    Qj)*(1+((Vj)*(log((1)/(1-p1hatmean))))))
R2UpperPconstantJmean <- (Vj*p1hatmean)/(Qj*((Vj*p1hatmean)
    +(1-p1hatmean)))

### Predicted R-squared from IC-Karanth (p constant) ###

R2LowerPconstantKmean <- ((Vk)*(log((1)/(1-p1hatmean))))/((
    Qk)*(1+((Vk)*(log((1)/(1-p1hatmean))))))
R2UpperPconstantKmean <- (Vk*p1hatmean)/(Qk*((Vk*p1hatmean)
    +(1-p1hatmean)))

### Predicted R-squared from IC - Jhala (p variable) ###

#R2LowerPvarJmean <- (a.avg*Vj*log((a.avg+b.avg)/(b.avg)))/(
    Qj*((a.avg*Vj*log((a.avg+b.avg)/(b.avg)))+(a.avg)+((muj+
    Vj)*(log((a.avg+b.avg)/(b.avg)))*(a.avg+b.avg)*(a.avg+b.
    avg+1)))) ---- Old calculation

R2LowerPvarJmean <- (Vj*((log((a.avg+b.avg)/(b.avg)))^2))/((
    Qj)*(((Vj)*((log((a.avg+b.avg)/(b.avg)))^2))+(log((a.avg+
    b.avg)/(b.avg)))+((muj+Vj)*((a.avg)/((b.avg)*(a.avg+b.avg
    +1))))))

R2UpperPvarJmean <- (a.avg*Vj*(a.avg+b.avg+1))/(Qj*((a.avg*
    Vj*(a.avg+b.avg+1))+(a.avg*b.avg)+(b.avg^2)+((muj+Vj)*b.
    avg)))

### Predicted R-squared from IC - Karanth (p variable) ###

#R2LowerPvarKmean <- (a.avg*Vk*log((a.avg+b.avg)/(b.avg)))/(
    Qk*((a.avg*Vk*log((a.avg+b.avg)/(b.avg)))+(a.avg)+((muk+
    Vk)*(log((a.avg+b.avg)/(b.avg)))*(a.avg+b.avg)*(a.avg+b.
    avg+1)))) ---- Old calculation

R2LowerPvarKmean <- (Vk*((log((a.avg+b.avg)/(b.avg)))^2))/((
    Qk)*(((Vk)*((log((a.avg+b.avg)/(b.avg)))^2))+(log((a.avg+
    b.avg)/(b.avg)))+((muk+Vk)*((a.avg)/((b.avg)*(a.avg+b.avg
    +1))))))

R2UpperPvarKmean <- (a.avg*Vk*(a.avg+b.avg+1))/(Qk*((a.avg*
    Vk*(a.avg+b.avg+1))+(a.avg*b.avg)+(b.avg^2)+((muk+Vk)*b.
    avg)))



##### Actual bootstraps #####
```

```
for (i in 1:BS){
        while (p1bs[i]<=0 | p1bs[i]>=1){
                p1bs[i] <- rnorm(1, mean=p1hatmean, sd=
                    TrueSEp1)
        }
        a[i] <- (p1bs[i])*((((p1bs[i])*(1-p1bs[i]))/(p1bs[i
            ]^2))-1)
        b[i] <- (1-p1bs[i])*((((p1bs[i])*(1-p1bs[i]))/(p1bs[
            i]^2))-1)

}
### R-squared estimates P-constant ###
## IC-Jhala ##

R2LowerPconstJ <- ((Vj)*(log((1)/(1-p1bs))))/((Qj)*(1+((Vj)*
    (log((1)/(1-p1bs))))))
R2UpperPconstJ <- (Vj*p1bs)/(Qj*((Vj*p1bs)+(1-p1bs)))

## IC-Karanth ##

R2LowerPconstK <- ((Vk)*(log((1)/(1-p1bs))))/((Qk)*(1+((Vk)*
    (log((1)/(1-p1bs))))))
R2UpperPconstK <- (Vk*p1bs)/(Qk*((Vk*p1bs)+(1-p1bs)))

### R-squared estimates P-variable ###

## IC-Jhala ##

## R2LowerPvarJ <- (a*Vj*log((a+b)/(b)))/(Qj*((a*Vj*log((a+b
    )/(b)))+(a)+((muj+Vj)*(log((a+b)/(b)))*(a+b)*(a+b+1))))
    ---- Old calculation

R2LowerPvarJ <- (Vj*((log((a+b)/(b)))^2))/((Qj)*(((Vj)*((log
    ((a+b)/(b)))^2))+(log((a+b)/(b)))+((muj+Vj)*((a)/((b)*(a+
    b+1))))))

R2UpperPvarJ <- (a*Vj*(a+b+1))/(Qj*((a*Vj*(a+b+1))+(a*b)+(b
    ^2)+((muj+Vj)*b)))

## IC-Karanth ##

##R2LowerPvarK <- (a*Vk*log((a+b)/(b)))/(Qk*((a*Vk*log((a+b)
    /(b)))+(a)+((muk+Vk)*(log((a+b)/(b)))*(a+b)*(a+b+1))))
    ---- Old calculation
R2LowerPvarK <- (Vk*((log((a+b)/(b)))^2))/((Qk)*(((Vk)*((log
    ((a+b)/(b)))^2))+(log((a+b)/(b)))+((muk+Vk)*((a)/((b)*(a+
    b+1))))))

R2UpperPvarK <- (a*Vk*(a+b+1))/(Qk*((a*Vk*(a+b+1))+(a*b)+(b
```

```
    ^2)+((muk+Vk)*b)))


################
avar <- sum((a-a.avg)^2)/(BS-1)
aSE <- sqrt(avar)

bvar <- sum((b-b.avg)^2)/(BS-1)
bSE <- sqrt(bvar)

cat("p1avg mean =", p1hatmean, "p1avg SE =", TrueSEp1)
cat("a.avg mean =", a.avg, "a.avg SE =", aSE)
cat("b.avg mean =", b.avg, "b.avg SE =", bSE)

R2LowerPconstJvar <- sum((R2LowerPconstJ-
    R2LowerPconstantJmean)^2)/(BS-1)
R2LowerPconstJSE <- sqrt(R2LowerPconstJvar)

R2UpperPconstJvar <- sum((R2UpperPconstJ-
    R2UpperPconstantJmean)^2)/(BS-1)
R2UpperPconstJSE <- sqrt(R2UpperPconstJvar)

R2LowerPconstKvar <- sum((R2LowerPconstK-
    R2LowerPconstantKmean)^2)/(BS-1)
R2LowerPconstKSE <- sqrt(R2LowerPconstKvar)

R2UpperPconstKvar <- sum((R2UpperPconstK-
    R2UpperPconstantKmean)^2)/(BS-1)
R2UpperPconstKSE <- sqrt(R2UpperPconstKvar)

cat("R2LowerPconstJhala mean = ",R2LowerPconstantJmean, "SE
    = ", R2LowerPconstJSE)
cat("R2UpperPconstJhala mean = ",R2UpperPconstantJmean, "SE
    = ", R2UpperPconstJSE)

cat("R2LowerPconstKaranth mean = ",R2LowerPconstantKmean, "
    SE = ", R2LowerPconstKSE)
cat("R2UpperPconstKaranth mean = ",R2UpperPconstantKmean, "
    SE = ", R2UpperPconstKSE)

R2LowerPvarJvar <- sum((R2LowerPvarJ-R2LowerPvarJmean)^2)/(
    BS-1)
R2LowerPvarJSE <- sqrt(R2LowerPvarJvar)

R2UpperPvarJvar <- sum((R2UpperPvarJ-R2UpperPvarJmean)^2)/(
    BS-1)
R2UpperPvarJSE <- sqrt(R2UpperPvarJvar)
```

```
R2LowerPvarKvar <- sum((R2LowerPvarK-R2LowerPvarKmean)^2)/(
    BS-1)
R2LowerPvarKSE <- sqrt(R2LowerPvarKvar)

R2UpperPvarKvar <- sum((R2UpperPvarK-R2UpperPvarKmean)^2)/(
    BS-1)
R2UpperPvarKSE <- sqrt(R2UpperPvarKvar)

cat("R2LowerPvarJhala mean = ",R2LowerPvarJmean, "SE = ",
    R2LowerPvarJSE)
cat("R2UpperPvarJhala mean = ",R2UpperPvarJmean, "SE = ",
    R2UpperPvarJSE)

cat("R2LowerPvarKaranth mean = ",R2LowerPvarKmean, "SE = ",
    R2LowerPvarKSE)
cat("R2UpperPvarKaranth mean = ",R2UpperPvarKmean, "SE = ",
    R2UpperPvarKSE)
```

Chapter 7

An attraction-repulsion type Bayesian spatially-explicit capture-recapture model to assess behavioural plasticity in carnivore populations

7.1 Introduction

The role of behavioural processes in the shaping of population dynamic parameters remains to be one of the fundamental unanswered questions in ecology (May, 1999). While answers to some of these questions have been answered in controlled laboratory settings (Rainey & Rainey, 2003), they remain to be ecological mysteries in the wild. We argue that this situation will persist as long as the rigid ecological divisions (Wiens, 1989; Coulson *et al.*, 2006) that separate individuals from populations exist. And many fascinating questions lie at this interface between understanding individuals and populations.

Behavioral ecologists have for the past 30-40 years been very interested in the concept and understanding of phenotypic plasticity of behaviour (Dingemanse *et al.*, 2010). They have been interested to know to what extent a species will change its behaviour based on the degree of environmental variation during its lifetime (Piersma & Drent, 2003). Evolutionary ecologists have been investigating the extent to which genotypes vary in order to accommodate for changes in the environment beyond an individual's lifetime, over generations (Scheiner, 1993; Via *et al.*, 1995). More recently, there has been much interest in such changes under a systemic environmental pressure such as climate change. And there is concern in some that rapid changes in the environment, owing to climate change and forest degradation, are fast outpacing the genetic adaptations to such changes (Chevin *et al.*, 2010), bringing back focus on phenotypic plasticity of behaviour. For example, a recent argument in favour of

synurbization (Luniak, 2004) brings to light how some species such as the black bird (*Turdus merula*) adapt their behaviour completely and even thrive in a completely urban setting, and even doing better than in "natural" conditions.

At the population level of organization, ecologists have been investigating and confronting long-term population dynamic patterns, such as the classic snow shoe hare (*Lepus americanus*)populations (Wolff, 1980), or the coupled dynamics of lemmings and vole populations (Oksanen & Oksanen, 1992) and even merely trying to fit long term dynamic patterns to noisy data (De Valpine, 2003; Knape & de Valpine, 2012). Hence, growth of knowledge in ecology, both in theoretical and empirical terms, has focussed almost exclusively at two distinct levels, pretending there are no interactions lying at the interface. In doing so, some of the most important processes pertinent to the feedback between these two levels remains as a major knowledge gap in ecology. For example, increase in population density (at the population level) could cause an alteration in individual behaviour (e.g., cooperation), or conversely, behavioural constraints (e.g., territoriality) could set the carrying capacity K in a population by acting as an additional resource constraint (Wolff, 1997). This would require, therefore, that these two levels be studied in unison rather than separately as has been the approaches until now.

Arguably, the major stumbling block in studying of wild populations is the sampling problem (Williams *et al.*, 2002) - the unavoidable fact that only a *proportion* of the total animals in a population are available for observation. This fundamental problem has led behavioural and population ecologists to use methods that are seemingly disparate in their approaches and motivation. To date, study of behaviour and individuals has relied on techniques that facilitate direct (e.g., ethograms, time and activity budget charts; sommer1991practical) or indirect (e.g., analysing telemetry or ringed data; (Dunn & Gipson, 1977)) methods to facilitate detailed behavioural observations. On the other hand, the study of populations has relied on estimation methods, such as mark-recapture approaches (Williams *et al.*, 2002; Amstrup *et al.*, 2005) for studies of state variables (such as abundance) or vital rates (such as survival and recruitment). Although mark-recapture approaches do, in some sense, track the fate of individual animals, and provides vital information in the form of parameter estimates of vital rates (see Karanth *et al.*, 2004, as an example) that would interest behavioural ecologists, such information can merely be *correlated* with the sort of questions that they would be interested in. And the only sort of behavioural questions in mark-recapture studies relates to phenomena such as "trap-shyness" or "trap-happiness" which are mere refinements in estimating detection probability.

With this background, we take a closer look at carnivore studies. Given their role as apex predators, carnivore biologists have, for long, been interested in the mating tactics and spacing rules governing the order Carnivora (Sandell, 1989). Apart

from having relatively much larger home range sizes, compared to those species lower in the food chain, there has been much excitement and charisma surrounding, for example, their behaviour, such as territoriality (e.g., Kruuk, 1978; Lindström, 1986; Sillero-Zubiri & Macdonald, 1998). Remarkably, however, they also exhibit a wide range of cooperative behaviours, for example, within the family *Felidae*, lions (*Panthera leo*) tend to form social groups whilst tigers (*Panthera tigris*) do not (Macdonald & Loveridge, 2010). And within the *Canidae*, we see a range of pack sizes and social structure (Macdonald & Sillero-Zubiri, 2004). While there have been many hypotheses (e.g., Macdonald, 1983), which revolve around some form of prey choice and distribution, to explain this wide variety of social behaviours within the landscape, the lack of a unifying approach to study these phenomena at the interface of behavioural and population ecology of carnivores has resulted in a less than satisfactory understanding of these processes.

However, the advent of spatial capture-recapture (SECR) models (Borchers & Efford, 2008; Royle *et al.*, 2009b,a; Gopalaswamy, 2013) and their explosive development over the past five years (Royle *et al.*, 2013b) has provided a valuable set of tools to aid in breaking down this artificial division that exists between behavioural and population ecology, especially in carnivores. The central motivation behind the development of SECR models was to reliably estimate animal density in capture-recapture studies using auxiliary information about the location of captured individuals. The basic hierarchical structure of the model (Royle *et al.*, 2009a) serves as a blueprint for us for further research and tackle some basic questions relating to carnivore space use patterns and formation of societies (Macdonald & Johnson, In Press).

The hierarchical nature of SECR models, distinguishes the *state process*(the manner in which animals are distributed in space and the process that ecologists are most interested in) from the *observation process*(the manner in which data are gathered through surveys, a reality that cannot be ignored). Royle & Dorazio (2008) provides a detailed account of this modeling philosophy for a range of ecological problems focusing on the process of estimation important parameters, such as abundance and occupancy. From the Bayesian version of the basic SECR model (Royle *et al.*, 2009a) has emerged a plethora of refinements, with an emphasis on the observation process (Sollmann *et al.*, 2011; Royle *et al.*, 2011b) and some simple modeling of the state process (Royle *et al.*, 2013a,c) with a basic aim of refining estimates of abundance. Two studies (Gardner *et al.*, 2009; Ergon & Gardner, 2013) go on to estimate vital rate parameters and dispersal parameters, relevant to long term population dynamic studies that involves some modeling of individual movement behaviours in populations. Recently, however, Reich & Gardner (2014) looked more closely at conspecific interactions by integrating a spatial point process called the

Strauss process (see Møller & Waagepetersen, 2004) into SECR models to quantify the degree of "territoriality". From the standpoint of answering ecological questions, this development (Reich & Gardner, 2014) can be very useful to investigate the notion of "tolerance" between individuals with respect to their organization in space. The idea being that some species in some specific ecological conditions may exhibit a higher degree of tolerance to conspecifics than others. However, the basic structure of the Strauss process provides little inference on the degree of "cooperation" or "attraction" among conspecifics, and this question has for long been of great interest to studying various ecological systems (Axelrod & Hamilton, 1981; Nowak, 2006).

In this study, we aim to develop and confront a more flexible spatially-explicit capture-recapture model involving conspecific interaction that will allow us to be able to ask multiple questions at the interface of behavioural and population ecology. For example, we may like to ask:

1. Will a species change its behaviour (i.e. will it become more tolerant of conspecifics) when resources are enriched causing an increase in population size? To what extent does a species exhibit behavioural plasticity in terms of accommodating conspecifics within its home ranges?

2. Can we actually estimate behavioural thresholds, such as 'switching' from repulsion to attraction, or vice versa, at specific environmental circumstances, say at varying degree of population sizes?

3. To what extent do behavioural constraints contribute to density-dependence in a species?

4. Can a species changed its behaviour at an evolutionary time scale? For example, can the sociobiology of tigers (solitary living) become like lions (group living) if ecological conditions (prey resources stay enriched for long periods of time)?

5. What behavioural thresholds come into play with respect to conspecific interactions that will force carnivores into conflict with humans?

The basis for answering these questions will lie in the development of an integrated behavioural-population ecology statistical model that can specifically estimate parameters of "attraction" and "repulsion" in a combined set up. Thus, motivated to tackle the above sort of questions, we developed an attraction-repulsion SECR statistical model that estimates specific parameters that aids inferences about such questions.

7.2 Material and methods

7.2.1 Notations and variables used

Table 7.1 lists the various notations and variables used in the development of the statistical model below.

7.2.2 Sampling scheme

We envision a standard spatial capture-recapture data set (see Gopalaswamy *et al.*, 2012d, for a full example) that is a 3-dimensional matrix, consisting of individuals ($i = 1, 2, 3 \ldots N$) x trap locations ($l = 1, 2, 3 \ldots L$) x sampling occasions ($t = 1, 2, 3 \ldots K$).

7.2.3 Model development

We follow the general hierarchical modeling philosophy (Royle & Dorazio, 2008) to go about the construction of our model. Accordingly, data obtained from a field experiment is viewed as a consequence of two distinct processes. Firstly, we have the *state process* - which in our case, refers to the true distribution of animals within a bounded region. There are usually two unknown questions associated with the state process - we do not know the true population size and we do not know the spatial locations of these animals. Secondly, we have the *observation process* - which in our case, refers to the manner in which data are gathered given the distribution of animals over a region of interest. Our goal therefore is to draw inferences about the state process, given that too often the resulting data on hand is obscured by the observation process.

We develop the basic Bayesian SECR model (Royle *et al.*, 2009a) with an emphasis on the state process to describe conspecific interactions (of attraction and repulsion). We preserve the same observation process from Royle *et al.* (2009a) on spatial capture-recapture data obtained from camera traps, in our model development.

First, we define a large statespace S around the trapping grid that is large enough to ensure no individual outside this region has any probability of being sampled during the period of observation (the sampling period). Since the population density is unknown, we assume the true population size at this unknown density within S to be N. In order to proceed with the estimation we define a much larger value M, called the data augmented value of abundance which will serve as the absolute upper limit to the unknown N. In order to simplify computation we divide S into a number of small pixels G, such that the resolution is fine enough to ensure that errors on animal locations relative to their intrinsic movement behaviour is relatively

small. The locations of each individual i, s_i, is also unknown. Additionally, when there are conspecific interactions s_i will depend in some way on the locations of all other individuals within S.

During the process of observation, the total number of individuals caught in the study is represented by n. This value is inflated with n_z, so that the data augmented population size $M = n + n_z$. The true population size N, therefore, will lie somewhere between n and M.

Thus, the true population size is defined by the model,

$$[N|M, \psi] \sim \text{Binomial}\,[M, \psi]$$

Here, ψ is the proportion of the data augmented population size that are real individuals. These N individuals are then arranged according to a defined spatial point process (Møller & Waagepetersen, 2004). A spatial point process can be simply defined as a statistical model that maps point patterns over space according to the probability function defined in the model. If there were to be no conspecific interaction, then each individual is independent of the other, and such was the basis of construction of the basic Royle *et al.* (2009a) SECR model. Since, in our case, we are defining an interaction, such a point process involving dependencies is termed as a Marked point process (Møller & Waagepetersen, 2004).

In our specific case, we define a conspecific environment in the statespace. In our framing of the problem, this environment can either be 'attractive' or 'repulsive', according to the model,

$$[\omega|\pi] \sim \text{Bernoulli}\,[\pi]$$

Here, π is the probability of the conspecific environment being *attractive* and $1 - \pi$ is the probability of the conspecific environment being *repulsive*. In our Bayesian approach, we can begin with a prior probability of say 0.5 for π during the process of estimation.

We next consider the arrangement of the N individuals within the statespace. If there is no conspecific interaction between individuals, the N individuals will occupy any one of the point locations with a probability of $1/G$, where G is the total number of discrete pixels that defines the statespace. However, with conspecific interactions, the probability field will likely be inhomogenous (Borchers & Efford, 2008). So, a new individual i will find an entry into one of the grid cells s_i in such an inhomogenous statespace according to a multinomial probability as follows,

$$[s_i|\omega] \sim \text{Multinomial}\,[\Phi_i(\omega)] \sim \text{Multinomial}\,[p_{\omega 1}, p_{\omega 2}, p_{\omega 3} \ldots p_{\omega G}]$$

Here, $\Phi_i(\omega) = (p_{\omega 1}, p_{\omega 2}, p_{\omega 3} \ldots p_{\omega G})$, is the probability field generated by the conspecific environment defined by the value of ω.

Importantly here, s_i should also depend on the population size N in some way. We can envision two hypothesis of behaviour here. First, if behavioural effects (e.g., territoriality) has an effect on the carrying capacity then there will be limits to the number of individuals that can enter the statespace. We do not explore these effects in our model development, although it may be easier to accommodate those effects directly in an alternative Bayesian SECR model (Royle *et al.*, 2013c). In our model development, the location of each member will be governed by $\Phi_i(\omega)$, which will be a function of parameters that lays down the "social rules" for the population of size N.

Similar to the observation process in spatial capture-recapture models (Borchers & Efford, 2008; Royle *et al.*, 2009a), wherein, the probability of detection of an individual declines as a function of distance from the trapping devise, such as a camera trap, the probability of an individual, say i, to occur in a particular grid cell will depend on how far away the grid cell is from a conspecific, say j. If the probability field is"attractive", then the probability of occurrence on a particular grid cell s_i will decline as a function of distance from j. Conversely, if the probability field is"repulsive", then the probability of occurrence of i on a particular grid cell will increase as a function of distance from a conspecific j.

The rate of decline in detection probability in SECR models is defined by the parameter σ. And this parameter is also related to movement of individuals. It can also be thought of as, crudely, the home (or activity) range radius of the individuals in a population during the period of sampling. In carnivores, it is commonly accepted that in their usual ranging behaviour, the animals also *patrol* these areas and defend their territories (Macdonald & Sillero-Zubiri, 2004; Macdonald & Loveridge, 2010). Therefore, in our parameterization of the problem we would like to investigate the behavioural effects of attraction or repulsion relative to the parameter σ. Here, we introduce another parameter κ. In order to define the interaction between conspecifics, instead of using only the parameter σ we will use the multiplied term $\kappa\sigma$. It is therefore clear here that when $\kappa = 1$, $\sigma = \kappa\sigma$ and this implies that the extent of influence of an individual i in defining the position of a conspecific j, is exactly equal to the relative amounts of time i spends in its activity range during the period of sampling. However, when $\kappa < 1$ we imply that the conspecific interaction is more concentrated at the activity centre than on the periphery of its activity range. In a repulsive field, such as in a territorial environment, a value of $\kappa < 1$ indicates that individuals are exhibiting some level of "tolerance" whilst sharing home ranges. Similarly, in an attractive field, a value of $\kappa < 1$ indicates that animals are more closely social than the natural amount that movement indicates. This distinction is important in distinguishing between two important biological phenomena - because, σ informs us *only* about resource use, while κ gives us the very specific

insight about attractive-repulsive behaviour based on the spatial arrangement of individuals in the statespace. And when $\kappa > 1$, we start approaching a situation where the role of conspecific interactions itself starts to become insignificant. In other words, when value of $\kappa >> 1$, we approach a situation relating to the more general spatial capture-recapture model (Royle *et al.*, 2009a), whereby, the probability of an individual occurring at any individual grid cell is not influenced by the location of conspecifics, irrespective of the nature of the field.

More generally, for an individual i located at s_i, the probability of observing this individual during sampling is given by the product of two probabilities: Prob(detection by a detecting device — location at s_i) x Prob(of obeying the social rules to reside at s_i). So, the probability of detection of an individual i located at s_i to be detected by a trap location l on a given sampling occasion t is given by:

$$\text{cloglog}\left(\pi_{il}\right) = \log\left(\lambda_0\right) - \frac{d_{il}^2}{2\sigma^2}$$

Here, λ_0 is the basal encounter rate when the individual i's activity centre is exactly on trap j. d_{il} is the distance between individual i and the trap location l.

This is the same observation structure as described in (Royle *et al.*, 2009a; Gopalaswamy *et al.*, 2012d). Most marked point processes described in the spatial point process statistical literature (Møller & Waagepetersen, 2004; Goldstein *et al.*, 2014) view the attraction-repulsion problem in relation to an estimable distance measure, as a yardstick, for defining attraction or repulsion. So, the conspecific interaction is defined as "attraction" if the distance between two individuals is less than the yardstick measure and is defined as "repulsion" if the distance between two individuals is less than the yardstick measure. This feature makes the problem *scale dependent*, so the notion of attraction or repulsion is based on the yardstick measure and the problem may be compared to the larger ecological problem of scale (Levin, 1995). In our specific problem, such a scale-dependent attraction-repulsion scenario may be sensitive to the choice of the size of the statespace and the yardstick itself may be an awkward way of setting up this problem. Instead, we propose an alternative view of the problem and define attraction $\propto {}^{1}/_{\text{repulsion}}$. For example, if the probability field is "repulsive" (i.e., $\omega = 0$), an individual i will have a higher probability of being located at grid cells further away from a conspecific j, and similarly, if the probability field is "attractive" (i.e., $\omega = 1$), an individual i will have a higher probability of being located at grid cells closer to a conspecific j relative to grid cells located further away. For a finite number of grid cells G, the probability field that is created in our construction is relative and *not* absolute. The biological point of being relative, is that behaviour is assumed to not impose a constraint for a new individual from entering the system, but only plays a role in the determination of where this new individual must be located.

We shall define $g = 1, 2, 3, 4 \ldots G$ to be the grid cells defined in the discrete statespace. Let us consider the situation where a new individual i has to be located on any g^{th} cell. Consider $j = 1, 2, 3, 4 \ldots J$ to be J conspecific located on J grid cells, given that a grid cell can host more than one individual as it's activity centre. Then, the probability of an individual occupying a grid cell g for the "attraction" model is then given by,

$$[p_{ig}|\omega = 1] = C_1 \times \frac{e^{-1/2\kappa^2\sigma^2\left(\sum_{j=1}^{N_{i-1}} d_{gj}^2\right)}}{\sum_{g=1}^{G} e^{-1/2\kappa^2\sigma^2\left(\sum_{j=1}^{N_{i-1}} d_{gj}^2\right)}}$$

Similarly, the probability of an individual occupying a grid cell g for the "repulsion" model is,

$$[p_{ig}|\omega = 0] = C_2 \times \frac{1}{[p_{ig}|\omega = 1]}$$

Here, C_1 and C_2 are simply normalizing constants to ensure that $p_i g$ remains as a probability and $\sum_{g=1}^{G} p_{ig} = 1$. And d_{gj} is the Euclidean distance from grid cell g to location of conspecific j. Here, N_{i-1} is the total number of individuals added to the population prior to the addition of the i^{th} individual.

In our estimation, therefore, we aim to achieve twin objectives of finding optimal locations for each real individual in the data augmented population based on the relative positions of conspecifics and the model describing social rules, and simultaneously, finding locations based on the observation likelihood generated by the spatial capture-recapture data.

7.2.4 Implementation of the model

We implement the model by developing a Bayesian Markov Chain Monte Carlo (MCMC) piece of code in the programming environment R (R Development Core Team, 2009) based on the Metropolis-Hastings algorithm (Bolker, 2008). The Appendix 7.7.1 contains the detailed code for this estimator.

7.2.5 Simulations

Simulations of marked point processes is generally not a straight forward exercise because the likelihoods are generally intractable(Møller & Waagepetersen, 2004). However, we simulate data sets by adding individuals to the state space, *sequentially*, as a Markov process. We recognize that such simulated data sets may not be an accurate description of the estimator developed above. However, we believe this approach is less of a problem if the carrying capacity in a population is *not* influenced by spacing patterns, which is the assumption in our case. Appendix 7.7.2 contains the detailed R code for the sequential simulator function.

Firgures 7.1 and 7.2 illustrate the spatial point patterns generated by such sequential deposition of individuals into the statespace. Figure 7.1 provides a repulsive field ($\omega = 0$), and figure 7.2 provides an attractive field ($\omega = 1$). These two are the state processes. Our simulator also provides a spatial capture-recapture data set for a given set of observation process parameters.

7.3 Results

We successfully developed both a *sequential* simulator and an estimator for analysing spatial capture-recapture data, when the state process is governed by an attraction-repulsion type marked point process. For a typical sample size of about 50 individuals, augmented to about 450 individuals in the statespace with approximately 2000 pixels in resolution, the estimator takes some 100-120 seconds per iteration in the MCMC chain. A dry run of 500 iterations demonstrated adequate mixing in the MCMC chain. An extensive simulation exercise and trials on a data set is yet to be undertaken.

7.4 Discussion

Many interesting ecological and evolutionary questions lie at the interface between the various levels of organization in ecology. One such interface lies between the domains of behavioural and population ecology (May, 1999). By committing research effort to any particular level without suitably mapping onto other levels (Coulson *et al.*, 2006) we preclude our ability to fully appreciate the complexity and interconnectivity in nature, which is one of the fundamental goals of ecological studies. Such a divisive approach also diverts attention of the quantitative ecologists into finding solutions to problems only within such specific levels, forfeiting several opportunities of investigations into the interconnectedness in nature (Mandelbrot, 1983).

In this context, and motivated by behaviour and population level studies within the order Carnivora, we aim to bring about a methodological synthesis between these two levels by developing a statistical estimator that incorporates the attraction-repulsion behaviour within a spatial capture-recapture model that estimates population density. The estimator will thus enable us to tackle several unanswered questions about how animals space themselves, and whether the much talked about phenomena of behavioural traits such as "territoriality" or "cooperation" (Maher & Lott, 1995), particularly among carnivore species, are natural phenomena or whether they are glorified in the context of population regulation due to resource depletion.

Within the realm of statistical ecology, spatial capture-recapture models (Borchers & Efford, 2008; Royle *et al.*, 2009a) are increasingly occupying a major portion in

the abundance estimation domain. Combined with statistical techniques, such as spatial point processes (Møller & Waagepetersen, 2004), SECR models can become a powerful tool to investigate all kinds of social arrangements in carnivores and simultaneously explore the link between behavioural mechanisms, such as the resource dispersion hypothesis (Macdonald, 1983; Macdonald & Johnson, In Press) or even alternative mechanisms such as the cub-defense hypothesis or prey renewal hypothesis (Revilla, 2003), and population dynamic mechanisms (Oksanen & Oksanen, 1992). The power of this model can be harnessed most effectively at large spatial and temporal scales. For example, Clutton-Brock & Sheldon (2010) list and explain the tremendous growth in ecological knowledge, especially with respect to birds such as the Florida scrub jay (*Aphelocoma coerulescens*), in long term individual-based studies. A limitation in most of these long term studies, is that the spatial scale of study is relatively small owing to the fact that at larger spatial scales we run into the sampling problem (Williams *et al.*, 2002). Our model can effectively deal with both the problem of real ecological patterns, and addressing the sampling problem effectively.

We have used an alternative treatment of marked point processes in the sense that our framing of the attraction-repulsion problem is different from the one that is set up in standard spatial point processes literature (Møller & Waagepetersen, 2004; Goldstein *et al.*, 2014). However, the development of the model is still in its preliminary stages. While we have developed a "sequential" simulator we recognize this is a very close but approximate solution to a true simulator. Further work is necessary, perhaps using an MCMC algorithm or a simulated-annealing approach (Bolker, 2008), to optimize the simulator to reflect a true simulation process. This will facilitate more accurate testing of the developed estimator and estimator performance can then be evaluated. The models will also have to be confronted with real spatial capture-recapture data from natural populations. We recommend that the model first be confronted with data only from the sex that most characterizes resource selection, for example, females in mammals.

From the standpoint of ecology and conservation of carnivores, this model serves to assess the behavioural plasticity of individuals and the response time to change in behaviour over a period of time. For example, even in successful conservation initiatives when key threats are addressed to population, it will take some time for individuals to respond to these changes. An increase in resources should cause a decrease in home ranges followed by the "tolerance" necessary to absorb new individuals into the population level. In the future, we also hope to assess species-specific models, building on the approaches developed here.

7.5 Tables

Table 7.1: Notations and variables used in the study

Symbol or Letter	Definition
S	the larger statespace
n	number of individuals caught in the study
N	true population size in the larger statespace
s_i	location of an individual i on the larger statespace
n_z	augmented number of individuals for the purpose of estimation
M	$= n + n_z$, absolute maximum number of individuals possible within the larger statespace
ψ	proportion of real individuals in the data augmented population M
G	total number of grid cell points in the within the larger statespace
Φ_i	$= p_{i1}, p_{i2} \dots p_{iG}$, is the probability field defined over G making some regions more attractive for a new individual i entering the statespace
λ_0	basal encounter rate, expected encounter rate of an individual i whose centre of activity is on a trap j
σ	rate of decline in the probability of detection of an individual animal i as it moves away from a trap j
ω	a Bernoulli indicator variable indicating whether Φ is (1 =‘attractive’) or (0 =‘repulsive)
κ	relative strength of conspecific interaction as compared to its usual movement or detection decline rate parameter σ

7.6 Figures

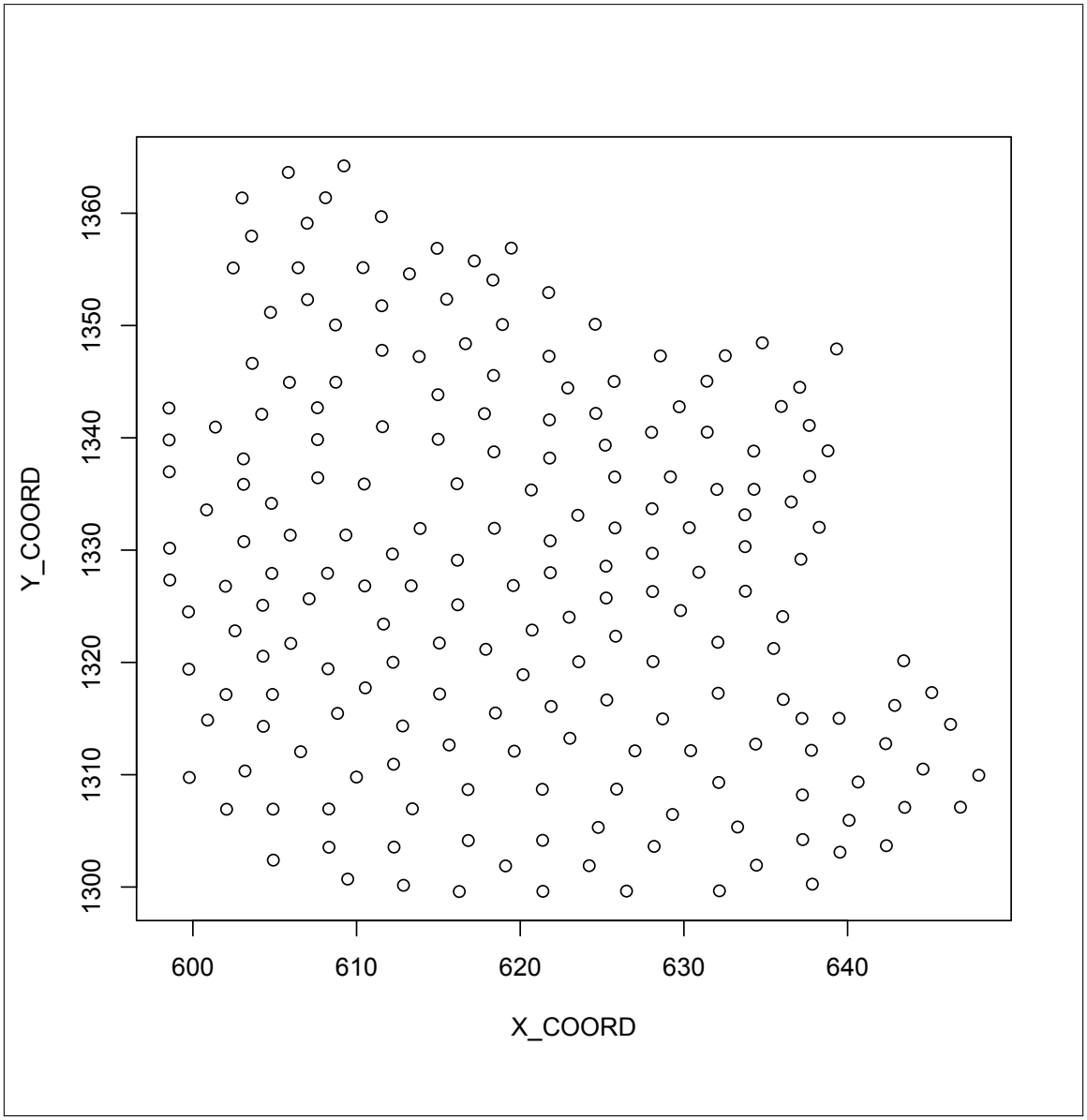


Figure 7.1: A simulation of the state space with a repulsive field $\omega = 0$. Here, $N = 200$, $\sigma = 0.7$ and $\kappa = 0.4$

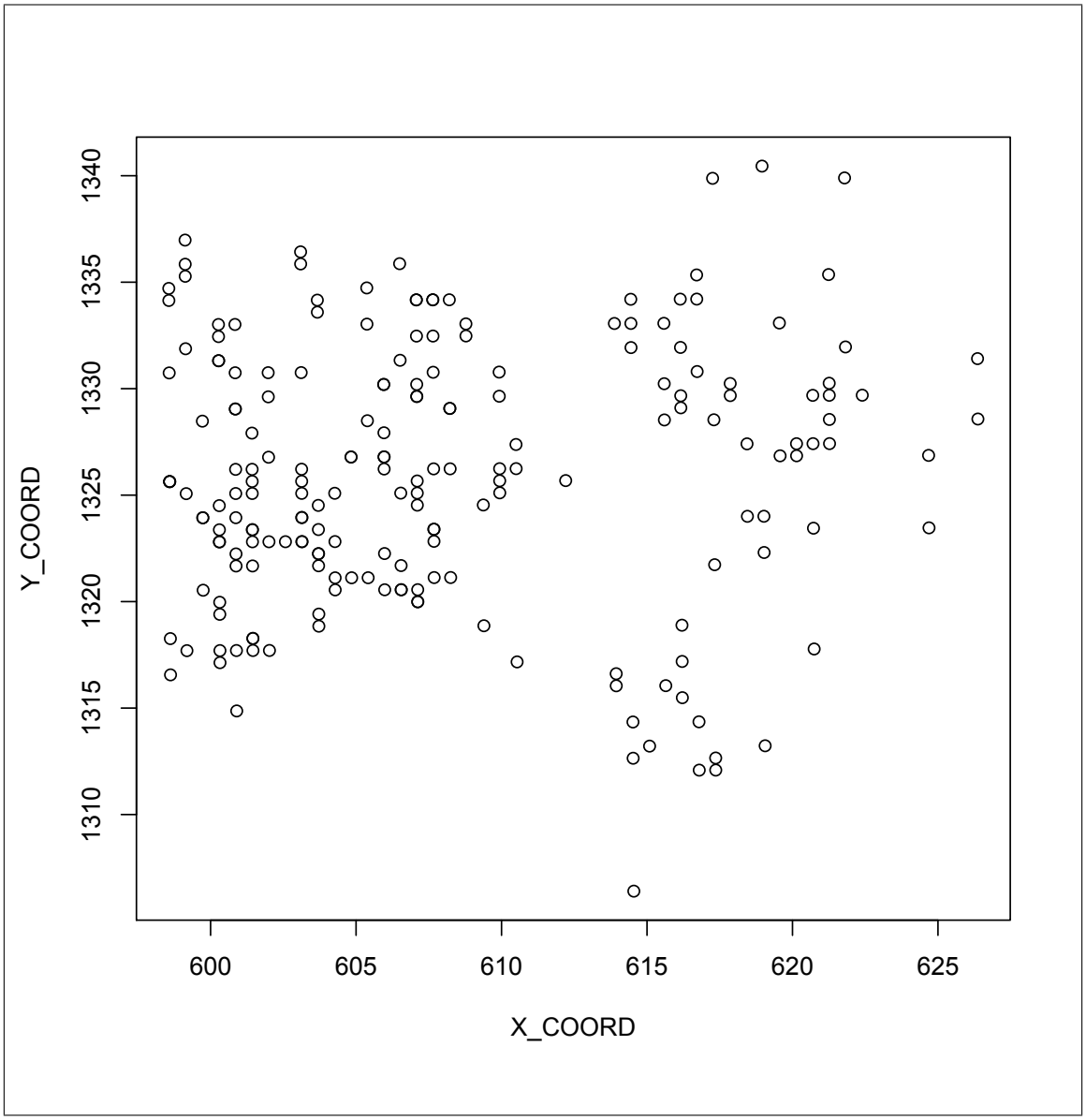


Figure 7.2: A simulation of the state space with an attractive field $\omega = 1$. Here, $N = 200$, $\sigma = 2.3$ and $\kappa = 0.4$

7.7 Appendices

7.7.1 Attraction-repulsion SECR model - R code

```
#### APPENDIX 7.7.1: Attraction-repulsion SECR model - R
    code function ####

SCRj1.fn <-
function (scrobj, ni = 10, burn = 2, skip = 1, nz = 200,
Msigma = 1,
Mb = 0,
Msex = 0,
Msexsigma = 0,
Mss=0,
Meff = 0,
Mtel = 0,
Mattrep = 0, # [AMG] This is the switch for attraction-
    repulsion code
theta = 1,
coord.scale = 1000, area.per.pixel = 1,
     thinstatespace = 10, maxNN = 20, dumprate = 1000)
{

     call <- match.call()
# [AMG] The scrobj object contains details about all input
    data including covariates
          if (Mattrep==1){
                Msigma=1
                Mb=0
                Msex=0
                Msexsigma=0
                Mss=0
                Meff=0
                Mtel=0
          } # [AMG] This is just to ensure that a simple model
              for attraction-repulsion runs without the other
             complexities. We can introduce complexities later
             .


     traps <- scrobj$traps # [AMG] Calls the trap matrix
        which includes trap operation details
     captures <- scrobj$captures # [AMG] Includes columns
        sessions, individual, occasion, trapid
     statespace <- scrobj$statespace # [AMG] grid cell no, x.
        coord, y.coord, habitat mask (1/0)
     alive <- scrobj$alive # [AMG] Individual x Occasions
        matrix (1 for alive, 0 for dead). Default is "all
        alive 1s"
# Mss = fit "Xd" covariate [AMG] This is the density
    covariate influencing the statespace
```

```
    Xd <- scrobj$Xd # [AMG] Default is 1s on all grid cells
# Meff = fit "Xeff" covariate [AMG] This is the effort
    covariate
     Xeff <- scrobj$Xeff
     Xsex <- scrobj$Xsex       # [AMG] A vector that contains
        a series of 1 or zero indicating sex-for all known
        individuals
     Xtel <- scrobj$Xtel       # [AMG] Related to
        incorporating telemetry data
     if(is.null(scrobj$Ytel)){
          loglik.tel<-function(bsigma,Ytel,Msexsigma){
              return(0)
              }
          }
     if(!is.null(scrobj$Ytel)){
        Ytel<- scrobj$Ytel
     if(is.null(Xtel)){
          loglik.tel<-function(bsigma,Ytel,Msexsigma){
              sbar<-attributes(Ytel)$sbar
              sex.tel<-attributes(Ytel)$sex

              sigma <- sqrt(1/(2*bsigma) )
              if(length(bsigma)==1){
return(              sum(     dnorm(Ytel[,1],sbar[,1][Ytel
    [,3]],sigma,log=TRUE) +
            dnorm(Ytel[,2],sbar[,2][Ytel[,3]],sigma,log=TRUE)
                                 ))
              }
              if(length(bsigma==2)){
                tmp<- sex.tel[Ytel[,3]]
               return(sum(dnorm(Ytel[,1],sbar[,1][Ytel[,3]],
                   sigma[tmp+1],log=TRUE) +
              dnorm(Ytel[,2],sbar[,2][Ytel[,3]],sigma[tmp+1],
                  log=TRUE)                ))
            }
          }
          ##tel.gr<-     make.statespace(ll = statespace,
              buffer = 0.01, nx = 40)



     }

if(!is.null(Xtel)){
# extract coordinates and covariate values here

}
```

```
    }


    captures <- cbind(captures[, 4], captures[, 2], captures
        [,
        3]) # [AMG] Re-ordering by Trap ID, Individual ID,
            Occasion No. captures[,1] includes session for
            future work on multi-session models.
    if (length(unique(captures[, 2])) != length(min(captures
        [,
        2]):max(captures[, 2]))) {
        cat("Error: individuals not numbered sequentially,
            renumbering them now",
            fill = TRUE)
        captures[, 2] <- as.numeric(factor(captures[, 2]))
    } # [AMG] Routine error-checking function
    Y <- captures
    traplocs <- traps[, 2:3] # [AMG] Just a two column list
        of trap location coordinates

    xrange <- max(traplocs[,1])-min(traplocs[,1])
    yrange <- max(traplocs[,2])-min(traplocs[,2])
if( (xrange + yrange) < coord.scale)
    {
cat("coord.scale appears to be too large, check this out,
    maybe set to 1", fill=TRUE)
        return(0)
    } #[AMG] coord.scale is the scaling parameter that
        scales the coordinates into smaller, manageable
        quantities. For example: default converts metres to
        kilometres.
    MASK <- matrix(1, nrow=nrow(traplocs), ncol=ncol(traps)
        -3)
    # MASK <- as.matrix(traps[, 4:ncol(traps)]) # [AMG]
        matrix of the trap operations mask (typically when
        camera traps don't work) no. of traps x no. of
        occasions
    nind <- max(Y[, 2]) # [AMG] Total number of individuals
        in study
    T <- dim(MASK)[2] # [AMG] 2 here refers to the second
        dimension (no of occasions)
    M <- nind + nz # [AMG] Total number of individuals (
        caught and augmented)
    ntraps <- nrow(traplocs) # [AMG] Total number of traps
    totalarea <- nrow(statespace) * area.per.pixel # [AMG]
        Combined area of larger state-space
    thinned <- seq(1, nrow(statespace), thinstatespace) # [
        AMG] Greater the thinstatespace, more coarse is the
        grid of potential activity centres.
```

```
statespace <- statespace[thinned, ] # [AMG] Statespace
    is now thinned out
Xd <- Xd[thinned] # [AMG] Statespace covariate has to
    also be thinned
goodbad <- statespace[, 3] # [AMG] Habitat mask where 1
    is good habitat and 0 is bad habitat
G <- statespace[, 1:2] # [AMG] Extracts only the
    coordinates columns from statespace
G <- G[goodbad == 1, ] # [AMG] Filters out those
    coordinates of pixels which have good quality habitat
Gunscaled <- G # [AMG] Is a variable that only has good
    habitat information
Xd <- Xd[goodbad == 1] # [AMG] Filters out density
    covariates on statespace only on permissible habitat
nG <- nrow(G) # [AMG] Is a count of number of grid
    points
new.area.per.pixel <- totalarea/nG # [AMG] IS NOT THIS
    UNNECESSARY? CHECK FOR USE LATER
# [AMG] Trap locations and the statespace is now re-
    scaled based on the coordinate scale. Default is
    1000, typically for converting metres to kilometres.
mgx <- min(traplocs[, 1])
mgy <- min(traplocs[, 2])
traplocs[, 1] <- (traplocs[, 1] - mgx)/coord.scale # [
   AMG] Brought to x-origin
traplocs[, 2] <- (traplocs[, 2] - mgy)/coord.scale # [
   AMG] Brought to y-origin
G[, 1] <- (G[, 1] - mgx)/coord.scale # [AMG] G is now
    rescaled (can take negative coordinates)
G[, 2] <- (G[, 2] - mgy)/coord.scale # [AMG] G is now
    rescaled (can take negative coordinates)


msk2 <- array(NA, c(nind + nz, T, ntraps)) # [AMG]
    Creates ntraps matrices of (nind+nz) x T. This the is
     masking for the known and augmented set of
    individuals.
for (i in 1:(nind + nz)) {
    msk2[i, 1:T, 1:ntraps] <- t(MASK[1:ntraps, 1:T])
} # [AMG] Replaces NAs with 1s and 0s containing trap
    operations information. Essentially, how many
    individuals are exposed on a specific occasion to a
    particular trap.
msk2 <- as.vector(msk2) # [AMG] Creates a long vector of
     the masking information. So 1/0 on 1:M on T=1 on
    trap=1, 1:M on T=2 on trap=1.....I:M on T=1 on trap
    =1, 1:M on T=1 on trap=2.
Ynew <- array(0, dim = c(nind, T, ntraps)) # [AMG] Array
     of dimension of caught individuals
```

```
Ynew[cbind(Y[, 2], Y[, 3], Y[, 1])] <- 1 # [AMG]
Y <- Ynew      # [AMG] Essentially, we fill up 0s to the
    old Y (or captures) object and reorder it by
    Individual, Occasion and TrapID.
Yaug <- array(0, dim = c(nind + nz, T, ntraps))
for (j in 1:nind) {
    Yaug[j, 1:T, 1:ntraps] <- Y[j, 1:T, 1:ntraps]
} # [AMG] Creates an augmented capture array. This has
    nind+nz individuals.

## [AMG] Code below incorporates the various types of
    covariates effort, sex-specific stuff and behavioural
     stuff

if (Meff ==1 ) {
    ######################!is.null(Xeff)) {
    Xeffnew <- array(0, dim = c(nind + nz, T, ntraps))
    for (j in 1:M) {
        Xeffnew[j, 1:T, 1:ntraps] <- t(Xeff)
    }
   ######################  Xeff.tf <- TRUE [AMG] Effort
      covariate incorporation such as setting baits in
      camera traps etc.
}
if (Meff ==0){
    Xeffnew <- array(0, dim = c(nind + nz, T, ntraps))
   ######################## Xeff.tf <- FALSE
}
Xeff <- Xeffnew
if (!is.null(Xsex)) {
    Xsexnew <- c(Xsex, rep(NA, nz))
    # [AMG] If sex is defined. NA for all augmented
        individuals.
}
if (is.null(Xsex)) {
    Xsexnew <- rep(0, nind + nz)
}
Xsex <- Xsexnew # [AMG] A long vector of M individuals
    with sex information
Xsex <- as.numeric(Xsex) # [AMG] Inserted 5 Aug 14
sex.naflag <- is.na(Xsex) # Determine all those
    individuals whose sex is unknown, marked with NA

prevcap <- array(0, c(nind + nz, T, ntraps))
for (i in 1:(nind)) {
    for (j in 1:ntraps) {
        tmp <- Yaug[i, 1:T, j]
        if (any(tmp == 1)) {
            fst <- min((1:T)[tmp == 1])
```

```
                if (fst < T)
                  prevcap[i, (fst + 1):T, j] <- 1 # [AMG]
                     prevcap has 1 on all occasions after
                     first capture on a particular trap j
              }
          }
      }
# [AMG] Here trap response is trap-specific. An animal is
    assumed to have a different probability of capture at a
    particular trap after first capture. Same animal is
    assumed not to be influenced by the first capture at one
    trap location when it encounters another trap at a
    different location.
     prevcap <- as.vector(prevcap)
     alive.trues <- array(1, c(nind + nz, T, ntraps)) # [AMG]
         1s put on all M individuals
     for (i in 1:nind) {
         for (t in 1:T) {
             alive.trues[i, T, 1:ntraps] <- alive[i, t]
         }
     } ###### [AMG] Records dead animals if any
     alive.trues <- as.vector(alive.trues) # [AMG] One long
        vector containing information on alive/dead animals
     aliveid <- alive.trues[msk2 == 1] # [AMG] Filters out
        alive/dead status only for relevant individual, traps
         and occasions that is operational
     arr.trues <- array(TRUE, c(nind + nz, T, ntraps)) # [AMG
        ] Generates an array of TRUEs for the augmented data
        set
     idx <- which(arr.trues, arr.ind = TRUE) # [AMG] idx
        simply is an array containing index number eg: first
        row is 1(individual no),1(occasion no),1(trap no)
        second is 2,1,1.....nrow(idx)=(nind+nz)*(T)*(ntraps)
     y <- as.vector(Yaug) # [AMG] Note this is not the same
        as Y
     y <- y[msk2 == 1 & alive.trues == 1] # [AMG] A condition
         imposed to filter out unexposed traps and dead
        animals from the augmented data set to avoid them to
        appear in the likelihood.
     Xeff <- as.vector(Xeff)
     Xeff <- Xeff[msk2 == 1 & alive.trues == 1]
     prevcap <- prevcap[msk2 == 1 & alive.trues == 1]
     indid.LM <- idx[msk2 == 1, 1] # [AMG] This is just a
        repeated set of individuals (augmented) T x ntraps
        times filtered by trap masking information.
     indid <- idx[msk2 == 1 & alive.trues == 1, 1] # [AMG]
        Repeated set of individuals (augmented) but dead ones
         ignored after death
     repid <- idx[msk2 == 1 & alive.trues == 1, 2] # [AMG]
```

```
   Repeated sets of occasion numbers
trapid <- idx[msk2 == 1 & alive.trues == 1, 3]

##### [AMG] From here we get into main function again
getNN <- function(maxNN, G) {
    nG <- nrow(G)
    NN <- matrix(NA, nrow = nG, ncol = maxNN) # [AMG]
       Matrix creating neighbours
    for (i in 1:nG) {
        od <- sqrt((G[i, 1] - G[, 1])^2 + (G[i, 2] - G[,
            2])^2)
        NN[i, 1:maxNN] <- order(od)[1:maxNN] # [AMG]
            This chooses the maxNN neighbours around each
             filtered pixel. Arranged in increasing order
             of distance from i.
    }
    numnn <- rep(0, nrow(NN))
    for (i in 1:nG) {
        for (j in 1:ncol(NN)) {
            if (any(NN[NN[i, j], ] == i, na.rm = TRUE))
              next
            else NN[i, j] <- NA
        } # [AMG] We ensure that if one grid cell, say
           no. 4, has no. 6 as its nearest neighbour,
           then no. 6 also has to have no. 4 as its
           nearest neighbour. Else, proposal
           distribution will violate the MCMC properties
            of having a jumping probability in one of
           the directions.
        numnn[i] <- sum(!is.na(NN[i, ]))
        NN[i, 1:numnn[i]] <- NN[i, !is.na(NN[i, ])] # [
           AMG] Stores all the neighest neighbours and
           fills NA to those that do not meet the
           symmetric distribution criteria of the
           proposal distribution.
        # NN[i, numnn[i]+1:ncol(NN)]<-NA # [AMG] 13Aug14
            - Inserted to avoid duplication of grid cell
             numbers in NN[i, ]
    }
    if (min(numnn) < 3)
        cat("State-space grid has isolated or nearly-
           isolated cells increase maxNN or modify state
           -space",
            fill = TRUE)
    out <- list(NN = NN, numnn = numnn)
    return(out)
} # [AMG] Produces main sites nearest neighbours and a
   vector of number of nearest neighbours for each grid
   cell point
```

```
    hld <- getNN(maxNN, G)
    NN <- hld$NN
    numnn <- hld$numnn
    # [AMG] We now start allocating centers to individuals.
        First for known ones (centers1). Initial locations
        for starting parameters is chosen based on the
        average coordinates of actual captures of known
        individuals.
    centers1 <- rep(NA, nind)
    for (i in 1:nind) {
        tt <- t(as.matrix(Yaug[i, , ]))
        tt <- row(tt)[tt == 1]
        xxx <- as.matrix(traplocs[tt, ], ncol = 2, byrow =
            FALSE)
        av.coord <- colSums(xxx)/nrow(xxx)
        dvec <- as.vector(e2dist(matrix(av.coord, ncol = 2),
            G))
        centers1[i] <- (1:length(dvec))[dvec == min(dvec)
            ][1]
    }
    # [AMG] The augmented set of individuals are allocated
        centers (centers2) randomly (sampled with replacement
        ) over the grid points.
    centers2 <- sample(1:nG, M - nind, replace = TRUE)
    centers <- c(centers1, centers2) # [AMG] Is the net
        collection of pixel numbers of the known as well as
        augmented set of individuals.
    S <- G[centers, ] # [AMG] S has information of the exact
        coordinates of those grid cells occupied by
        individuals (both known and augmented). Number of
        rows is length(centers) = M. If the same center is
        occupied by more than one individual, the pixel
        numbers are allocated as, for example, 1045.1, 1045.2
        etc.

# [AMG] Code below sets all the initial values before
    analysis
    if (Msexsigma == 0)
        bsigma <- 2
    if (Msexsigma == 1)
        bsigma <- c(3, 3)

    update.theta <- FALSE
    if (is.na(theta)) {
        theta <- 0.75
        update.theta <- TRUE
    } # [AMG] If user is uncertain about detection function
        model, which is determined by theta, then they may
        state theta=NA. In which case, theta needs to be
```

```
    determined.
  loglam0 <- log(0.3)
  beta.behave <- 0
  beta.sex <- 0
  beta1 <- 0
  lam0 <- exp(loglam0)
  psi.sex <- mean(Xsex, na.rm = TRUE) # [AMG] Is the
    starting value of sex ratio. Produces a warning at
    the moment, without as.numeric. Included this earlier
     now.
  pz<-0.32
  # [AMG] Just a vector of 1s indicating all true
    individuals present (known and unknown). [AMG] 1 Aug
    2014 Changed psi to pz. pz becomes probz later. psi
    remains consistent with usage of psi in Royle et al
    (2009)
  z <- c(rep(1, nind), rbinom(nz, 1, pz))
  psi <- sum(z)/M   ## [AMG] Ensuring consistency in
    definition of psi.

########### [AMG] Attraction/repulsion code initialization
   ############
###### Setting starting values for these parameters
   #########
                #### Set initial condition as attraction (1=
                   attract, 0=repulse) ####
                omega <- 0
                ## Just a multiplier to ensure easy
                   computation in case we get into 0s##
        alpha0 <- 3.0
        ## Defining a column maximizer function ##
        apply.sum <- function(A) apply(A, 2, sum)
        ## Initial condition for "stretch" parameter - this
           parameter ranges from 0 to infinite ##
        stretch = 0.4

### A function to generate log(phi) ###
    likpos.fn<-function(omega, theta, bsigma, S, z, stretch)
       {
        bsigmas <- bsigma*(1/(stretch^2))
        ## Obtain index of real individuals
        real<-which(z %in% 1)
        ## Create matrix of all distances against
        allDist <- e2dist(G,S[real,])
        # This is the g(ij) function of all possible
           distances from conspecifics
        allDistfn <- exp(-(allDist^theta)*bsigmas)
        # Calculate all alphas, the expected encounter rates
           . Same dimensions as allDist and allDistfn (G x S
```

```
        [real,])
    alphai <- alpha0*(rowSums(allDistfn)-allDistfn[,1:
        length(real)])

    # Calculate the probabilities
    if (omega=1){
            # PATTalphai <- 1-exp(-alphai) # For
                attraction
            # sumalphai <- colSums(alphai)
            # cat(PATTalphai)
            alphavec <- diag(alphai[centers[real],])
            sumProbATT <- apply.sum(alphai)
            # This gives us absolute probabilities by
                normalizing
            PATTreals <- alphavec/sumProbATT
            logProbs <- log(PATTreals)
            logProbs
            # cat(PATTreals)
            # PATTreals[PATTreals==0] <- .Machine$double
                .xmin
    } else {
            # Now for repulsion
            alphaiInv <- 1/alphai
            alphavecInv <- diag(alphaiInv[centers[real
                ,])
            sumProbREP <- apply.sum(alphaiInv)
            # REPcounts <- 1/PATTreals
            PREPreals <- alphavecInv/sumProbREP
            # REPcounts/sum(REPcounts)
            # PREPreals[PREPreals==0] <- .Machine$double
                .xmin
            # Obtain the appropriate probabilities (
                attraction/repulsion)
            # Gprobs <- omega*(PATTreals)+(1-omega)*(
                PREPreals)
            # Take log of these "relative" probabilities
            logProbs <- log(PREPreals)
            logProbs
    }
}

#
    ##########################################################################


if (sum(sex.naflag) > 0)
    Xsex[sex.naflag] <- rbinom(sum(sex.naflag), 1, psi.
        sex) # modified 03/17/14 JFG - changed mean binom
         prob to mean of known pop. instead of .65
```

```
#
# obsolete
#   if (is.null(Xd)) {
#       Xd <- rep(1, nG)
#   }

  beta.den <- 0

#### [AMG] Creating the likelihood function
    lik.fn <- function(lpc, y1) {
        llvector.new <- -1 * exp(lpc) # [AMG] This is for
            all the possibilities(animals that are caught and
             the augmented set of animals that are exposed to
             occasions and traps). log[L(1-success)]=log[L(
            failure)]
        part2 <- exp(exp(lpc[y1])) - 1
        part2[part2 == 0] <- .Machine$double.eps # [AMG]
            Ensure that small numbers are created in part2 of
             the likelihood, instead of becoming 0. This is
            to ensure that those captures have a >0
            probability of capture.
        llvector.new[y1] <- llvector.new[y1] + log(part2) #
            [AMG] log[L(success)]
        llvector.new
    } # [AMG] Total likelihood vector. Log-probability of 0
       (failure) and log-probability of 1 (success)


    trapgridbig <- traplocs[trapid, ] # [AMG] Coordinates of
        trap ID repeated many times.
    y1 <- y == 1 # [AMG] Produces a y1 matrix of TRUE/FALSE
       depending on whether y=1/not.
    c1 <- (S[indid, 1] - trapgridbig[, 1])^2 # [AMG]
       Multiple entries of individual locations to trap
       locations - X distance
    c2 <- (S[indid, 2] - trapgridbig[, 2])^2 # [AMG]
       Multiple entries of individual locations to trap
       locations - Y distance
    gof.new <- gof.data <- rep(NA, (ni - burn)/skip) # [AMG]
        For goodness of fit stats
    out <- matrix(NA, nrow = (ni - burn)/skip, ncol = 18) #
       [AMG] Output matrix
    dimnames(out) <- list(NULL, c("bsigma", "sigma", "
       bsigma2",
        "sigma2", "loglam0","lam0", "beta.behave", "beta1(
           effort)", "beta.sex",
        "psi", "psi.sex", "Nsuper", "theta", "beta.density",
        "D","beta(habitat)", "omega", "kappa"))
    zout <- matrix(NA, nrow = (ni - burn)/skip, ncol = M)
```

```
Sout <- matrix(NA, nrow = (ni - burn)/skip, ncol = M)
# matrix to store likelihood output for Bayes factors
LLout<-matrix(NA, nrow = (ni-burn)/skip,ncol=2)

### [AMG] Begins the calculations
m <- 1
LM1 <- LM2 <- matrix(0, nrow = M, ncol = length(indid.LM
    )/M) # [AMG] ncol reflects trap effort (no. of
    occasions x ntraps(exposed)).
ones <- rep(1, ncol(LM1)) # [AMG] These are the traps
    and occasions that all augmented M individuals are
    exposed to.
if (Msexsigma == 0)
    lp.sigma <- Msigma * bsigma * (c1 + c2)^theta # [AMG
        ] Non sex-specific detection probability based on
         distance from traps.
if (Msexsigma == 1)
    lp.sigma <- bsigma[Xsex[indid] + 1] * (c1 + c2)^
        theta # [AMG] Sex-specific detection probability
        based on traps.
acc.count <- 0
delta <- 0.1 # [AMG] Tuning parameter for bsigma
    proposal distributions

################################ [AMG] MCMC loop begins here
    ############################
for (i in 1:ni) {
    cat("iter: ", i, fill = TRUE)
    lp <- loglam0 + Mb * beta.behave * prevcap - lp.
        sigma +
        Meff* beta1 * Xeff + Msex * beta.sex * Xsex[
            indid] # [AMG] Log-likelihood at current
            parameter values
    loglam0c <- rnorm(1, loglam0, 0.15) # [AMG] Proposal
        for loglam0
    lpc <- loglam0c + Mb * beta.behave * prevcap - lp.
        sigma +
        Meff*beta1 * Xeff + Msex * beta.sex * Xsex[indid
            ] # [AMG] Log-likelihood at proposed
            parameter values (change in loglam0)
    llvector <- lik.fn(lp, y1) # [AMG] log-likelihood
        vector with current parameters
    llvector.new <- lik.fn(lpc, y1) # [AMG] log-
        likelihood of proposed parameters
    LM1[aliveid == 1] <- llvector.new # [AMG] Proposal
        likelihood.
    LM2[aliveid == 1] <- llvector # [AMG] Current
        likelihood
```

```
if (runif(1) < exp(sum(((LM1[z == 1, ] - LM2[z == 1,
    ]) %*% ones)))) {
    loglam0 <- loglam0c
    lam0 <- exp(loglam0)
    llvector <- llvector.new
    lp <- lpc
    LM2 <- LM1
}
 # cat("ERROR TRAP 1") ### [AMG] These are error
    traps. More will follow. Uncomment these to find
     out where the problems lie in.
if (update.theta) {
    lp <- loglam0 + Mb * beta.behave * prevcap - lp.
       sigma +
        Meff* beta1 * Xeff + Msex * beta.sex * Xsex[
           indid]
    thetac <- rnorm(1, theta, 0.1)
    if (thetac >= 0.5 & thetac <= 1) {
        if (Msexsigma == 0)
          lp.sigmac <- Msigma * bsigma * (c1 + c2)^
             thetac
        if (Msexsigma == 1)
          lp.sigmac <- bsigma[Xsex[indid] + 1] * (c1
              +
            c2)^thetac
        lpc <- loglam0 + Mb * beta.behave * prevcap
           -
          lp.sigmac + Meff*beta1 * Xeff + Msex *
             beta.sex *
          Xsex[indid]
        llvector <- lik.fn(lp, y1)
        llvector.new <- lik.fn(lpc, y1)
        LM1[aliveid == 1] <- llvector.new
        LM2[aliveid == 1] <- llvector
        lphi1 <- likpos.fn(omega, thetac, bsigma, S,
            z, stretch)
        lphi2 <- likpos.fn(omega, theta, bsigma, S,
           z, stretch)

        if (runif(1) < exp(sum(rowSums(LM1[z == 1,
           ])+(Mattrep*lphi1)) - sum(rowSums(LM2[z
           == 1, ])+(Mattrep*lphi2)))){
          theta <- thetac
          lp.sigma <- lp.sigmac
          llvector <- llvector.new
          lp <- lpc
          LM2 <- LM1
          lphi2 <- lphi1
        }
```

```
    }
}
 # cat("ERROR TRAP 2")
if (Msexsigma == 0) {
    bsigmac <- exp(rnorm(1, log(bsigma), delta))
    lp.sigmac <- Msigma * bsigmac * (c1 + c2)^theta
    lpc <- loglam0 + Mb * beta.behave * prevcap - lp
        .sigmac +
         Meff*beta1 * Xeff + Msex * beta.sex * Xsex[
            indid]
    llvector.new <- lik.fn(lpc, y1)
    LM1[aliveid == 1] <- llvector.new
    lphi1 <- likpos.fn(omega, theta, bsigmac, S, z,
        stretch)
    lphi2 <- likpos.fn(omega, theta, bsigma, S, z,
        stretch)
                # cat(lphi1)
                # cat(lphi2)
    mh.ratio <- exp(sum(rowSums(LM1[z == 1, ])+(
        Mattrep*lphi1)) - sum(rowSums(LM2[z == 1, ])
        +(Mattrep*lphi2)) +
         (Mtel*(loglik.tel(bsigmac,Ytel,Msexsigma) -
            loglik.tel(bsigma,Ytel,Msexsigma))))
    # cat(mh.ratio)
    if (runif(1) < mh.ratio){
         lp.sigma <- lp.sigmac
         bsigma <- bsigmac
         llvector <- llvector.new
         lp <- lpc
         LM2 <- LM1
         lphi2 <- lphi1
         acc.count <- acc.count - 1
    }
    else {
         acc.count <- acc.count + 1
    }
}
 # cat("ERROR TRAP 3")
if (Msexsigma == 1) {
    bsigmac <- c(exp(rnorm(1, log(bsigma[1]), 2 *
        delta)),
         bsigma[2])
    lp.sigmac <- bsigmac[Xsex[indid] + 1] * (c1 + c2
        )^theta
    lpc <- loglam0 + Mb * beta.behave * prevcap - lp
        .sigmac +
         Meff*beta1 * Xeff + Msex * beta.sex * Xsex[
            indid]
    llvector.new <- lik.fn(lpc, y1)
```

```
            LM1[aliveid == 1] <- llvector.new
            mh.ratio<-                exp(sum(((LM1[z == 1, ] -
                LM2[z ==
                                                  1, ]) %*% ones)
                                                    ) +
                Mtel*(loglik.tel(bsigmac,Ytel,Msexsigma) -
                    loglik.tel(bsigma,Ytel,Msexsigma))  )

            if (runif(1) < mh.ratio){
                lp.sigma <- lp.sigmac
                bsigma <- bsigmac
                llvector <- llvector.new
                lp <- lpc
                LM2 <- LM1
            } # [AMG] If proposal for bsigma[1] works, then
                proceed to other parameters.
            else { # [AMG] Otherwise propose for bsigma[2].
            }
            bsigmac <- c(bsigma[1], exp(rnorm(1, log(bsigma
                [2]),
                2 * delta)))
            lp.sigmac <- bsigmac[Xsex[indid] + 1] * (c1 + c2
                )^theta
            lpc <- loglam0 + Mb * beta.behave * prevcap - lp
                .sigmac +
                Meff*beta1 * Xeff + Msex * beta.sex * Xsex[
                    indid]
            llvector.new <- lik.fn(lpc, y1)
            LM1[aliveid == 1] <- llvector.new
mh.ratio<-  exp(sum(((LM1[z == 1, ] - LM2[z ==1, ]) %*%
    ones)) +
             Mtel*(loglik.tel(bsigmac,Ytel,Msexsigma) -
                 loglik.tel(bsigma,Ytel,Msexsigma))  )
            if (runif(1) < mh.ratio) {
                lp.sigma <- lp.sigmac
                bsigma[2] <- bsigmac[2]
                llvector <- llvector.new
                lp <- lpc
                LM2 <- LM1
            }
        }
        cat("DELTA: ", delta, fill = TRUE)
        if (any(bsigma < 0)) {
            cat("negative bsigma....", fill = TRUE)
            return(0)
        }
         # cat("ERROR TRAP 4")
        if (Msex == 1) {
            beta.sexc <- rnorm(1, beta.sex, 0.1)
```

```
    lpc <- loglam0 + Mb * beta.behave * prevcap - lp
        .sigma +
         Meff*beta1 * Xeff + Msex * beta.sexc * Xsex[
            indid]
    llvector.new <- lik.fn(lpc, y1)
    LM1[aliveid == 1] <- llvector.new
    if (runif(1) < exp(sum(((LM1[z == 1, ] - LM2[z
       ==
         1, ]) %*% ones)))) {
         beta.sex <- beta.sexc
         llvector <- llvector.new
         lp <- lpc
         LM2 <- LM1
    }
}
 # cat("ERROR TRAP 5")
if (Mb == 1) {
    beta.behave.c <- rnorm(1, beta.behave, 0.1)
    lpc <- loglam0 + Mb * beta.behave.c * prevcap -
        lp.sigma +
         Meff*beta1 * Xeff + Msex * beta.sex * Xsex[
            indid]
    llvector.new <- lik.fn(lpc, y1)
    LM1[aliveid == 1] <- llvector.new
    if (runif(1) < exp(sum(((LM1[z == 1, ] - LM2[z
       ==
         1, ]) %*% ones)))) {
         beta.behave <- beta.behave.c
         llvector <- llvector.new
         lp <- lpc
         LM2 <- LM1
    }
}
# cat("ERROR TRAP 6")
if (Meff == 1){
    beta1c <- rnorm(1, beta1, 0.1)
    lpc <- loglam0 + Mb * beta.behave * prevcap - lp
        .sigma +
         Meff*beta1c * Xeff + Msex * beta.sex * Xsex[
            indid]
    llvector.new <- lik.fn(lpc, y1)
    LM1[aliveid == 1] <- llvector.new
    if (runif(1) < exp(sum(((LM1[z == 1, ] - LM2[z
       ==
         1, ]) %*% ones)))) {
         beta1 <- beta1c
         llvector <- llvector.new
         lp <- lpc
         LM2 <- LM1
```

```
            }
        }
        # cat("ERROR TRAP 7")
### Updating purely positional parameters ###
## Updating purely attraction/repulsion parameters ###
                if (Mattrep == 1){
                        # Updating omega (the parameter that
                            defines whether we have
                            attraction/repulsion)
                        omegac <- 1-omega
                        lphi1 <- likpos.fn(omegac, theta,
                            bsigma, S, z, stretch)
                        lphi2 <- likpos.fn(omega, theta,
                            bsigma, S, z, stretch)
                        cat("OMEGA Candidate = ",omegac)
                        # cat((lphi1))

                        cat("OMEGA Current = ", omega)
                        # cat((lphi2))
                        if (runif(1) < exp(sum(lphi1)-sum(
                            lphi2))){
                                omega <- omegac
                                lphi2 <- lphi1
                        }

                        # Updating kappa or "stretch
                            parameter" (the force of
                            attraction/repulsion between
                            individuals)

                        stretchc <- rnorm(1, stretch, 0.1)

                        ## 0 = infinitely sharp force
                        while (stretchc<0){
                                stretchc <- rnorm(1, stretch
                                    , 0.1)
                                }

                        lphi1 <- likpos.fn(omega, theta,
                            bsigma, S, z, stretchc)
                        if (runif(1) < exp(sum(lphi1-lphi2))
                            ){
                                stretch <- stretchc
                                lphi2 <- lphi1
                                }
                }

        # cat("ERROR TRAP 8")
########### [AMG] Critical point of intervention for
```

```
attraction-repulsion code ##########

    if (Mattrep==1){
        # Create a temporary vector of existing real
            individuals
        tmp.exist <- z
        # Segregate those individuals that are
            undetected and flip reals to ghosts and
            vice versa
        tmp.undetected <- c(rep(1,nind),1-z[(nind+1)
            :M])
        # Update real/ghost status to undetected
            animals sequentially
        # cat("ERROR TRAP 8a")
        for (j in (nind+1):M){
                zc <- z
                zc[j] <- tmp.undetected[j]
                lphi1 <- likpos.fn(omega, theta,
                    bsigma, S, zc, stretch)
        if (runif(1) < exp(sum(rowSums(LM2[zc == 1,
            ])+(Mattrep*lphi1)) - sum(rowSums(LM2[z
            == 1, ])+(Mattrep*lphi2)))) {
          z <- zc
          lphi2 <- lphi1
          }
                }
        psi <- rbeta(1, 1 + sum(z), 1 + M - sum(z))
        } else {

      probz <- exp(rowsum(llvector[indid > nind],
          indid[indid >
      nind])) # [AMG] Here indid[indid>nind] is the
          grouping variable.
        pz <- (probz * psi)/(probz * psi + (1 - psi)
            ) # [AMG] Is nothing but {P(uncaught|real
            )*P(real)}/{[P(uncaught|real)*P(real)]+P(
            ghost)}, where P(Ghost)=P(1-real)
        z[(nind + 1):M] <- rbinom(M - nind, 1, pz) #
            [AMG] New individuals come into the
            picture. Here, each individual's
            probability of being selected into the
            real population set depends on its
            existing position, thereby giving it a
            certain likelihood for getting caught and
             its probability of being real. So each
            individual has a different probability (
            so pz is a vector)
        psi <- rbeta(1, 1 + sum(z), 1 + M - sum(z))
            # psi gets updated here.
```

```
}
# cat("ERROR TRAP 9")

if (!is.null(Xsex)) {
    tmp.sex <- Xsex
    tmp.sex[sex.naflag] <- 1 - Xsex[sex.naflag] # [
       AMG] Flip the sex.
    if (Msexsigma == 0)
        lp.sigmac <- Msigma * bsigma * (c1 + c2)^
           theta
    if (Msexsigma == 1)
        lp.sigmac <- bsigma[tmp.sex[indid] + 1] * (
           c1 +
          c2)^theta
    lpc <- loglam0 + Mb * beta.behave * prevcap - lp
       .sigmac +
        Meff*beta1 * Xeff + Msex * beta.sex * tmp.
           sex[indid]
    llvector.new <- lik.fn(lpc, y1)
    lik.othersex <- exp(rowsum(llvector.new, indid))
    lik.sex <- exp(rowsum(llvector, indid))
    prior.curr <- (psi.sex^Xsex) * ((1 - psi.sex)^(1
       -
        Xsex)) # [AMG] Only place we use priors
           other than flat priors.
    prior.cand <- (psi.sex^tmp.sex) * ((1 - psi.sex)
       ^(1 -
        tmp.sex))
    swtch <- sex.naflag & (runif(M, 0, 1) < ((lik.
       othersex *
        prior.cand)/(lik.sex * prior.curr)))
    Xsex[swtch] <- 1 - Xsex[swtch] # [AMG] Switch
       the sex of those individuals whose sex is
       unknown and likelihood x prior of candidate
       sex is more appropriate than likelihood x
       prior of current sex.
    psi.sex <- rbeta(1, 0.1 + sum(Xsex[z == 1]), 0.1
       +
        sum(z) - sum(Xsex[z == 1])) # [AMG] psi.sex
           gets updated here.
    if (Msexsigma == 0)
        lp.sigma <- Msigma * bsigma * (c1 + c2)^
           theta
    if (Msexsigma == 1)
        lp.sigma <- bsigma[Xsex[indid] + 1] * (c1 +
           c2)^theta
    lp <- loglam0 + Mb * beta.behave * prevcap - lp.
       sigma +
        Meff*beta1 * Xeff + Msex * beta.sex * Xsex[
```

```
            indid]
      llvector.new <- lik.fn(lp, y1)
      LM1[aliveid == 1] <- llvector.new
      llvector <- llvector.new
      LM2 <- LM1
} # [AMG] updation of psi.sex occurs here.
# cat("ERROR TRAP 10")
# newcenters <- trunc(runif(M, 0, numnn[centers])) +
    1 # [AMG] This is where new centers come into
    the picture
newcenterchoice <- ceiling(runif(M, 0, numnn[centers
    ])) # [AMG] Changed based on SPACECAP code

newcenters <- NN[cbind(centers, newcenterchoice)] #
    [AMG] Allocates a new center among choices of
    centers
qnew <- 1/numnn[centers] # [AMG] Probability of new
    neighbour being selected
qold <- 1/numnn[newcenters] # [AMG] Probability of
    going back to old neighbour
Sc <- G[newcenters, ] # [AMG] Coordinates of new
    centers
c1c <- (Sc[indid, 1] - trapgridbig[, 1])^2
c2c <- (Sc[indid, 2] - trapgridbig[, 2])^2
if (Msexsigma == 0)
     lp.sigmac <- Msigma * bsigma * (c1c + c2c)^theta
if (Msexsigma == 1)
     lp.sigmac <- bsigma[Xsex[indid] + 1] * (c1c +
        c2c)^theta
lpc <- loglam0 + Mb * beta.behave * prevcap - lp.
    sigmac +
     Meff*beta1 * Xeff + Msex * beta.sex * Xsex[indid
        ]
llvector.new <- lik.fn(lpc, y1)
LM1[aliveid == 1] <- llvector.new
lphi1 <- likpos.fn(omega, theta, bsigma, Sc, z,
    stretch)
likdiffS <-c(rep(0,M))
likdiffS[z==1] <- (rowSums(LM1[z==1,])+(Mattrep*
    lphi1)) - (rowSums(LM2[z==1,])+(Mattrep*lphi2))
# likdiff[z == 0] <- 0 ### [AMG] This is unnecessary
     if z==1 is incorporated in previous statement.
    All ghosts are simply made to jump.
logprior.new <- Xd[newcenters] * beta.den
logprior.old <- Xd[centers] * beta.den
likdiffS <- likdiffS + (log(qold)-log(qnew)) + (
    logprior.new -
     logprior.old) # [AMG] qold/qnew accounts for
        ratio of (probability of going back to old):(
```

```
            probability of new neighbour being selected).
             Likelihood ratio x ratio of jumping
            probabilities x ratio of prior probabilities.
      accept <- runif(M) < exp(likdiffS)
      cat("accept rate: ", mean(accept), fill = TRUE) # [
         AMG] Provides the mean number of individuals that
          jump
      S[accept, ] <- Sc[accept, ]
      centers[accept] <- newcenters[accept]
      c1 <- (S[indid, 1] - trapgridbig[, 1])^2
      c2 <- (S[indid, 2] - trapgridbig[, 2])^2
      LM2[accept, ] <- LM1[accept, ]
      lphi2 <- likpos.fn(omega, theta, bsigma, S, z,
         stretch)
      # cat("ERROR TRAP 11")
      real <- which(z %in% 1)
      S[accept,]
if(Mss==1){
      beta.den.c <- rnorm(1, beta.den, 0.25)
      numerator.new <- exp(Xd * beta.den.c)
      loglik.new <- Xd[centers] * beta.den.c - log(sum(
         numerator.new))
      numerator.old <- exp(Xd * beta.den)
      loglik.old <- Xd[centers] * beta.den - log(sum(
         numerator.old))
      if (runif(1) < exp(sum(loglik.new) - sum(loglik.old)
         )) {
          beta.den <- beta.den.c
      } # [AMG] Here beta.den is updated
      }
      if (Msexsigma == 0)
          lp.sigma <- Msigma * bsigma * (c1 + c2)^theta
      if (Msexsigma == 1)
          lp.sigma <- bsigma[Xsex[indid] + 1] * (c1 + c2)^
             theta
      lp <- loglam0 + Mb * beta.behave * prevcap - lp.
         sigma +
          Meff*beta1 * Xeff + Msex * beta.sex * Xsex[indid
             ]
      llvector <- lik.fn(lp, y1)
      LM2 <- LM1

      # cat("ERROR TRAP 12")
      ###
      ###
      ###
      ###
      if ((i > burn) & (i%%skip == 0)) {
          sigma <- sqrt(1/(2 * bsigma))
```

```
        if (Msexsigma == 0) {
            sigmatmp <- c(sigma, sigma)
            bsigmatmp <- c(bsigma, bsigma)
        }
        else {
            sigmatmp <- sigma
            bsigmatmp <- bsigma
        } # [AMG] bsigma is converted to sigma as in
            Borchers and Efford (2008)
        logmu <- loglam0 + Mb * beta.behave * prevcap -
            lp.sigma +
            Meff*beta1 * Xeff + Msex * beta.sex * Xsex[
                indid]
        mu <- (1 - exp(-exp(logmu))) * z[indid] # [AMG]
            Mean expectation of encounters for gof
            calculations

#
    ##########################################################################

        # JFG Added Mar '14. These lines calculate the
            probability of observing the data given the
            parameter values
        # Make an array that stores P(1-y[i,j,k])
        mu.array <- array(1-mu, dim=c(max(indid), max(
            repid), max(trapid)))
        # Change values to P(y[i,j,k]) when/where
            captures occurred
        mu.array[as.matrix(captures[,c(2,3,1)])] <- 1 -
            mu.array[as.matrix(captures[,c(2,3,1)])]
        # Convert to log scale
        logmu.array = log(mu.array)
        # How many infinite values?
        n.nolik = sum(is.infinite(logmu.array))
        # Replace -Inf values with sufficiently small
            number
        logmu.array[is.infinite(logmu.array)] = .Machine
            $double.min.exp
        # Report sum of log-likelihoods and the number
            of problem pts.
        LLout[m,] = c(sum(logmu.array), n.nolik)
#
    ##########################################################################


        newy <- rbinom(length(mu), 1, mu)
        gof.stats <- cbind(y, newy, mu)
        gof.stats <- aggregate(gof.stats, list(indid),
            sum)
```

```
            gof.data[m] <- sum((sqrt(gof.stats[, 2]) - sqrt(
                gof.stats[,
                  4]))[z == 1]^2)
            gof.new[m] <- sum((sqrt(gof.stats[, 3]) - sqrt(
                gof.stats[,
                  4]))[z == 1]^2)
            density <- sum(z)/totalarea
            zout[m, ] <- z
            Sout[m, ] <- centers
            out[m, ] <- c(bsigmatmp[1], sigmatmp[1],
                bsigmatmp[2],
                 sigmatmp[2], log(lam0), lam0, beta.behave,
                    beta1, beta.sex,
                 psi, psi.sex, sum(z), theta, beta.den,
                    density,NA, omega, stretch)
            print(out[m, ])
            if (m%%dumprate == 0) {
            }
            m <- m + 1
        }
    }
    # JFG 04/17/13 edited so that psi.sex is included if
        Msexsigma = 1.
    parms.2.report <- c(TRUE, TRUE,
                        Msexsigma == 1,
                        Msexsigma == 1, TRUE, TRUE, Mb == 1,
                        Meff ==1, Msex == 1, TRUE, (Msex ==
                            1|Msexsigma==1),
        TRUE, update.theta, Mss == 1, TRUE,FALSE)
    # JFG edited to return likelihood object created above
    out <- list(mcmchist = out, G = G, Gunscaled = Gunscaled
        ,
         traplocs = traplocs, Sout = Sout, zout = zout,
            likelihood = LLout, statespace = statespace,
         gof.data = gof.data, gof.new = gof.new, call = call,
         parms2report = parms.2.report)
    class(out) <- c("scrfit", "list")
    return(out)
}
```

7.7.2 Attraction-repulsion SECR model - R code for simulation

```
#### APPENDIX 7.7.2: Attraction-repulsion SECR model - R
    code for simulation ####

sim.data <-
function (N=50, sigma = 0.3, loglam0 = log(0.35),
```

```
        K = ncol(traps)-3,
    statespace ,
    traplocs ,
    Xss = NULL, alpha1 = 0,
    coord.scale = 1000,  Ntel=0, nfixes=0, theta=1, omega=1,
        stretch=1)
{
        alpha0=3
    bsigma = 1/(2*((stretch*sigma)^2))
    statespace[, 1:2] <- statespace[, 1:2]/coord.scale
    traplocs <- traplocs/coord.scale
    ntraps <- nrow(traplocs)
    G <- statespace[, 1:2]
    goodbad <- statespace[, 3]
    G <- G[goodbad == 1, ]
    nG <- nrow(G)
    if (is.null(Xss))
        Xss <- rep(1, nG)
    #cat("ERROR TRAP 1")
    centers <- c(rep(0,N))
    # Generate an equiprobability surface to begin with
    Equiprob <- 1/nG
    Gprobs <- c(rep(Equiprob,nG))
    #cat("ERROR TRAP 2")
    #sprobs <- exp(alpha1 * Xss[goodbad == 1])
    #sprobs <- sprobs/sum(sprobs)


    # Introduce a population of individuals sequentially
    for (count in 1:N){

        # Probability of including an individual is driven
            by resources.
                # Introduce one individual at a time in this
                    attraction/repulsion scheme
                intro <- rmultinom(1, 1, prob=Gprobs)
                # Check which grid cell did it occupy
                centers[count] <- which(intro %in% 1)
                # Find out distances from of each grid point
                    with the individuals introduced
                cat("Introducing individual", count, fill=
                    TRUE)
                GridDistance <- e2dist(G[centers[1:count],],
                    G)
                # Find out the exponential of these
                    distances based on declining function
                GridDistfn <- exp(-(GridDistance^theta)*
                    bsigma)
                # Generate the poisson counts at various
```

```
                    grid points
                Gcounts <- alpha0*colSums(GridDistfn)
                Gcounts[Gcounts==0] <- .Machine$double.xmin
                PATTcounts <- Gcounts/max(Gcounts)
                PATTcounts[PATTcounts==0] <- .Machine$double
                    .xmin
                REPcounts <- 1/PATTcounts
                PREPcounts <- REPcounts/max(REPcounts)
                PREPcounts[PREPcounts==0] <- .Machine$double
                    .xmin
                # Generate probability surface
                Gprobs <- omega*(PATTcounts)+(1-omega)*(
                    PREPcounts)
                Gprobs[Gprobs==0] <- .Machine$double.xmin
                # if (count==250){
                        # cat(Gprobs)
                        #cat(1-exp(-(Gcounts)))
                #         plot(1:nG, Gprobs)
                #}
  }
  cat("Realized number of individuals=", N)
  #cat("ERROR TRAP 4")
  # centers <- sample(1:nG, N, replace = TRUE, prob =
      sprobs)
  S <- G[centers, ]
  #par(mar=c(3,3,3,6))
  #spatial.plot(statespace, Xss, cx=2, add=FALSE)
  # spatial.plot(statespace, Gprobs, cx=2, add=FALSE)
  # points(S, pch=20)
     # par(mfrow=c(1,2))
  # scatterplot3d(x=G[,1], y=G[,2], z=Gprobs)
   plot(S)
  Y <- array(NA, c(N, K, ntraps))
  #Ytel<- matrix(NA, nrow=nG, ncol=Ntel)
  Ytel<-NULL
  sex<-rbinom(N,1,.5)  # sex ratio = .5
  #cat("ERROR TRAP 5")
  if(length(sigma)==2){
    sigma.v<- sigma[sex+1]

}
else{
    sigma.v<-rep(sigma,N)
    }

  MASK <- as.matrix(traps[,4:ncol(traps)])
  for (i in 1:N) {
      lp <- loglam0 - (1/(2*sigma.v[i]^2)) * ((S[i, 1] -
          traplocs[, 1])^2 +
```

```
            (S[i, 2] - traplocs[, 2])^2)
        lambda<- exp(lp)
        pcap <- 1 - exp(-1 * lambda)
        for (t in 1:K) {
            ttt <- rbinom(ntraps, 1, pcap)
            Y[i, t, 1:ntraps] <- ttt*MASK[,t]
        }

        if(Ntel>0 & i<= Ntel){
            ## Xss needs in encounter model too
            log.lambda<-
-(1/(2*sigma.v[i]^2)) * ((S[i, 1] - G[, 1])^2 + (S[i, 2] - G
    [, 2])^2) +alpha1*Xss
           # rsf.probs<- exp(log.lambda)/sum(exp(log.lambda)
              )
           # Ytel[,i]<- rmultinom(1,nfixes,rsf.probs)
           Ytel<- rbind(Ytel,cbind( rnorm(nfixes,S[i,1],
              sigma.v[i]),
                                    rnorm(nfixes,S[i,2],
                                       sigma.v[i]),i))
        }

    }

    # attr(Ytel,"sex")<- sex[1:Ntel]
     #cat("ERROR TRAP 6")

     ndet <- apply(Y, c(1, 2), sum)
     ind.captured <- apply(ndet, 1, sum)

     cat("total captures: ", sum(Y), fill = TRUE)
     cat("total individuals: ", sum(ind.captured > 0), fill =
         TRUE)
     #cat("ERROR TRAP 7")
     Y <- Y[ind.captured > 0, , ]
     sex<-sex[ind.captured>0]
     #MASK <- matrix(1, ntraps, K)
     #MASK <- as.matrix(traps[,4:ncol(traps)])
     traplocs <- traplocs * coord.scale
     trapout <- cbind(1:nrow(traplocs), traplocs, MASK)
     dimnames(trapout) <- list(1:nrow(traplocs), c("LOC_ID",
        "X_Coord",
         "Y_Coord", as.character(1:ncol(MASK))))
     #cat("ERROR TRAP 8")
##     write.csv(trapout, "traps.csv", row.names = FALSE)
     o <- NULL
     for (i in 1:dim(Y)[1]) {
         z <- Y[i, , ]
         o <- rbind(o, cbind(i, row(z)[z > 0], col(z)[z > 0])
```

```
            )
    }
    y <- cbind(o[, 3], o[, 1], o[, 2])
    y <- y[order(y[, 1]), ]
    dimnames(y) <- list(1:nrow(y), c("trapid", "individual",
        "occasion"))
#     write.csv(y, "captures.csv", row.names = FALSE)
#     write.csv(statespace, "statespace.csv", row.names =
    FALSE)
    # list(Y = Y, MASK = MASK, traplocs = traplocs, Xss = Xss
        , Ytel=Ytel,
            #Xsex=sex, statespace=statespace)
    list(captures=y, traps=traps, statespace=statespace)
}
```

Chapter 8

Conclusions

In the preceding chapters of the thesis I demonstrated that the assessment of parameters in typical, empirical source-sink dynamics (Pulliam, 1988) or metapopulation dynamics (Hanski & Hanski, 1999) dynamics is not a straightforward puzzle to solve, especially when they involve animals which are wide-ranging, occur at low densities and are often cryptic - epitomized by the emblematic tiger. However, if this larger biological problem can be broken down into specific parameter estimation goals of (1) assessing state variables and (2) vital rates (Williams *et al.*, 2002), the specific findings of this thesis and the related studies in the appendices, successfully resolves the methodological issues pertinent to (1).

The methodological 'toolbox' emerging out of this thesis can directly be used to assess the extent of patchiness (Levin, 1976; May, 1978) via the overdispersion parameter k within large carnivore landscapes. Karanth *et al.* (2011b) demonstrates the existence of such an environment using tiger occupancy probabilities, which are firmly tied to the available prey base at the landscape scale and consequently advances the source-sink view of the larger tiger metapopulation dynamics problem. The findings, of course, will naturally lead to the next research step of estimating critical vital rates relevant to source-sink dynamics (Pulliam, 1988; Watkinson & Sutherland, 1995) in such real world large carnivore landscapes.

More specifically, in *Chapter 2*, some of the key milestones of spatial capture-recapture methods. Given its explosive development, we will most defintely see many extensions to these basic models in the future. We have already witnessed the application of SECR models on data from camera trap surveys of large cats (Royle *et al.*, 2009b,a), mist-netting surveys of birds (Borchers & Efford, 2008), acoustic signal surveys of birds(Dawson & Efford, 2009), hair-snare surveys of large carnivores (Gardner *et al.*, 2010b) and for all kinds of search-encounter surveys (Royle *et al.*, 2011b). We have also seen useful extensions to improve estimation of density from SECR models by incorporating covariates (Sollmann *et al.*, 2011). More recently, we witnessed the exploration of resource selection functions and landscape connectivity in SECR models (Royle *et al.*, 2013a). I posit that the utility of SECR models can

go far beyond the objective of merely estimating animal density. The SECR approach now provides a framework to investigate how behavioural mechanisms drive population dynamics in wild animals. This has long been a big knowledge gap in ecology (May, 1999) but has so far been very challenging to confront. SECR approaches have opened doors by providing unifying models that estimate parameters about behavioural processes (such as home range sizes) with population processes (such as survival and recruitment), amidst the field problem of sampling processes (Thompson, 2002; Williams *et al.*, 2002). For example, telemetry studies have been widely used in ecology to study individuals and obtain answers to questions about size of breeding female home ranges and how they change over time (see Goodrich *et al.*, 2010, for an example with tigers). These behavioural mechanisms, almost certainly, influence vital rates, and consequently population size. But population level parameters are usually estimated by completely different approaches, such as camera-trapping (see Karanth & Nichols, 1998; Karanth *et al.*, 2006; O'Connell *et al.*, 2010b, for an example with tigers). Now, with the advent of SECR models, we have a convincing way of integrating knowledge from individual-based studies and population-based studies with a combined model. Consequently, I see enormous potential for the use of sophisticated spatial point processes (Møller & Waagepetersen, 2004) within SECR models to answer such interesting questions, not only with tigers but with other carnivore species as well.

In the context of tigers, I see tremendous potential of their application in tiger source populations, as has been highlighted as a critical conservation need of the hour (Walston *et al.*, 2010). Rigorously designed SECR experiments, can also yield vital information on the movement rate parameter, σ. This parameter can be of direct use to conservation managers practicing science-based management practice in the context of structured decision making (Martin *et al.*, 2009). This is because, when conservation interventions are to be assessed on large carnivore populations and carrying capacity changes, say an increase of decrease in prey availability, I would expect a time lag for carnivores to respond in terms of their changes in densities. However, changes in the parameter σ could be more immediately relevant to assess whether conservation interventions are working or not.

Similarly, the most reliable way is to accurately estimate tiger prey densities to first determine potential source populations, especially where direct killing of tigers (either due to poaching or retaliatory killings) may be rampant (Karanth *et al.*, 2004; Walston *et al.*, 2010; Sharma *et al.*, 2014). But in areas where prey densities are low (<3 animals/km^2, most parts of south east Asia), and standard techniques such as distance sampling using line transects (Karanth *et al.*, 2004; Buckland *et al.*, 2001; Kumar, 2011) are not feasible to employ, an alternative technique is required. *Chapter 3* attempts to address this need and we estimate densities of four large

ungulate species from simple, cost-effective sign surveys using the 'AOS' model in a region where ungulate densities are relatively low. The results from this experiment show that it is important for investigators to incorporate biological knowledge of a target species in the design of such a sign survey. Of critical importance is the reasonably accurate specification of expected daily range sizes of animals to establish cell sizes and consideration of the length of spatial replicates. However, simulation results show that although underestimation or overestimation of values assumed for radius of activity can bias estimates of density, the direction of this bias appears predictable. This consistency in direction of bias makes this approach a useful tool, at least as an 'index', for monitoring ungulate densities across space and time because detection probability is still estimated. On a related note, if accurate information on R, or the number of cells expected to be visited by each animal group is based on prior knowledge (e.g., via radio-telemetry), this information can be profitably incorporated into the survey design. We emphasize that the AOS model approach works better when expected densities of ungulates are low enabling the capturing of heterogeneous detection rates among cells that is critical for the model to work. Another large site in the same region Rayar (2010) has also confirmed the reliability of this approach by including additional covariates. This approach can also potentially be extended from sign surveys to other animal detection techniques such as camera traps, scent stations, and track-plates, which are currently used to assess populations of other animal taxa (e.g., Nag, 2008). I was quite excited to use the findings from this ungulate study to a completely different study system and context. By including two variants of the basic model (Royle & Nichols, 2003), this approach was successfully applied to assess better prevalence in wildlife diseases in a typical host-pathogen system (Lachish *et al.*, 2012, Appendix K).

I next moved to the problem of taking cutting-edge methods to practicing field conservationists, ecologists and park managers by developing a software program called SPACECAP, described in *Chapter 4* of this thesis. Increasingly, the use of Bayesian inference methods (Gelman *et al.*, 1996; Bolker, 2008; Royle & Young, 2008) are becoming popular due to their inherent capacity to solve complex problems with powerful MCMC algorithms (Tierney, 1994), computational power (Venables & Ripley, 2002) and flexibility to tailor the modelling to account for particular ecological and sampling situations (e.g., Gardner *et al.*, 2010a; Sollmann *et al.*, 2011). However, these methods seem very distant from practicing field biologists. This situation results in a large time lag between development of methods to their implementation. This program aims to address this conservation need by making spatially-explicit capture recapture models accessible to most with the development of a user-friendly interface and providing an output in an easily reportable manner.

Although a typical analysis using SPACECAP now takes much longer than a comparable analysis using DENSITY (Efford, 2004), we do not view this as a serious practical disadvantage, in the case of sampling tigers, considering that it takes 45–65 days of field work and several weeks of organising and checking the data prior to analysis. We envisage that the modeling framework of Royle *et al.* (2009a) used in SPACECAP for analysing animal detection data, especially from camera trap surveys, provides the opportunity to incorporate auxiliary information in the form of 'capture' data on the same individuals in the form of photographs taken by tourists, animals killed in conflict situations or by poachers and even live tracking of animals tagged by radio-frequency identification tags, radio-collars or satellite-collars to improve accuracy and precision of the abundance estimates. The explicit recognition of the fact that most animals have biologically defined home ranges in the statistical modelling allows for a more realistic characterisation of the spatial ecology of the constituent individuals within a population. This realistic biological treatment, in turn, is likely to provide more reliable estimates of animal abundance and density. Given the simplicity of its use, this package has already been recommended as one of the tiger density estimation packages by the National Tiger Conservation Authority, Govt. of India.

In *Chapter 5*, we confronted the problem of estimating tiger density using data from more than one survey: (1) a photographic capture-recapture survey and (2) a fecal DNA capture-recapture survey by proposing three plausible models and described two ways of updating our knowledge on the central parameter (tiger density) of interest. The motivation was to strengthen our inference on tiger density by combining information and increase precision of estimates. We also demonstrate two approaches to combining information in data sets and our result emphasizes the point that it is worth all the effort to incorporate prior knowledge, where available, in the design and conduct of ecological studies, more generally. We expect that such statistical approaches which favor combining multiple sources of data, such as the integrated population models (Abadi *et al.*, 2010b,a) and this approachs, present the way forward in density estimation of rare and elusive carnivores. Spatially-explicit capture-recapture models are now being increasingly used in density estimation for various taxa and survey methods (Borchers & Efford, 2008; Royle *et al.*, 2009a; Gardner *et al.*, 2009, 2010a; Royle *et al.*, 2011b). This study is the first example where data sets from multiple sources have been brought together to bear on the density estimation question at hand using SECR models. This approach of combining information on common parameters from multiple sources, without having to assume that the spatial arrangement of animals will remain the same between the two survey processes, has broad ecological applications. For instance, during the same year but in different seasons, overall animal density and abundance could still

be the same, but resource-patches could vary seasonally (Macdonald & Johnson, In Press). In the example with tigers, and more generally for strongly territorial carnivores, it might be worth accounting for the fact that carnivore spatial arrangements are likely to be stable over long time periods.

I shift scales in *Chapter 6* to address an issue connecting abundance estimation at the landscape and the tiger metapopulation scale. Here, we statistically and empirically investigate inferences from index-calibration experiments, which are popular methods used for assessing animal abundance, and more specifically, tiger abundance at landscape scales. The exact roles of key sampling processes (imperfect detection p, spatial heterogeneity in sampling V and sampling uncertainties of population estimates Q) on the coefficient of determination R^2 generated by simple index-calibration experiments (Eberhardt & Simmons, 1987) are revealed. Because we have now made clear that any combination among these three uncontrollable sampling processes can lead to high or low R^2 values, we provide a major step forward to address the question of how field studies can exhibit such a wide variation in degree of inferential success in such index-calibration experiments. Our study shows that such direct index-calibration exercises are worth considering only if p^* is very high and non-varying, V is very high, Q is very low and μ is also low. Meeting these assumptions are highly unrealistic in the field. We therefore strongly recommend against wasting limited conservation and scientific resources in such direct index-calibration exercises. We recommend only those models which explicitly distinguishes the observation process parameter (especially p) from the state process, for example, see Conroy *et al.* (2008) and more recently, Chandler & Royle (2013) which propose alternative ways of estimating abundance at large geographic scales by explicitly estimating detection probability within a joint model. More recently, Efford & Fewster (2013), in their study of the skink *Oligosoma infrapunctatum*, demonstrate a possible way of estimating abundance at large spatial scales by using suitable covariates within a spatial capture-recapture framework. It is still relevant, however, that even with these newer techniques the sampling design should consist of a large V, a Q very close to 1 (using rigorous methods of estimating densities), and applying accurate covariates to describe the detection probability parameter to reduce its unaccounted for heterogeneity. Although, this development was largely motivated by the tiger studies discussed here, the parameters that determine the R^2 in simple linear regression problems is relevant to a wide variety of macroecological studies from examinations of speciation or diversity patterns and processes (Signor & Lipps, 1982; Tarver *et al.*, 2007) to identifying disease hotspots (McClintock *et al.*, 2010; Lachish *et al.*, 2012), and more readily, to questions on long-term population dynamics (de Valpine & Hastings, 2002; De Valpine, 2003). Further, simple regression models are used widely across a range of disciplines (covering the

entire spectrum from the sciences to social sciences) to assess linear correlations between two quantities of interest. The implications of our study may find relevance to similar sampling situations in other disciplines too.

Finally, but although in a very primitive stage of model development, I extend spatial capture-recapture models to permit inferences about behavioural plasticity in carnivores in *Chapter 7*. In specific, the model estimates a parameter ω, which informs us whether the environment is of an attractive or repulsive type. Another parameter κ also informs us about the degree of interaction between conspecific interactions. In addition to all the parameters associated with estimating abundance, these two parameters can provide us with vital keys to understanding how behavioural processes shape population dynamic parameters, which remains to be an ecological mystery (May, 1999).

Inherently tied to these methodological issues around estimating basic parameters in assess tiger metapopulation dynamics are also issues related to human-wildlife conflict around the edges of source tiger populations. In this context, I got the opportunity to explore methodological issues related to this conservation interface, which resulted in the following outputs: Karanth *et al.* (Appendix J, Appendix M, Appendix N, respectively 2012, 2013a,b,c). In particular, I explored the use of spatial krigging and model selection and applied them to such problems.

Most of the questions within my thesis focussed on the estimation of state variables, such as abundance and occupancy of tigers and their prey, at small and large scales. Indeed, this was the limit set on the sort of questions posed and estimating vital rates at the metapopulation level is still some time away. Interestingly, however, during the course of the study, anecdotal evidences of long range tiger dispersal events were observed and reported Patil *et al.* (2011, Appendix I) and providing hope that simple, but rigorous cost-effective camera traps surveys can yield vital information about tiger metapopulation dynamics, which seemed inconceivable in the past.

Appendix A

A tiger cannot change its stripes: using a three-dimensional model to match images of living tigers and tiger skins

Downloaded from rsbl.royalsocietypublishing.org on 8 June 2009

biology letters

A tiger cannot change its stripes: using a three-dimensional model to match images of living tigers and tiger skins

Lex Hiby, Phil Lovell, Narendra Patil, N. Samba Kumar, Arjun M. Gopalaswamy and K. Ullas Karanth

Biol. Lett. 2009 **5**, 383-386 first published online 11 March 2009
doi: 10.1098/rsbl.2009.0028

References This article cites 4 articles, 1 of which can be accessed free
http://rsbl.royalsocietypublishing.org/content/5/3/383.full.html#ref-list-1

Subject collections Articles on similar topics can be found in the following collections

ecology (628 articles)

Email alerting service Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Biol. Lett.* go to: **http://rsbl.royalsocietypublishing.org/subscriptions**

This journal is © 2009 The Royal Society

Downloaded from rsbl.royalsocietypublishing.org on 8 June 2009

biology letters
Population ecology

Biol. Lett. (2009) **5**, 383–386
doi:10.1098/rsbl.2009.0028
Published online 11 March 2009

A tiger cannot change its stripes: using a three-dimensional model to match images of living tigers and tiger skins

Lex Hiby[1,*], Phil Lovell[2], Narendra Patil[3], N. Samba Kumar[3], Arjun M. Gopalaswamy[3] and K. Ullas Karanth[3]

[1]*Conservation Research Ltd, 110 Hinton Way, Great Shelford, Cambridge CB22 5AL, UK*
[2]*Sea Mammal Research Unit, Gatty Marine Laboratory, University of St Andrews, St Andrews, Fife KY16 8LB, UK*
[3]*Wildlife Conservation Society—India Programme, Centre for Wildlife Studies, 1669, 31st Cross, 16th Main, Banashankari 2nd Stage, Bangalore 560 070, India*
**Author for correspondence (lexhiby@gmail.com).*

The tiger is one of many species in which individuals can be identified by surface patterns. Camera traps can be used to record individual tigers moving over an array of locations and provide data for monitoring and studying populations and devising conservation strategies. We suggest using a combination of algorithms to calculate similarity scores between pattern samples scanned from the images to automate the search for a match to a new image. We show how using a three-dimensional surface model of a tiger to scan the pattern samples allows comparison of images that differ widely in camera angles and body posture. The software, which is free to download, considerably reduces the effort required to maintain an image catalogue and we suggest it could be used to trace the origin of a tiger skin by searching a central database of living tigers' images for matches to an image of the skin.

Keywords: tiger; camera trap; pattern matching

1. INTRODUCTION

Camera traps are being used to monitor tiger numbers in Asia's remaining forests (Karanth *et al.* 2004, 2006). The traps detect and photograph animals moving along the forest tracks. Non-occluded sections of their coat patterns can be used to individually identify the tigers photographed. The resulting individual 'capture histories' are used in capture–recapture models (Williams *et al.* 2002) to estimate population size and other parameters relevant to planning and evaluation of conservation measures: population density; survival; recruitment; and temporary and permanent emigration (Karanth *et al.* 2006). The proportion of surviving wild tiger populations represented in photographic identification (photo-id) catalogues is therefore increasing and is set to increase further as imaging technology improves.

Where populations are small or highly fragmented, new images can be compared by eye with existing ones to identify new individuals and record where and when each was photographed. But as the photo-id catalogue grows with population turnover and immigration from adjacent sites, the effort required to search the entire catalogue for matches to each new image becomes excessive. The number of visual comparisons required can be reduced substantially by calculating similarity scores between each new image and the images by which existing individuals are represented in the catalogue. Visual comparisons can then be restricted to those individuals for which the maximum score exceeds a threshold value, as exemplified by the work of Kelly (2001) for cheetahs and Karlsson *et al.* (2005) for grey seals.

Camera trap images of tigers are characterized by a wide range of camera angles and body posture. To allow similarity scores to be calculated between such images, we developed a three-dimensional surface model of a tiger. The software, first developed for matching images of grey seals (Hiby & Lovell 1990), fits the model to each new image and scans the stripe pattern from the region of the image lying beneath a predefined region of the model, so that the resulting pattern sample is largely unaffected by camera angle and body posture. A pattern comparison algorithm then calculates similarity scores between the new and existing pattern samples. Scanning the pattern via a three-dimensional model, in effect, unwraps it from the curved surface of the tiger's flank and thus also allows a camera trap image of a living tiger to match an image of its skin laid out flat.

Using camera trap images from the Nagarhole and Bandipur tiger reserves in Karnataka, India, we demonstrate the resulting reduction in effort required to generate capture histories from a large database of camera trap images. We also suggest that the software could be used as a forensic tool to trace the origin of tiger skins.

2. MATERIAL AND METHODS

The surface model of the tiger, stored as a text file of three-dimensional coordinates, extends from the shoulders to the base of the tail, undercutting the fore and hind limbs. Figure 1 shows the model fit to a camera trap image and the pattern sample scanned from the flank region. The surface model vertices are the endpoints of rays emanating at intervals from an internal spine, the ray lengths at each interval forming a cosine series function of angle from the dorsal midline, so that initially the model is bilaterally symmetrical. The series coefficients are fixed and were determined by least squares fit to a cloud of surface points generated by photogrammetry, using photographs of a zoo tiger. The computer program initially scales, translates and rotates the model by minimizing the sum of squared distances from shoulder, hip and tail base points on the model to corresponding points on the image, positioned manually on the screen by the user. The model can also twist and bend and thicken or thin differentially along its long axis. A set of four bi-quadratic splines determine the bending and twisting angles and the expansion factors along the model spine. The spline node coordinates introduce a further set of parameters that are then used in conjunction with the scale, translation and rotation parameters to optimize the fit of the projected model to image margins marked by the user.

The same procedure is used to scan left and right flank pattern samples from a photograph of a skin laid flat, except that the surface model used is planar and is not permitted to bend or twist. Thus, as for the live animal photographs, the pattern samples are largely unaffected by camera angle or the scale and orientation of the image.

The pattern sample scanned via the model from the new image is then compared with the database of previous samples. Two algorithms are used to calculate a similarity score between each pair of pattern samples. We omit the details here because a variety of algorithms are available for comparing patterns, but note that, by combining algorithms that are to some extent complementary, it is

Received 14 January 2009
Accepted 9 February 2009

This journal is © 2009 The Royal Society

Downloaded from rsbl.royalsocietypublishing.org on 8 June 2009



Figure 1. The three-dimensional model fitted to a camera trap image of a tiger. The yellow dots placed on the screen by the user indicate the position of shoulder, hip and tail base points and the red and blue dots indicate the upper and lower margins of the image, respectively (note the second tiger behind the first, which might make automated segmentation of the first image from the background difficult).

possible to reduce the risk of obtaining a low score between matching patterns (i.e. those scanned from images of the same tiger). One of the algorithms is robust to noise in the image but fails when regions of the image are occluded, and the other is robust to occluded regions but sensitive to noise.

The final similarity score is calculated as the posterior probability that the patterns are from images of the same tiger, given the values returned by the two algorithms. The prior probability that two randomly chosen images are of the same tiger could be adjusted to reflect the recorded sex, age and location of the images, but in this case is set simply to the reciprocal of an assumed, local population size, N. If S_c and S_n are the values returned by the two algorithms, the posterior probability that the patterns are from images of the same tiger is given by

$$P(\text{match}|S_c,S_n)=\frac{P(S_c,S_n|\text{match})\times P(\text{match})}{P(S_c,S_n)}$$
$$=\frac{P(S_c,S_n|\text{match})\times P(\text{match})}{P(S_c,S_n|\text{match})\times P(\text{match})+P(S_c,S_n|\text{non-match})\times P(\text{non-match})}$$
$$=\frac{1}{1+P(S_c,S_n|\text{non-match})\times(N-1)/P(S_c,S_n|\text{match})},$$

where $P(S_c,S_n|\text{non-match}|)$ and $P(S_c,S_n|\text{match})$ are the joint densities of S_c and S_n for non-matching and matching pattern pairs, and are estimated from the frequency distributions of the scores that each member of the pair accumulated as it was compared with the existing catalogue. The value of N affects the posterior probability but not the rank order of the existing individuals. In addition to combining the values into a single similarity score, the posterior probability reduces the expected rank of individuals that tend to achieve higher than average algorithm values (those represented only by low-quality images).

The posterior match probability can be compared with a threshold value, so that only those catalogue individuals from which at least one pattern sample exceeds the threshold are checked visually against the new images. However, in the tiger system, all catalogue animals are ranked in order of decreasing maximum posterior probability, so that the user has the option to search the entire catalogue and any existing match will be found as soon as possible.

3. RESULTS

Camera trap images from the Nagarhole and Bandipur tiger reserves in India were used to test the performance of the system. From 1990 to 2007, in ongoing studies of tiger ecology (Karanth *et al.* 2004, 2006), 227 tigers were identified by patterning on both sides, 37 on the left only and 34 on the right only, thus from 264 to 298 individuals, based purely on careful visual comparison of the images. Pattern samples were scanned via the three-dimensional model from multiple images of each tiger. The similarity score was calculated for every possible pairing of the samples. For each matching pair, that score was ranked against the scores that each member of the pair obtained with non-matching individuals (selecting the highest scoring image of the non-matching individuals). Ninety-five per cent of the matching pairs achieved rank 1 and all were in the top five rank positions. Pairs achieving rank 1 differed in camera angle by up to 66° and by up to 7 years in the date the photographs were taken. The test results are similar to those returned by the system when used with cheetah photographs and are better than those for grey seals. They indicate its ability to locate a match to a new image if the matching individual is currently represented by a single image in the catalogue. The risk of the matching individual failing to achieve a high rank is greatly reduced if multiple images are retained in the database. As yet, the ability of the system to match photographs of cubs has not been tested.

Images of three skins confiscated from poachers were also obtained from 2006 to 2008 (figure 2). Each provided pattern samples from the left and right flanks, which were compared with the catalogue. Five of the six samples generated similarity scores with catalogue images of the living tiger at rank position 1. The sixth sample was taken from the right flank of the skin image in figure 2*c*, in which the skin is partially folded and failed to achieve a high rank position.

4. DISCUSSION

In the current system, the pattern sample is scanned from each new image via a user interface. The time

Downloaded from rsbl.royalsocietypublishing.org on 8 June 2009

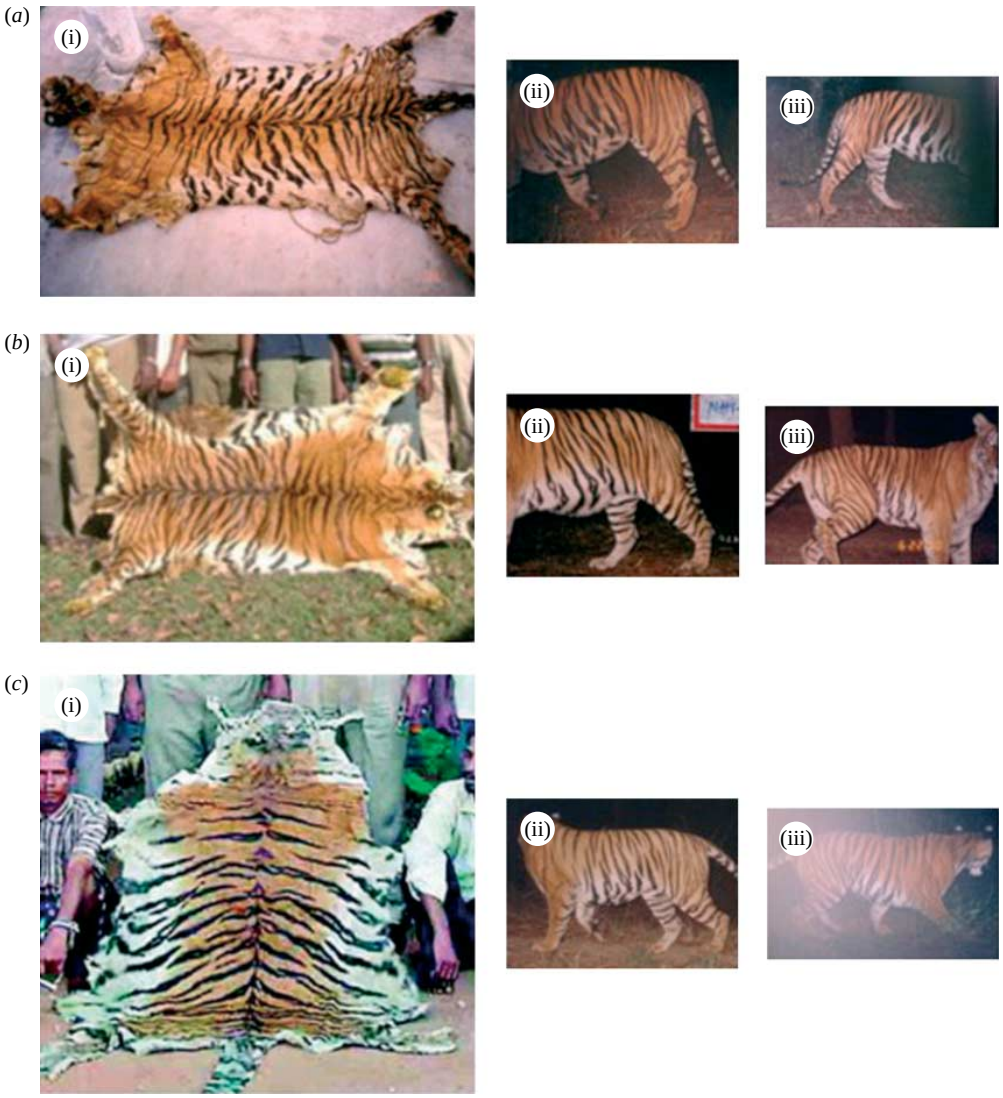


Figure 2. (*a–c*) Images of (i) tiger skins confiscated locally and (ii) left and (iii) right camera trap images of the living tigers taken in the reserves from which the animals were poached. Five of the six pattern samples taken from the skin images generated the highest similarity score with an image of the correct animal from the catalogue. The pattern on the right side of the tiger (*c*(i)) skin image was not scored as similar to the right side of the camera trap image of that tiger (iii).

required to complete that step is not significant in comparison with the effort required to obtain the image and is far less than the 30–45 min currently required for a purely visual search of the photo-id catalogue. However, the user interface introduces an element of subjectivity into the model fitting, and hence a risk of reducing the similarity scores between matching pattern samples if different users are involved. Recent advances in automated detection of animals in video images (Burghardt 2008) suggest that the model fitting step could be further automated.

The algorithms used to compare the pattern samples are also unlikely to be optimal. Nevertheless, the results from the Nagarhole and Bandipur images suggest that the current system is almost certain to place a matching individual from the catalogue at or near the top of the list of similarity scores.

This study also suggests that if copies of camera trap images were accumulated in a central database, an image of a skin that had been taken from one of the tigers represented in that database could be traced within a few minutes to where and when the living animal was last recorded. Such images could be sent from anywhere in the world and would therefore not be restricted to skins located within India. Furthermore, the database need not be restricted to camera trap images and could include, for example, images taken by tourists.

The system has been developed for other uniquely marked species such as leopards, zebras and salamanders. We believe that it provides a tool that is widely applicable for both scientific study and conservation of rare or threatened species.

Development of the software, which is available as a free download from www.conservationresearch.co.uk, was partially funded by the Tigers Forever Project of the Panthera Foundation and the Wildlife Conservation Society, New York. Facilities for obtaining the images used

Downloaded from rsbl.royalsocietypublishing.org on 8 June 2009

for the construction of the three-dimensional surface model were provided by the Thrigby Hall Zoo, Norfolk, England. The Centre for Wildlife Studies, Bangalore and the Wildlife Conservation Society, India Programme provided images, local resources and staff time to support this study.

Burghardt, T. 2008 Visual animal biometrics. PhD thesis, University of Bristol, UK.

Hiby, L. & Lovell, P. 1990 Computer-aided matching of natural markings: a prototype system for grey seals. *Rep. Int. Whaling Comm*. (Special issue) **12**, 57–61.

Karanth, K. U., Nichols, J. D., Kumar, N. S., Link, W. A. & Hines, J. E. 2004 Tigers and their prey: predicting carnivore densities from prey abundance. *Proc. Natl Acad. Sci. USA* **101**, 4854–4858. (doi:10.1073/pnas.0306210101)

Karanth, K. U., Nichols, J. D., Kumar, N. S. & Hines, J. E. 2006 Assessing tiger population dynamics using photographic capture–recapture sampling. *Ecology* **87**, 2925–2937. (doi:10.1890/0012-9658(2006)87[2925:ATPDUP]2.0.CO;2)

Karlsson, O., Hiby, L., Lundberg, T., Jussi, M., Jussi, I. & Helander, B. 2005 Photo-identification, site fidelity, and movement of female grey seals (*Halichoerus grypus*) between haul-outs in the Baltic Sea. *Ambio* **34**, 628–634. (doi:10.1639/0044-7447(2005)034[0628:PSFAMO]2.0.CO;2)

Kelly, M. J. 2001 Computer-aided photograph matching in studies using individual identification: an example from Serengeti cheetahs. *J. Mammal.* **82**, 440–449. (doi:10.1644/1545-1542(2001)082<0440:CAPMIS>2.0.CO;2)

Williams, B. K., Nichols, J. D. & Conroy, M. J. 2002 *Analysis and management of animal populations*. New York, NY: Academic Press.

Appendix B

Evaluation of non-invasive genetic sampling methods for estimating tiger population size

Biological Conservation 142 (2009) 2350–2360

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Evaluation of non-invasive genetic sampling methods for estimating tiger population size

Samrat Mondol [a], K. Ullas Karanth [b,c], N. Samba Kumar [b,c], Arjun M. Gopalaswamy [b,c], Anish Andheria [d], Uma Ramakrishnan [a,*]

[a] National Centre for Biological Sciences, TIFR, GKVK Campus, Bellary Road, Bangalore 560 065, India
[b] Wildlife Conservation Society, India Program, 1669, 31st Cross, 16th Main, Banashankari 2nd Stage, Bangalore 560 070, India
[c] Centre for Wildlife Studies, 26-2, Aga Abbas Ali Road (Apt: 403), Bangalore 560 042, India
[d] Post-graduate Program in Wildlife Biology and Conservation, National Centre for Biological Sciences, TIFR, GKVK Campus, Bellary Road, Bangalore 560 065, India

ARTICLE INFO

Article history:
Received 11 February 2009
Received in revised form 12 May 2009
Accepted 17 May 2009
Available online 17 June 2009

Keywords:
Fecal DNA
Genetic capture–recapture sampling
Panthera tigris
Population estimation
Individual identification
Elusive carnivores
Scats
Carnivore conservation

ABSTRACT

There is often a conservation need to estimate population abundances of elusive, low-density, wide-ranging carnivore species. Because of logistical constraints, investigators often employ non-invasive 'captures' that may involve 'genetic' or 'photographic' sampling in such cases. Established capture–recapture (CR) methods offer a powerful analytical tool for such data. In this paper, we developed a rigorous combination of captive, laboratory and field-based protocols for identifying individual tigers (*Panthera tigris*) from fecal DNA. We explored trade-offs between numbers of microsatellite loci used for reliable individual identifications and the need for higher capture rates for robust analyses. Our field surveys of scats were also specifically designed for CR analyses, enabling us to test for population closure, estimate capture probabilities and tiger abundance. Consequently, we could compare genetic capture estimates to results of a 'photographic capture' study of tigers at the same site. The estimates using the heterogeneity model (M_h-Jackknife) for fecal DNA survey were [M_{t+1} = 26; $\hat{p} = 0.09$ and $\hat{N}$ ($S\hat{E}[\hat{N}]$)=66 (12.98)] in close agreement with those from the photographic survey [(M_{t+1} = 29; $\hat{p} = 0.04$ and $\hat{N}$ ($S\hat{E}[\hat{N}]$) = 66 (13.8)]. Our results revealed that designing field surveys of scats explicitly for CR data analyses generate reliable estimates of capture probability and abundance for elusive, low density species such as tigers. The study also highlights the importance of rigorous field survey and laboratory protocols for reliable abundance estimation in contexts where other approaches such as camera-trapping or physical tagging of animals may not be practical options.

Crown Copyright © 2009 Published by Elsevier Ltd. All rights reserved.

1. Introduction

Abundance (population size) is a state variable of primary interest to scientists and conservationists because of its decisive influence on ecological and behavioral attributes, and, thus the potential viability of any animal population (Williams et al., 2002). However, abundance is a particularly difficult parameter to estimate for low density, wide-ranging, elusive, carnivore species, most of which are threatened (Schipper et al., 2008). For example, wild tigers (*Panthera tigris*) have suffered a massive range contraction of ~93% in historical times, with a global population size <3000–4000 animals surviving in fragmented local populations (Sanderson et al., 2006; Morell, 2007; Ranganathan et al., 2008). However, it is not practical to employ invasive physical tagging methods to derive tiger density estimates because of logistical difficulties, high costs and small numbers of captures possible (Karanth and Nichols, 2002). Consequently, non-invasive, photographic capture–recapture methods were initially developed by Karanth (1995) and thereafter refined (Karanth and Nichols, 1998, 2002) for estimating tiger numbers. Since then photographic CR sampling has been successfully employed on tigers (Karanth et al., 2004b; Wegge et al., 2004; Kawanishi and Sunquist, 2004; Simchareon et al., 2007) and other elusive felids like leopards (Henschel and Ray, 2003), jaguars (Silver et al., 2004) and ocelots (Trolle and Kery, 2003). The photographic CR method takes advantages of the fact that in all these species, individuals are identifiable from natural markings. However, they do not work for species in which individuals are not uniquely identifiable. Other potential disadvantages of photographic capture methods include: requirement of large number of camera traps, scarcity of skilled personnel, vulnerability of cameras to theft, vandalism, animal damage or adverse weather, and, potentially low animal densities or wariness resulting in small number of detections.

In this context, non-invasive genetic typing of DNA extracted from animal hair or scat has emerged as an alternative option for

* Corresponding author. Tel.: +91 80 23666030; fax: +91 80 65626304.
E-mail address: uramakri@ncbs.res.in (U. Ramakrishnan).

0006-3207/$ - see front matter Crown Copyright © 2009 Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.biocon.2009.05.014

capture–recapture sampling of populations of such rare, endangered or cryptic species (Waits, 2004). Genetic identification methods have often been simply used to count the minimum number of individuals or as part of pilot studies (e.g. Sloane et al., 2000; Creel et al., 2003; Fickel and Hohmann, 2006; Bhagavatula and Singh, 2006). However, since not all individuals in the population may be detected, such estimates are of limited utility (Williams et al., 2002).

Several different approaches have been used to estimate animal abundance from genetic data, in the face of such imperfect detections. Principally, these include rarefaction curve from unique genotypes (Kohn et al., 1999; Eggert et al., 2003; Wilson et al., 2003; Frantz et al., 2003, 2004), genetic capture–recapture method (Banks et al., 2002; Mowat and Paetkau, 2002; Boersen et al., 2003; Solberg et al., 2006) using Jackknife estimators (Flagstad et al., 2004), Lincoln–Peterson estimator (Triant et al., 2004) under closed (Rudnick et al., 2008) or open population (Prugh et al., 2005) capture–recapture (CR) models or under 'capture with replacement' methods (Miller et al., 2005; Zhan et al., 2006; Puechmaille and Petit, 2007). Different estimation methods have been reviewed from both theoretical (McKelvey and Schwartz, 2004; Lukacs and Burnham, 2005) and empirical (Petit and Valiere, 2006; Piggott et al., 2006; Dreher et al., 2007) perspectives. Among these, closed and open model capture–recapture (CR) methods have been the most well-founded theoretically and empirically (Williams et al., 2002; Amstrup et al., 2005).

Estimation methods based on 'genetic captures' require data from multiple hypervariable loci for unambiguous identification of individual animals (Taberlet and Luikart, 1999; Mills et al., 2000; Lukacs and Burnham, 2005; Waits and Paetkau, 2005; Broquet et al., 2007). However, getting reliable genotypes at sufficient loci for reliable individual identification poses a challenge. For example, amplifying all ten loci from a degraded genetic sample is much harder than amplifying just five loci. This requirement of increased genetic information lowers the number of identifiable individuals counted, thereby reducing numbers of 'individual captures' available for CR analysis, with consequent trade-offs between degree of accuracy of individual identification and statistical needs of robust CR modeling.

In this study, we addressed key methodological issues related to field survey design, laboratory protocols and CR analyses relevant to genetic capture–recapture sampling of tigers based on DNA extracted from scats. Our study was carried out in the 880 km^2 Bandipur Tiger Reserve (76°12′–76°46′ E and 11°37′–11°57′ N) in Karnataka state, India. This rugged landscape is drained by several rivers and varies in altitude (680–1454 m), annual rainfall (625–1250 mm) and mean monthly temperature (18 °C–29 °C). Its deciduous forests support high densities of ungulate prey species (~35.2 animals/km^2), and consequently, a large tiger population (Karanth et al., 2004b). Additionally, long-term photographic CR monitoring in this site provides independent, robust estimates of tiger abundance, allowing us to specifically investigate the possible trade-offs between the individual identification accuracy and CR.

Our objectives were: (1) develop and validate a genetic identification protocol for tigers using known distinct individuals (2) employ field survey protocols explicitly designed for robust capture–recapture analyses under various plausible models (Williams et al., 2002; Karanth et al., 2004a, 2006) (3) evaluate fecal DNA-based tiger abundance estimates in comparison to standard photographic capture–recapture methods, and, (4) explore methodological trade-offs between the twin objectives of attaining greater certainty in individual identifications by using more loci versus achieving higher capture and recapture rates for subsequent analyses. In addition to estimation of tiger numbers, we believe our results have wider relevance for non-invasive, fecal DNA based population assessments in many other scarce, low density, elusive, wide-ranging animal species.

2. Materials and methods

2.1. Field surveys for collection of scats

The study area of 671 km^2, located within Bandipur contained 18 'search routes' spatially distributed to ensure that the scats of each individual tiger within the area had probability of being detected. This was ensured by embedding the surveyed area within a camera trap array area that used in earlier photographic capture–recapture studies (Karanth et al., 2004a,b) designed to avoid large 'holes' in which an individual tiger had no exposure to detection (Fig. 1). These surveys followed dirt roads and trails that are known to be used as regular travel routes and regularly marked by scat deposits by tigers (Smith et al., 1989; Karanth and Sunquist, 2000). Each route was sampled by a team of two trained trackers at seven day intervals. Thus the entire study area was surveyed over six successive days (followed by a day of rest) by three teams, which covered a distance of 235.4 km, collecting 'fresh' tiger scats judged to be less than 7-days-old based on physical appearance and moisture content (Andheria, 2006). The survey was repeated over six consecutive weeks in the same order to ensure that each 'search route' was visited regularly once a week. The survey yielded a total of 63 samples.

In terms of capture-history data structure (Otis et al., 1978; Williams et al., 2002), tiger scats collected during days 1–6 of the survey were assigned to the first 'sampling occasion', days 8–13 to the second sampling occasion, and so on, to attain a total of six sampling occasions over 42 days. Thus our survey met two key requirements of standard capture–recapture analyses: (1) full spatial coverage of the area in each sampling occasion ensuring exposure of all individual tigers present to potential 'captures', and, (2) a sufficiently short survey period that reasonably met the assumption of population closure (Karanth and Nichols, 1998).

2.2. Validation and development of laboratory protocols

Blood samples were collected from 14 captive tigers from three different zoos in southern India. These animals included close relatives such as siblings, and were used for standardization of the final set of loci used for individual identification. Additionally, scats were collected from 13 separately housed tigers to test the effects of different sources of DNA (i.e. blood, or fresh scats <12 h old) on the amplification process and associated error estimates.

To test the efficiency of the microsatellite loci from scats of wild tigers, we collected 10 fresh scats deposited the same night, but sufficiently far apart within the Bandipur–Nagarahole landscape, allowing us to assume reasonably that they came from different individuals within the same genetic population. These scats were used to standardize protocols but not in the abundance estimation. These field scat samples were carefully collected to avoid contamination and preserved in absolute alcohol for DNA extraction. All samples were stored in ethanol as opposed to other storage mediums (e.g silica gel) following recommendations in Murphy et al., 2002.

2.3. DNA extraction and species identification

DNA was extracted in duplicate from all scat samples using commercially available QIAamp DNA Stool mini kit (QIAGEN Inc.) following the manufacturer's instructions with slight modifications. Around 180–200 mg of sample from the outer parts of the faeces was used for each extraction. The first overnight incubation

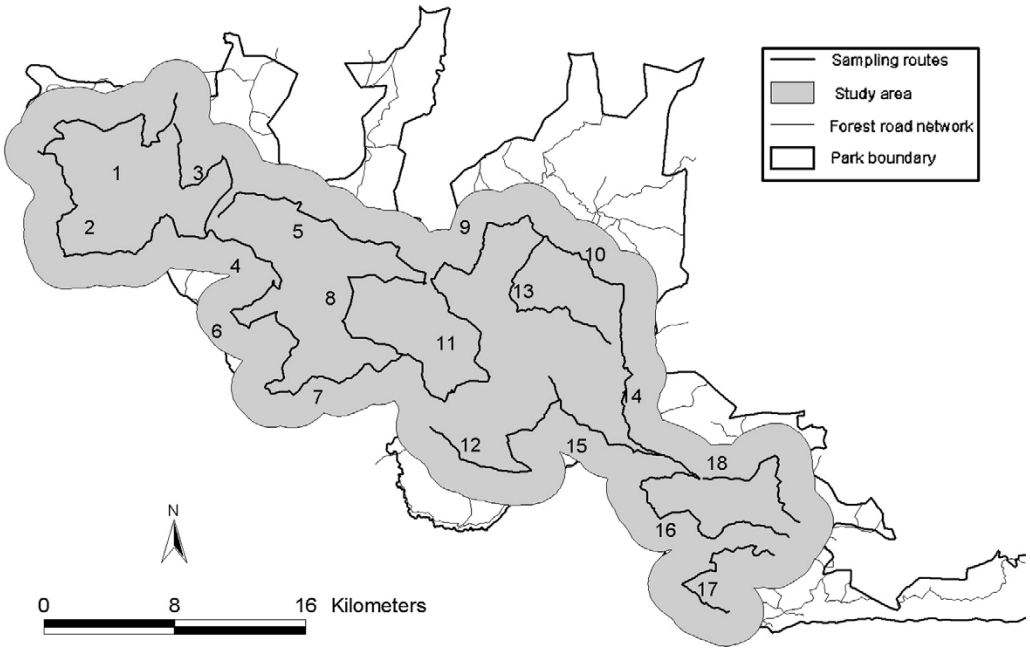


Fig. 1. Study area showing the 18 sampling routes in Bandipur National Park used for collecting tiger scat samples.

was conducted at room temperature with 1 ml ASL buffer, followed by a second incubation at 65 °C for 60 min. Four micrograms of carrier RNA (Poly-A from NEB) was added to increase potential DNA yield from faecal samples as recommended by Kishore et al., 2006. DNA Elution was done with 120 μl of TE buffer (pH 7.8). All extractions (in sets of ten samples) included an extraction control to monitor contamination. DNA was extracted in a separate, pre-PCR laboratory space (no prior extractions from mammalian samples had been conducted in this space). DNA was extracted from blood using QIAamp DNA Tissue Kit (QIAGEN Inc.) following manufacturer's instructions.

All scat samples collected in the wild (both during the Bandipur survey and field test) were first screened for species identification using species-specific primers developed earlier (Mukherjee et al., 2007). Despite the ability of trackers to distinguish tiger scats from those of leopards based on size and other characteristics (Karanth and Sunquist, 2000), as a further precaution we used DNA-based comparisons (Davison et al., 2002), As a result, we ensured that only genuine tiger scats received further analyses.

2.4. Microsatellite primer selection

We selected 33 loci characterized for either domestic cat (Menotti-Raymond et al., 1999) or different tiger subspecies including *P. t. sumatrae* (Williamson et al., 2002), *P. t. amoyensis* (Zhang et al., 2005) and *P. t. tigris* (Bhagavatula and Singh, 2006). The loci were selected based on their polymorphic information content and expected heterozygosity.

2.5. PCR standardization and genotyping

All PCR standardizations were initially conducted using the blood DNA samples. For all standardizations amplification was carried out in 10 μl reaction volumes containing 5 μl Qiagen multiplex PCR buffer mix (QIAGEN Inc.), 0.2 μM labeled forward primer (Applied Biosystems), 0.2 μM unlabelled reverse primer, 4 μM BSA and 1 μl of the DNA extract. For scat DNA, 3 μl of the extract was used in the reaction mixture. The temperature regime included an initial denaturation (94 °C for 15 min); 45 cycles of denaturation (94 °C for 30 s), annealing (T_a for 45 s) and extension (72 °C for 45 s); followed by a final extension (72 °C for 30 min) in an Eppendorf thermocycler. Post-temperature standardization, primers with same annealing temperatures were optimized for multiplex reactions from the blood and faecal DNA samples. For multiplex reactions, the PCR conditions were identical to uniplex ones. We also optimized the number of PCR cycles, which were 30 for blood and 45 for faecal samples. PCR negatives were incorporated in all reaction setups to monitor contamination. The PCR products were visualized in 2% agarose gels. One microlitre of the amplified product was added into 12 μl of formamide (Applied Biosystems) and 0.5 μl of ROX 500 size standard (Applied Biosystems) and then run into an automated sequencer ABI3100XL (Applied Biosystems). Microsatellite alleles were scored with GENEMAPPER version 4.0 (Applied Biosystems).

To obtain reliable genotypes we repeated the genotyping procedure thrice from all blood DNA samples. For all the faecal samples, the genotyping procedure was performed four times. Average amplification success (over individuals) for a given locus was calculated as the percent positive PCR (Broquet et al., 2007). In other words, amplification success was calculated as the number of times a particular locus showed amplification for all samples in all the trials. This definition of amplification success allows us to distinguish it from sample quality. Microsatellite loci that revealed positive amplification were tested for 13 scat samples collected from 13 captive individuals. The results from blood and scat genotyping were compared to evaluate the suitability of these particular microsatellites. Further, we quantified the average allelic dropout (across individuals) and the average number of false alleles (or genotyping success) as the number of dropouts or the number of false alleles over the total number of amplifications respectively from both kinds of samples. We compared the genotypes obtained from the blood and scat samples and calculated the $P_{ID(obs)}$ and the $P_{ID(sibs)}$ based on the captive sample genotypes (blood and scat genotypes were the same) using GIMLET (Valière,

Table 1
Genetic variability at 30 microsatellites for 14 captive individual tigers (blood samples).

Locus	Allele size range(bp)	Blood amplification success (for faeces)	T_a	Allelic dropout	False alleles	H_e	H_o	No. of alleles	Reference
TIGER6	221–243	100 (40)	59	0	0	0.82	0.79	9	1
6HDZ170	214–226	100 (52)	55	0	0	0.78	0.5	7	2
FCA391[a]	197–217	100 (100)	56	0	0	0.77	0.71	6	3
FCA672[a]	94–112	100 (100)	56	0	0	0.73	0.93	5	3
F42	206–230	100 (32)	56	0	0	0.73	0.5	6	3
FCA205[a]	102–116	100 (100)	56	0	0	0.69	0.93	5	3
F115[a]	188–204	100 (100)	56	0	0	0.69	0.79	4	3
6HDZ007	227–241	100 (65)	60	0	0	0.68	0.36	4	2
FCA441[a]	146–174	100 (100)	51	0	0	0.64	0.64	5	3
TIGER22[a]	124–130	100 (100)	60	0	0	0.64	0.64	4	1
FCA126[a]	126–152	100 (100)	56	0	0	0.61	0.71	4	3
FCA232[a]	100–108	100 (100)	56	0	0	0.61	0.93	4	3
F41[a]	194–202	100 (100)	56	0	0	0.58	0.43	3	3
FCA309[a]	94–98	100 (100)	56	0	0	0.59	0.79	3	3
FCA506	206–214	100 (42)	52	0	0	0.57	0.5	3	3
FCA230[a]	101–107	100 (100)	52	0	0	0.57	0.64	3	3
TIGER23	262–314	100 (17)	55	0	0	0.55	0.21	6	1
FCA628[a]	090–096	100 (100)	56	0	0	0.52	0.57	4	3
FCA453[a]	156–176	100 (100)	59	0	0	0.49	0.29	4	3
FCA164[a]	74–84	100 (100)	59	0	0	0.5	1	2	3
FCA279[a]	98–100	100 (100)	52	0	0	0.5	1	2	3
FCA001[a]	136–140	100 (100)	54	0	0	0.5	0.5	2	3
FCA304[a]	120–138	100 (100)	56	0	0	0.46	0.14	4	3
FAC052[a]	108–114	100 (100)	55	0	0	0.46	0.71	2	3
E21B	159–167	100 (78)	55	0	0	0.35	0.36	3	4
6HDZ056	174–176	100 (67)	56	0	0	0.34	0	2	2
TIGER24	176–178	100 (90)	51	0	0	0.24	0	2	1
FCA069[a]	96–106	100 (100)	56	0	0	0.2	0.21	3	3
FCA090	106–116	100 (92)	51	0	0	0.14	0.14	3	3
FCA170	76	100 (77)	55	0	0	0	0	1	3

1 – Zhang et al. (2005).
2 – Williamson et al. (2002).
3 – Menotti-Raymond et al. (1999).
4 – Bhagavatula and Singh (2006).
[a] Loci amplified from faecal samples of captive individuals.

2002). Ten loci were chosen finally for further analysis based on their amplification success in the faecal DNA samples (see results). These 10 microsatellite loci were combined into six PCR reactions (information provided in Table 2).

2.6. Data validation

In this study, we maximized data quality using a modified multiple-tube approach monitored by the quality index approval. All the faecal samples from the study area that showed positive amplification for at least three loci after first genotyping were selected and the whole process was repeated thrice, independently. After scoring the alleles with GENEMAPPER version 4.0 (Applied Biosystems) the genotypes from four different amplifications were compared and the quality index value was assigned as per Miquel et al. (2006). In other words, samples that revealed the identical genotype in all four repetitions had a quality index of one, compared to samples that yielded three identical genotypes out of the four amplifications (quality index = 0.75).

2.7. Individual identification

We used our set of 10 microsatellites to create genotype profiles for all the samples. We then compared these sample profiles using the identity analysis module in program CERVUS (Marshall et al., 1998), which identifies samples with identical genotypes for the specified number of loci. We explain the approach that we have taken to individual identification in the context of a five loci criteria, although the approach was similar for the three and seven loci datasets as well. Identical genotypic profiles (based on five or more loci) were used to identify multiple instances of the same individual. Because we included incomplete profiles (using an approach similar to that described by Frantz et al., 2003), we scored individuals as the same if they had identical genotypes for at least five or

Table 2
Genetic variability at 10 microsatellites for ten field test samples (Scat samples) of wild tigers from Bandipur National Park and Nagarahole National Park.

Locus	Product size range (bp)	Amplification success (%)	Allelic dropout (%)	False alleles	H_e	H_o	No. of alleles	Multiplex sets for PCR
FCA453	192–200	71	0	0	0.62	0.14	3	Singleplex
FCA391	205–217	71	0.035	0	0.54	0.29	3	
FCA628	86–108	100	0	0	0.64	0.71	3	Set 1
FCA205	100–108	100	0	0	0.78	1	5	
FCA126	140–146	86	0	0	0.64	0.86	4	Set 2
F41	112–124	100	0.035	0	0.36	0.29	3	
FCA232	98–106	100	0	0	0.44	0.43	3	Set 3
FCA441	146–158	71	0	0	0.18	0.14	2	Singleplex
FCA672	100–102	100	0	0	0.46	0.71	2	
F115	192–196	100	0	0	0.49	0.29	2	Set 4

more common loci. Such analyses then allow us to discern both the number of unique individuals as well as the number of recaptured individuals. We re-iterate that although there is a possibility that the incomplete genotypes might belong to a new individual, this method of grouping them with matching samples ensure conservative population estimation (Frantz et al., 2003) by minimizing individuals created through erroneous, multi-locus genotypes (non-existent individuals).

We similarly identified unique and recaptured individuals based on seven (genotypes identical for at least seven loci) and three loci (genotypes identical for at least three loci) for population estimation.

2.8. Analysis of 'tiger capture' data

The collection of scats following an explicit survey design (Karanth and Nichols, 2002; Andheria, 2006) enabled us to use standard closed model capture–recapture analyses (see reviews by Otis et al., 1978; Williams et al., 2002; Amstrup et al., 2005).

Systematic and rigorous validation and development of genetics-based identifications permitted us to achieve a high degree of certainty in individual identification of tigers even from among wild tigers that could be potential siblings, allowing us to avoid capture–recapture models with additional nuisance parameters that account for such uncertainties (see Waits et al., 2001; Lukacs and Burnham, 2005 for details).

Each genetic capture of an individual tiger was assigned to the appropriate sampling occasion (as described earlier) to generate individual capture histories in the standard X matrix format (Otis et al., 1978). These capture histories were analyzed using the software program CAPTURE (Rexstad and Burnham, 1993; White et al., 1982). Earlier work (Karanth and Nichols, 2002; Karanth et al., 2004b) shows that closed capture–recapture model M_h, which accounts for heterogeneity of capture probabilities among individuals, is often most useful for tiger populations. Because of its stability and better convergence properties during iterative computations, and, based on reasonable results in the past with tiger data (Karanth and Nichols, 2002) we preferred the M_h Jackknife estimator (Burnham and Overton, 1978) implemented in program CAPTURE over the M_h models implemented in program MARK (White and Burnham, 1999) wherein heterogeneity is handled using the method of finite mixtures (Pledger, 2000). Additionally, we also report results of analyses using the simpler M_0 (null) model to enable comparisons, although this model is not robust to violation of assumptions of invariant capture probabilities.

Finally, we also compared these faecal DNA-based tiger abundance estimates to abundances derived from a 48-day photographic capture–recapture survey of tigers conducted 15 weeks earlier in the same area. The photographic CR survey involved 118 trap locations set along some of the same scat sampling routes (See Fig. 2). It expended 1246 camera trap nights of effort across 12 sampling occasions. The survey resulted in 35 photographic 'captures' of 29 distinct individuals (Karanth et al., 2008, unpublished report), from the same tiger population that we subsequently sampled using genetic captures. The methods used in the photographic capture–recapture survey are described in detail elsewhere (Karanth and Nichols, 1998, 2002; Karanth et al., 2004a, 2004b).

3. Results

3.1. Species identification from the field samples

Genetic screening showed that from among 73 scat samples collected in the wild (63 from Bandipur survey and 10 field test samples), 68 scats were of tigers (all 10 field test samples and 58 survey samples), with the remaining 5 scats being determined to be from leopards.

3.2. Amplifying microsatellites from DNA samples of Indian tigers

While reporting our results, we have, respectively, used the following terms; "captive samples" (for blood and scat samples from zoo tigers), "field test samples" (for scat samples obtained from putatively different individuals from regional tiger population)

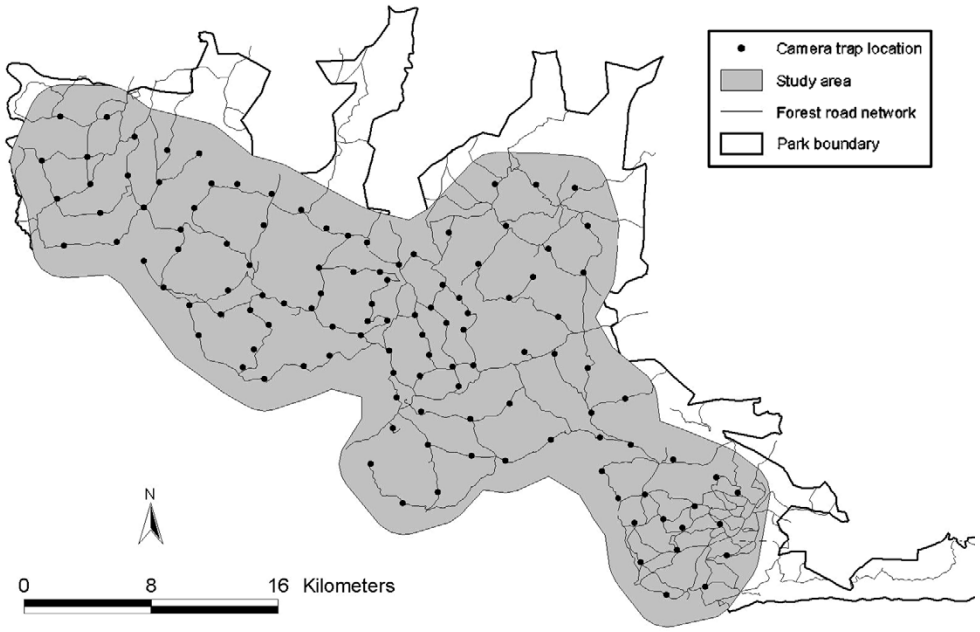


Fig. 2. Study area showing the Camera trap locations in Bandipur National Park for the photographic capture–recapture survey.

and "Bandipur samples" for scat samples obtained from our field survey.

A total off 33 microsatellite primers were tested for the captive tiger blood DNA samples, of which 30 primers showed amplification (Table 1). The microsatellite loci varied from extremely polymorphic (locus Tiger6, nine alleles; loci FCA164 and FCA279, $H_o = 1$) to being monomorphic (locus FCA170, one allele; loci 6HDZ056, Tiger24 and FCA170, $H_o = 0$). Results for captive tiger scat DNA samples revealed that amplification success was higher for loci with shorter amplicon size. Out of these 30 loci, 19 loci showed 100% amplification success from these scat samples (Table 1). The remaining loci showed amplification success range from 17 to 92% for the 13 scat samples. Further observation reveals that loci with larger amplicons (generally <200 bp) produce low amplification success (in this case, highest of 65%). Since none of the amplifiable loci showed allelic dropout, we infer that for captive scat DNA samples (in which DNA quality is presumed high, because of their freshness), amplification success depends only on amplicon size of the locus.

For the 10 field test samples from wild tigers in the region, all 19 amplifiable loci were tested. For individual identification we require reliable data from maximum number of loci for further analyses towards population estimation. Out of the 19 loci tested, all produced reliable data but only 10 of the 19 tested loci showed an amplification success higher than 70%. These ten loci thus were used in this study and results from these 10 loci are presented in Table 2.

We calculated the probability of identity curves based on these 10 loci for both captive and field test samples (Fig. 3). This graph highlights the $P_{ID(sibs)}$ for three, five and seven most informative loci, and reveals that the $P_{ID(sibs)}$ was higher for the captive samples compared to the field test samples, and that it increases to 0.008 and 0.0003 for field test and captive individuals respectively for seven loci.

Fresh Bandipur scats (50) were amplified for the above 10 loci. Amplification success for these samples was high at >74% and allelic dropout was less than 1.7% (see Table 3). For all the ten loci presented in Table 3, the number of alleles observed for the tiger samples from Bandipur survey was higher than that observed for the captive or field test tigers. Genotyping success for the Bandipur tigers ranged from 98 to 100%.

3.3. Individual identification and data validation

Of the 58 tiger scats collected in the survey, only 50 produced genetic data. Only 38 of these 50 samples produced five or more loci genotypes with quality index >0.75. Twelve of these samples produced data for all ten loci. We could generate data from nine, eight, seven and five loci genotypes profile from 11, six, five and four samples, respectively. Before conducting individual identification, we investigated how the probability of identity changes as a function of number of loci for the Bandipur samples. In order to do so, we used the conservative cutoff of seven loci suggested by Fig. 3 (both field test and captive). Seven loci (complete genotypic profiles) were used to identify 25 different individuals. Analysis with GIMLET resulted in a probability of identity for siblings ($P_{ID(sibs)}$) value of 0.0005 for 10 loci (Fig. 3) for the Bandipur scat samples.

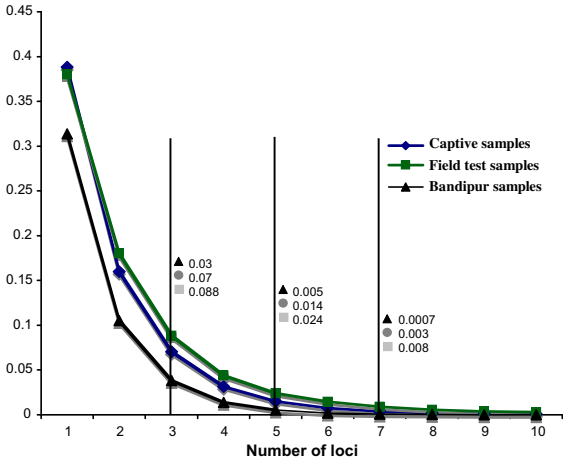


Fig. 3. A plot showing the effect of number of microsatellites on the probability of identity assuming all individuals are siblings $P_{ID(sibs)}$ for a set of captive ($n = 14$), field test ($n = 10$) and Bandipur samples ($n = 25$). In this plot the number of loci are taken in the order of decreasing information contents. In other word, the locus having the highest power to discriminate among the individuals is the first one and the least one is in the end. The order of the loci (1–10) is as follows: FCA628, FCA205, FCA441, FCA672, FCA126, FCA232, F41, FCA391, FCA453 and F115.

For capture–recapture analyses we set a conservative cutoff level for probability of identity, using prior knowledge that tiger population size was likely to be <80 individuals in the study area based on earlier work (Karanth et al., 2008, unpublished report). We used a $P_{ID(sibs)}$ value of 5×10^{-3}, which corresponds to one in every 200 tigers as a minimum for individual identification based on suggestions of earlier researchers (Waits et al., 2001; Miller et al., 2005). Fig. 3 reveals that to achieve this degree of certainty, five loci are required for the Bandipur samples, suggesting Bandipur tigers are more polymorphic than tigers involved in our 'captive test' or 'wild' field test samples.

After grouping the incomplete genotypes with the complete ones we identified 26 unique genotype profiles from the samples based on a five loci criteria. Fig. 4 highlights the individual and locus-specific quality indices for these 26 unique individuals. All five loci were among the eight most informative loci (based on the $P_{ID(sibs)}$ analyses, loci FCA628, FCA205, FCA126, FCA672, FCA232, FCA441, F41 and FCA391) of the complete ten loci panel. (See Fig. 4)

Table 3
Genetic variability at 10 microsatellites for 50 wild samples (scat) from Bandipur National Park.

Locus	Product size range (bp)	Amplification success (%)	Allelic dropout (%)	False alleles	H_e	H_o	No. of alleles
FCA453	180–200	74	0	0	0.74	0.19	5
FCA391	189–217	78	0.017	0	0.68	0.4	7
FCA628	86–110	96	0.007	0	0.88	0.7	12
FCA205	94–112	100	0.013	0	0.85	0.94	9
FCA126	132–154	96	0.015	0	0.79	0.91	11
F41	112–144	85	0	0	0.74	0.4	8
FCA232	96–124	96	0	0	0.77	0.6	11
FCA441	138–166	96	0	0	0.78	0.53	8
FCA672	082–104	96	0.007	0	0.76	0.81	9
F115	192–212	85	0.008	0	0.66	0.45	6

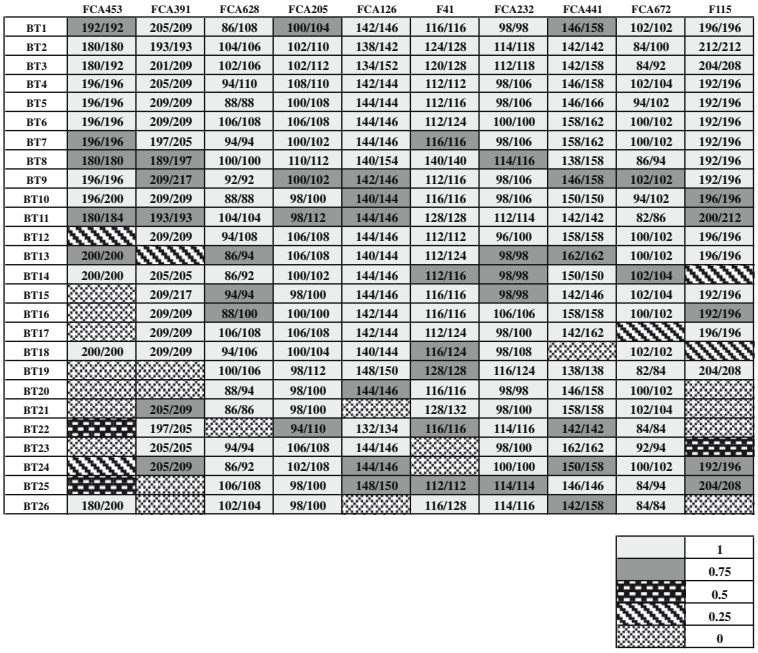


Fig. 4. A figure showing the quality index in a matrix form for all loci (in columns) across all the 26 different individuals (rows) based on five loci criteria. The quality index value of each cell in this matrix can be 0, 0.25, 0.5, 0.75 and 1 (as the genotypes are repeated four times). A quality index of 0.5 represents a scenario where two out of four repetitions resulted in identical genotypes. Note that all sets of five loci (out of the ten loci) have the same cumulative $P_{ID(sibs)}$ value (0.005). Validated genotypes (having quality index of $\geqslant$0.75) are shown for all the different individuals.

Similarly, based on the three loci criteria we were able to identify 29 different genotype profiles. The number of individuals increases when three loci were used because for this analysis we could use more number of samples (n = 41) in comparison to the other two cases (n = 38 for five loci and n = 34 for seven loci).

3.4. Model choice and tiger abundance estimates

Because of the short survey duration of 42 days, we could reasonably assume demographic closure (Otis et al., 1978; Karanth and Nichols, 1998), with no births, deaths, immigration or emigration. Results of the closure test (Otis et al., 1978) in program CAPTURE (z = 1.39, p = 0.92), as well result of the Stanley and Burnham (1999) closure test in program CloseTest (χ^2 = 3.65, df = 4, p = 0.46), supported our assumption of demographic closure during the genetic sampling survey (Table 4).

Following standard practice (Waits et al., 2001; Miller et al., 2005) in genetic studies and based on our results (Tables 1–3), we could assume that individual tigers were identified with sufficient certainty from Bandipur samples when five loci were used in the process. This process resulted in 'captures' of 26 distinct individual tigers during our survey. We analyzed their capture histories (Otis et al., 1978) using the closed model M_h Jackknife estimator implemented in program CAPTURE (Rexstad and Burnham, 1993). In addition, we also analyzed capture histories based on genetic recaptures using three and seven loci, to test the effects of the number of loci on population estimates.

Table 4
Results of the Otis et al. (1978) and Stanley and Burnham (1999) tests for demographic closure.

Dataset	Program CAPTURE[a]		Program CloseTest[b]		
	Z-statistic	P-value	χ^2 statistic	df	P-value
Genetic captures: 3 loci	1.19	0.88	3.40	6	0.76
Genetic captures: 5 loci	1.39	0.92	3.65	4	0.45
Genetic captures: 7 loci	0.97	0.83	3.36	4	0.50
Photographic captures	−1.22	0.11	3.22	9	0.95

[a] Based on Otis et al. (1978).
[b] Based on Stanley and Burnham (1999).

Although the overall model comparison results based on discriminant function test scored the M_0 null model higher than model M_h (score of 1.0 versus 0.77) we note that heterogeneity is likely to exist in tiger populations because of social spacing mechanisms (Karanth and Sunquist, 2000) and differential exposure of tigers to survey routes (Karanth and Nichols, 1998; Borchers and Efford, 2008; Royle et al., 2009). Therefore, we believe the results using the heterogeneity model M_h are more reliable. The tiger abundance estimates from both CAPTURE analyses (based on genetic captures using five loci) are reported in Table 5. These results have generated tiger abundance estimates of $\hat{N}$ ($S\hat{E}[\hat{N}] = 47(11.21)$) for the 'null' model, and, $\hat{N}$ ($S\hat{E}[\hat{N}] = 66(12.98)$) for the 'heterogeneity' model, respectively. We also report results generated by heterogeneity models under the method of finite mixtures (Pledger, 2000) implemented in program MARK (Table 5), although we rely on results generated by the M_h Jackknife estimator in further discussions of our results.

To validate these DNA-based tiger abundance estimates, we compared them to estimates from an earlier photographic capture–recapture survey (Karanth et al., 2008, unpublished data) of the same tiger population (Table 6). These comparisons show that whereas DNA studies (using five loci) identified 26 unique individual tigers in the sampled population, the photographic survey had yielded 29 individuals. As with the genetic survey, the photographic capture survey results also supported the assumption of population closure (Table 4). Both models M_0 and M_h-Jackknife, in program CAPTURE, had high model selection test scores (0.97 versus 0.94) for the photographic-capture data set. The average tiger abundance estimates from photographic capture–recapture surveys were $\hat{N}$ ($S\hat{E}[\hat{N}]$) = 81 (27.7) and $\hat{N}$ ($S\hat{E}[\hat{N}]$) = 66 (13.8) tigers, under the M_0 and M_h models, respectively (Table 6).

Table 5
Abundance estimation of tigers in Bandipur National Park by capture–recapture sampling using DNA collected from scats using five microsatellite loci for individual identification. Models from programs CAPTURE and MARK.

Individuals caught M_{t+1}	Total captures n	CAPTURE models			MARK models	
		Parameter	M_0	M_h-Jackknife	M_0	M_h-*2 point mixture*
26	35	Average per sample capture probability $\hat{p}$	0.13	0.09	0.12	$0.06_{\text{Group 1}}$ $0.44_{\text{Group 2}}$
		Total abundance, $\hat{N}(S\hat{E}[\hat{N}])$	47 (11.21)	66 (12.98)	47 (11.42)	74 (47.49)

Models: M_0 – Otis et al. (1978).
M_h-Jackknife – Burnham and Overton (1978).
M_h-2 point mixture – Pledger (2000).

Table 6
Abundance estimation of tigers in Bandipur National Park by photographic capture–recapture sampling. Models from programs CAPTURE and MARK.

Individuals caught M_{t+1}	Total captures n.	CAPTURE models			MARK models	
		Parameter	M_0	M_h-Jackknife	M_0	M_h-*2 point mixture*
29	35	Average per sample capture probability $\hat{p}$	0.04	0.04	0.04	$0.51_{\text{Group 1}}$ $0.04_{\text{Group 2}}$
		Total abundance, $\hat{N}(S\hat{E}[\hat{N}])$	81 (27.7)	66 (13.81)	81 (28.17)	81 (28.17)

4. Discussion

4.1. Protocol for individual identification

In comparison to other studies involving tiger individual identification using genetic data (Xu et al., 2005; Bhagavatula and Singh, 2006) we used a large panel of 30 microsatellite loci for the initial assessment of variability/microsatellite suitability. This allowed us to select the combination of loci that result in lower probability of identity for siblings. This is particularly relevant given that tigers have relatively low genetic diversity (Luo et al., 2004), making the individual identification more demanding than other species possessing higher genetic variation. Molecular screening of a larger number of loci did not significantly improve our P_{ID} (0.005, with six loci (Bhagavatula and Singh, 2006 versus five loci in our study). However, the amplification success of our selected loci was high (>85% for the Bandipur samples in our study compared to 60% in Bhagavatula and Singh, 2006) and allelic dropout was lower (between 0.5 and 1.5%).

Under field conditions the 'freshness' and 'quality' of tiger scats can vary greatly among samples. As a result, scats from captive tigers yield higher quantity and quality of DNA. Our results support this hypothesis, as the captive scats amplified more loci (19 versus 10) than the field test scats, and amplification success was also higher on an average (100% versus 90%).

Using distinct captive and wild individuals for investigating probability of identity revealed that wild test samples were less polymorphic than captive individuals, possibly because of high outbreeding in zoo tigers. Polymorphism was highest for the Bandipur samples, possibly because we had relatively higher numbers of individuals (25 versus 10 and 13) from the Bandipur survey population.

The quality index measure suggested by Miquel et al. (2006) was initially proposed for pilot studies on genotype data quality/reliability, and allows better data quality management within a study and comparability across studies. To our knowledge, ours is the first genetics-based individual identification study for any elusive large carnivore, in which the quality index approach has been used for data validation.

4.2. Modeling and estimation of abundance from tiger capture data

We used well-founded standard capture–recapture models (Otis et al., 1978; Williams et al., 2002; Amstrup et al., 2005) in which the influence of factors such as time, trap–response and individual heterogeneity on tiger capture probabilities can be assessed as opposed to the null model which assumes that these factors do not affect capture probabilities. Our results reveal that various CR analytical approaches produced similar estimates even when five loci are used in genetic identifications. However, we note that unlike with physical tagging or photographic captures, in the case of 'genetic captures', individual capture probabilities may additionally vary depending both on number of scats produced and the quality of DNA available in the scats collected non-invasively from different individuals (Broquet et al., 2007). We believe that capture–recapture analyses, which can explicitly deal with potentially complex forms of individual heterogeneity (such as the Jackknife estimator under model M_h in program CAPTURE), might be more appropriate for analysis of genetic capture data from tigers. Our analyses also demonstrated the robustness (Otis et al., 1978) in the parameter estimates under the M_h-Jackknife model to small variations in recapture rates, when we compare results from the photographic and genetic surveys using the same model (see Tables 5 and 6).

Since the precision of individual identification is highly dependent on the number of loci used in comparisons (through their effect on P_{ID}), we explored the effects of using different numbers of loci on tiger population estimates (Table 7). As expected, using very few loci (three loci) results in underestimates of population size due to low probability of identity for siblings resulting from 'shadow effect'. Interestingly, using five versus seven loci does not result in very different population estimates, despite use of seven loci reducing the number of identified individuals to 25. Our results are thus in contrast to those of Creel et al. (2003) who found that increasing number of loci for individual identification could result in artificially creating new individuals due to genotyping errors. Since the number of 'recaptures' is also reduced as the number of loci used for individual identification is increased, the estimated mean population size should not differ greatly, especially under the robust M_h-Jackknife model. These results demonstrate that five loci may be sufficient to accurately identify individual tigers in the context of abundance estimation.

4.3. Utility of genetic capture–recapture sampling for estimating tiger abundance

Overall, we note that the tiger abundance estimates derived from genetic CR sampling closely match abundance estimates generated from photographic CR sampling, which is a standard

Table 7
Abundance estimation of tigers in Bandipur National Park by capture–recapture sampling using DNA collected from scats using 3 and 7 microsatellite loci for individual identification. Models from programs CAPTURE and MARK.

No. of loci	Individuals caught M_{t+1}	Total captures n	CAPTURE models			MARK models	
			Parameter	M_0	M_h-Jackknife	M_0	M_h-*2 point mixture*
3	29	40	Average per sample capture probability $\hat{p}$	0.13	0.11	0.13	$0.46_{\text{Group 1}}$ $0.09_{\text{Group 2}}$
			Total abundance, $\hat{N}(S\hat{E}[\hat{N}])$	49 (10.49)	62 (11.76)	49.44 (10.65)	61.93 (24.03)
7	25	31	Average per sample capture probability $\hat{p}$	0.09	0.08	0.09	$0.55_{\text{Group 1}}$ $0.09_{\text{Group 2}}$
			Total abundance, $\hat{N}(S\hat{E}[\hat{N}])$	57 (18.36)	66 (13.04)	57 (18.74)	57 (18.74)

method successfully used for estimating abundances in populations of tigers (Karanth and Nichols, 1998; Karanth et al., 2004a; Wegge et al., 2004; Kawanishi and Sunquist, 2004; Simchareon et al., 2007) and other elusive carnivores (Henschel and Ray, 2003; Trolle and Kery, 2003; Silver et al., 2004). In this context, our results are in contrast to several other studies, where genetic capture–recapture yielded higher (Zhan et al., 2006; Creel et al., 2003; Fickel and Hohmann, 2006) or lower (Wilson et al., 2003) abundance estimates compared to more traditional survey methods. We further note that photographic capture–recapture estimates of tiger abundance reported here have additionally been validated against independent estimates based on density of principal ungulate prey and food intake rates of tigers (Karanth et al., 2004b). Overall carefully conducted non-invasive genetic-capture sampling leading to high degree of certainty in individual animal identifications appears to be a reliable approach for estimating abundances of tigers and other species. Furthermore, at a metapopulation level, for estimating long-distance dispersal movement, genetic sampling of individual tigers appears to be a potentially very attractive tool, because photographic capture surveys are difficult to conduct at regional spatial scales.

We see genetic sampling as a valuable additional tool for estimating tiger abundance in situations such as the Russian Far East, Sundarban mangrove swamps and rainforests of Souteastern Asia, where densities are low and camera trapping is often impractical due to various environmental and logistical constraints. Even where camera trap methods work well, faecal genetic sampling may provide additional data on age–sex structure and relatedness among individuals, information useful for answering some scientific or management questions.

However, with faecal DNA capture methods there is some potential for a particular age–sex class being virtually undetectable on account of behavioral reasons. For example, if transient individual tigers very rarely defecate on regular travel routes to avoid detections by territorial residents, scat sampling may underestimate that segment of the population. However, we note that trained dogs have been employed to increase detections rates of carnivore scats (Wasser et al., 2004; Nagata et al., 2005; Kerley and Salkina, 2006), thus offering an avenue to overcome such biases.

We have used conventional capture–recapture models in our analyses here, particularly relying on the M_h Jackknife estimator that performed well in earlier studies of tigers (Karanth et al., 2004b). However, we note that recently developed spatial capture–recapture models (Royle and Young, 2008; Borchers and Efford, 2008; Royle et al., 2009) could potentially offer a superior analytical framework for data analysis eventually when they are explicitly designed for scat detection surveys. Therefore, designing scat sampling surveys explicitly to meet capture–recapture modeling assumptions, particularly via the construction of sampling occasions that generate additional information on sampling effort is advantageous (Williams et al., 2002; Amstrup et al., 2005; Royle et al., 2009) over ad hoc surveys employed in many genetic sampling studies (Fickel and Hohmann, 2006; Bhagavatula and Singh, 2006). We believe that greater integration with formal capture–recapture sampling approaches will allow non-invasive genetic methods to become an even more important tool in the future for research on tigers and other endangered species.

In the broader context of animal abundance estimation, genetic-capture sampling provides a useful alternative wherever photographic sampling or conventional tagging studies cannot be employed for reasons such as lack of distinguishing natural marks on study species, logistical difficulties of physical captures, low-animal density, lack of skilled survey personnel or camera equipment, potential for camera theft/damage and sometimes even survey cost considerations. Genetic capture method may also be easier to apply in broader surveys that target multiple species, as opposed to camera-trap or physical tagging surveys, which are typically tailored specifically for one or two species at a time. The problem of evasive animal response to conventional traps or cameras by some elusive species can also be overcome with faecal DNA surveys.

Acknowledgements

We are grateful to the forest department of Karnataka for according permissions to carry out the field work in Bandipur. We also thank the local forest officials and staff at Bandipur for their assistance. We thank the forest department of Kerala for permission to obtain genetic samples from captive tigers at the Thrissur zoo. We acknowledge the veterinarians and support staff at the Zoos in Mysore, Bannerghatta and Thrissur who helped with genetic sampling of captive tigers. Our special thanks go to late Krishna Narain as well as to Dr. Abraham David for their assistance in obtaining genetic samples and to Raghavendra Mogaroy, Narendra Patil and Devcharan Jathanna for help with analytical work. We thank JD Nichols, S Mukherjee, J Krishnaswamy and A Kumar for helpful discussions. We thank the NCBS sequencing facility and specifically S Gyanam for help with genotyping. SM and UR were partially supported by a grant from the Wildlife Conservation Society. This research was funded by a startup grant to UR from the National Centre for Biological Sciences and a grant from the Department of Science and Technology (Government of India) to KUK and UR. Wildlife Conservation Society India Program and Centre for Wildlife Studies provided the services of KUK, NSK and AMG as well as substantial logistical and material support for the fieldwork. We thank the two anonymous referees for their constructive comments and helpful suggestions, which greatly improved this manuscript.

References

Amstrup, S.C., McDonald, T.L., Manly, B.F.J., 2005. Handbook of Capture–Recapture Analysis. Princeton University Press, Princeton, NJ.

Andheria, A.P., 2006. Assessment of diet and abundance of large carnivores from field surveys of scats. M.Sc. Thesis, Manipal University.

Banks, S.C., Piggott, M.P., Hansen, B.D., Robinson, N.A., Taylor, A.C., 2002. Wombat coprogenetics: enumerating a common wombat population by microsatellite analysis of faecal DNA. Australian Journal of Zoology 50, 193–204.

Bhagavatula, J., Singh, L., 2006. Genotyping faecal samples of Bengal tiger (*Panthera tigris tigris*) for population estimation: a pilot study. BMC Genetics 7, 48.
Boersen, M.R., Clark, J.D., King, T.L., 2003. Estimating black bear population density and genetic diversity at Tensas River, Louisiana using microsatellite DNA markers. Wildlife Society Bulletin 31, 197–207.
Borchers, D.L., Efford, M.G., 2008. Spatially explicit maximum likelihood methods for capture–recapture studies. Biometrics 64, 377–385.
Broquet, T., Menard, N., Petit, E., 2007. Noninvasive population genetics: a review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. Conservation Genetics 8, 249–260.
Burnham, K.P., Overton, W.S., 1978. Estimation of the size of a closed population when capture probabilities vary among animals. Biometrika 65, 625–633.
Creel, S., Spong, G., Sands, J.L., Rotella, J., Zeigle, J., Joe, J., Murphy, K.M., Smith, D., 2003. Population size estimation in Yellowstone wolves with error-prone noninvasive microsatellite genotypes. Molecular Ecology 12, 2003–2009.
Davison, A., Birks, J.D.S., Brookes, R.C., Braithwaithe, T.C., Messenger, J.E., 2002. On the origin of faeces: morphological versus molecular methods for surveying rare carnivores from their scats. Journal of Zoology 257, 141–143.
Dreher, B.P., Winterstein, S.R., Scribner, K.T., Lukacs, P.M., Etter, D.R., Rosa, G.J.M., Lopez, V.A., Libants, S., Filcek, K.B., 2007. Noninvasive estimation of black bear abundance incorporating genotyping errors and harvested bear. Journal of Wildlife Management 71, 2684–2693.
Eggert, L.S., Eggert, J.A., Woodruff, D.S., 2003. Estimating population sizes for elusive animals: the forest elephants of Kakum National Park, Ghana. Molecular Ecology 12, 1389–1402.
Fickel, J., Hohmann, U., 2006. A methodological approach for non-invasive sampling for population size estimates in wild boars (*Sus scrofa*). European Journal of Wildlife Research 52, 28–33.
Flagstad, Ø., Hedmark, E., Landa, A., Broseth, H., Persson, J., Andersen, R., Segeratrom, P., Ellegren, H., 2004. Colonization history and non-invasive monitoring of a re-established wolverine population. Conservation Biology 18, 676–688.
Frantz, A.C., Pope, L.C., Carpenter, P.J., Roper, T.J., Wilson, G.J., Delahay, R.J., Burke, T., 2003. Reliable microsatellite genotyping of the Eurasian badger (*Meles meles*) using faecal DNA. Molecular Ecology 12, 1649–1661.
Frantz, A.C., Schaul, M., Pope, L.C., Fack, F., Schley, L., Muller, C.P., Roper, T.J., 2004. Estimating population size by genotyping remotely plucked hair: the Eurasian badger. Journal of Applied Ecology 41, 985–995.
Henschel, P., Ray, J., 2003. Leopards in African Rainforests: Survey and Monitoring Techniques, Global Carnivore Program, Wildlife Conservation Society. (<http://www.savingwildplaces.com/media/file/low-leopard.pdf>).
Karanth, K.U., Nichols, J.D., 1998. Estimating tiger densities in India from camera trap data using photographic captures and recaptures. Ecology 79, 2852–2862.
Karanth, K.U., Sunquist, M.E., 2000. Behavioral correlates of predation by tiger, leopard and dhole in Nagarahole, India. Journal of Zoology 250, 255–265.
Karanth, K.U., Nichols, J.D., 2002. Monitoring Tigers and their Prey: A Manual for Researchers, Managers and Conservationists in Tropical Asia. Centre for Wildlife Studies, Bangalore.
Karanth, K.U., Nichols, J.D., Kumar, N.S., 2004a. Photographic sampling of elusive mammals in tropical forests. In: Thompson, W.L. (Ed.), Sampling Rare or Elusive Species. Island Press, Washington, DC, pp. 229–247.
Karanth, K.U., Nichols, J.D., Kumar, N.S., Link, W.A., Hines, J.E., 2004b. Tigers and their prey: predicting carnivore densities from prey abundance. Proceedings of National Academy of Sciences of USA 101, 4854–4858.
Karanth, K.U., 1995. Estimating tiger *Panthera tigris* populations from camera-trap data using capture recapture models. Biological Conservation 71, 333–338.
Karanth, K.U., Nichols, J.D., Kumar, N.S., Hines, J.E., 2006. Assessing tiger population dynamics using photographic capture–recapture sampling. Ecology 87, 2925–2937.
Karanth, K. U., Kumar, N. S., Gopalaswamy, A., Srinivas, V., 2008. Distribution and Dynamics of Tiger and Prey Populations in Karnataka. Centre for Wildlife Studies (Unpublished report).
Kawanishi, K., Sunquist, M.E., 2004. Conservation status of tigers in a primary rainforest of Peninsular Malaysia. Biological Conservation 120, 329–344.
Kerley, L.L., Salkina, G.P., 2006. Using scent-matching dogs to identify individual Amur tigers from scats. Journal of Wildlife Management 71, 1349–1356.
Kishore, R., Hardy, W.R., Anderson, V.J., Sanchez, B.S., Buoncristiani, M.R., 2006. Optimization of DNA extraction from low-yield and degraded samples using the Biorobot EZ1 and Biorobot M48. Journal of Forensic Science 51, 1055–1061.
Kohn, M.H., York, E.C., Kamradt, D.A., Haught, G., Sauvajot, R.M., Wayne, R.K., 1999. Estimating population size by genotyping faeces. Proceedings of Royal Society B 266, 657–663.
Lukacs, P.M., Burnham, K.P., 2005. Review of capture–recapture methods applicable to noninvasive genetic sampling. Molecular Ecology 14, 3909–3919.
Luo, S-J., Kim, J-H., Johnson, W.E., et al., 2004. Phylogeography and genetic ancestry of tigers (*Panthera tigris*). PLoS Biology 2, 2275–2293.
Marshall, T.C., Slate, J., Kruuk, L.E.B., Pemberton, J.M., 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology 7, 639–655.
McKelvey, K.S., Schwartz, M.K., 2004. Providing reliable and accurate capture mark-recapture estimates in a cost–effective way. Journal of Wildlife Management 68, 453–456.
Menotti-Raymond, M., David, V.A., Lyons, L.A., Schaffer, A.A., Tomlin, J.L., Hutton, M.K., O'Brien, S.J., 1999. A genetic linkage map of microsatellites of the domestic cat (*Felis catus*). Genomics 57, 9–23.
Miller, C.R., Joyce, P., Waits, L.P., 2005. A new method for estimating the size of small populations from genetic mark-recapture data. Molecular Ecology 14, 1991–2005.
Mills, L.S., Citta, J.J., Lair, K., Schwartz, M., Tallmon, D., 2000. Estimating animal abundance using non-invasive DNA sampling: promise and pitfalls. Ecological Applications 10, 283–294.
Miquel, C., Bellemain, E., Poillot, C., Bessiere, J., Durand, A., Taberlet, P., 2006. Quality indexes to assess the reliability of genotypes in studies using noninvasive sampling and multiple-tube approach. Molecular Ecology Notes 6, 985–988.
Morell, V., 2007. Can the wild tiger survive? Science 317, 1312–1314.
Mowat, G., Paetkau, D., 2002. Estimating marten *Martes americana* population size using hair capture and genetic tagging. Wildlife Biology 8, 201–209.
Mukherjee, N., Mondol, S., Andheria, A., Ramakrishnan, U., 2007. Rapid multiplex PCR based species identification of wild tigers using non-invasive samples. Conservation genetics 8, 1465–1470.
Murphy, M.A., Waits, L.P., Kendall, K.C., Wasser, S.K., Higbee, J.A., Bogden, R., 2002. An evaluation of long-term preservation methods for brown bear (*Ursus arctos*) faecal DNA samples. Conservation Genetics 3, 435–440.
Nagata, J., Aramilev, V.V., Belozor, A.B., Sugimoto, T., McCullough, D.R., 2005. Fecal genetic analysis using PCR–RFLP of cytochrome *b* to identify sympatric carnivores, the tiger *Panthera tigris* and the leopard *Panthera pardus*, in far eastern Russia. Conservation Genetics 6, 863–866.
Otis, D.L., Burnham, K.P., White, G.C., Anderson, D.R., 1978. Statistical inference from capture data on closed animal populations. Wildlife Monographs 62, 1–135.
Petit, E., Valiere, N., 2006. Estimating population size with noninvasive capture-mark-recapture data. Conservation Biology 20, 1062–1073.
Piggott, M.P., Banks, S.C., Stone, N., Banffy, C., Taylor, A.C., 2006. Estimating population size of endangered brush-tailed rock-wallaby (*Petrogale penicillata*) colonies using faecal DNA. Molecular Ecology 15, 81–91.
Pledger, S., 2000. Unified maximum likelihood estimates for closed capture–recapture models using mixtures. Biometrics 56, 434–442.
Prugh, L.R., Ritland, C.E., Arthur, S.M., Crebs, C.J., 2005. Monitoring coyote population dynamics by genotyping faeces. Molecular Ecology 14, 1585–1596.
Puechmaille, S.J., Petit, E.J., 2007. Empirical evaluation of non-invasive capture–mark–recapture estimation of population size based on a single sampling session. Journal of Applied Ecology 44, 843–852.
Ranganathan, J., Chan, K.M.A., Karanth, K.U., Smith, J.L.D., 2008. Where can tigers persist in the future? A landscape-scale, density-based population model for the India subcontinent. Biological Conservation 141, 67–77.
Rexstad, E., Burnham, K.P., 1993. Users' guide for interactive program CAPTURE. Abundance estimation of closed animal populations. Colorado State University, Fort Collins, CO, USA.
Royle, A.J., Young, K.V., 2008. A hierarchical model for spatial capture-recapture data. Ecology 89, 2281–2289.
Royle, A.J., Nichols, J.D., Karanth, K.U., Gopalaswamy, A.M., 2009. A hierarchical model for estimating density in camera trap studies. Journal of Applied Ecology 46, 118–127.
Rudnick, J.A., Katzner, T.E., Bragin, E.A., DeWoody, J.A., 2008. A non-invasive genetic evaluation of population size, natal philopatry, and roosting behavior of non-breeding eastern imperial eagles (*Aquila heliaca*) in central Asia. Conservation Genetics 9, 667–676.
Sanderson, E., Forrest, J., Loucks, C., et al. (2006) Setting Priorities for the Conservation and Recovery of Wild Tigers: 2005–2015, The Technical Assessment. WCS, WWF, Smithsonian and NFWF-STF, New York, Washington DC, USA.
Schipper, J. et al., 2008. The status of the world's land and marine mammals: diversity, threat, and knowledge. Science 141, 67–77.
Simchareon, S., Pattanavibool, A., Karanth, K.U., Nichols, J.D., Kumar, N.S., 2007. How many tigers (*Panthera tigris*) are there in Huai Kha Khaeng wildlife sanctuary, Thailand? An estimate using photographic capture–recapture sampling. Oryx 41, 1–7.
Silver, S.C., Ostro, L.E.T., Marsh, L.K., Maffei, L., Noss, A.J., Kelly, M.J., Wallece, R.B., Gomez, H., Ayala, G., 2004. The use of camera traps for estimating jaguar (*Panthera onca*) abundance and density using capture/recapture analysis. Oryx 38, 1–7.
Sloane, M.A., Sunnucks, P., Alpers, D., Beheregaray, L.B., Taylor, A.C., 2000. Highly reliable genetic identification of individual northern hairy-nosed wombats from single remotely collected hairs: a feasible censusing method. Molecular Ecology 9, 1233–1240.
Smith, J.L.D., McDougal, C., Miquelle, M., 1989. Scent marking in free-ranging tigers, *Panthera tigris*. Animal Behaviour 37, 1–10.
Solberg, K.H., Bellemain, E., Drageset, O.M., Taberlet, P., Swenson, J.E., 2006. An evaluation of field and non-invasive genetic methods to estimate brown bear (*Ursus arctos*) population size. Biological Conservation 128, 158–168.
Stanley, T.R., Burnham, K.P., 1999. A closure test for time-specific capture-recapture data. Environmental and Ecological Statistics 6, 197–209.
Taberlet, P., Luikart, G., 1999. Non-invasive genetic sampling and individual identification. Biological Journal of Linnean Society 68, 41–55.
Triant, D.A., Pace, R.M., Stine, M., 2004. Abundance, genetic diversity and conservation of Louisiana black bears (*Ursus americanus luteolus*) as detected through noninvasive sampling. Conservation Genetics 5, 647–659.
Trolle, M., Kery, M., 2003. Estimation of ocelot density in the Pantanal using capture-recapture analysis of camera trapping data. Journal of Mammalogy 84, 607–614.

Valière, N., 2002. GIMLET, a computer program for analyzing genetic identification data. Molecular Ecology Notes 2, 377–379.
Waits, L.P., Luikart, G., Taberlet, P., 2001. Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Molecular Ecology 10, 249–256.
Waits, L.P., 2004. Using noninvasive genetic sampling to detect and estimate abundance of rare wildlife species. In: Thompson, W.L. (Ed.), Sampling Rare or Elusive Species: Concepts, designs, and techniques for estimating population parameters. Island press, pp. 211–228. .
Waits, L.P., Paetkau, D., 2005. Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. Journal of Wildlife Management 69, 1419–1433.
Wasser, S.K., Davenport, B., Ramage, E.R., Hunt, K.E., Parker, M., Clarke, C., Stenhouse, G., 2004. Scat detection dogs in wildlife research and management: application to grizzly and black bears in the Yellowhead Ecosystem, Alberta, Canada. Canadian Journal of Zoology 82, 475–492.
Wegge, P., Pokheral, C.P., Jnawali, S.R., 2004. Effects of trapping effort and trap shyness on estimates of tiger abundance from camera trap studies. Animal Conservation 7, 251–256.
White, G.C., Burnham, K.P., 1999. Program MARK: survival rate estimation from both live and dead encounters. Bird Study 46, 120–139.
White, G.C., Anderson, D.R., Burnham, K.P., Otis, D.L., 1982. Capture–Recapture Removal Methods for Sampling Closed Populations. Los Alamos National Laboratory publication no. LA-8787-NERP. Los Alamos, USA.
Williams, B.K., Nichols, J.D., Conroy, M.J., 2002. Analysis and Management of Animal Populations. Academic Press, San Diego.
Williamson, J.E., Heubinger, RM., Sommer, J.A., Louis, E.E., Barber, R.C., 2002. Development and cross-species amplification of 18 microsatellite markers in the Sumatran tiger (*Panthera tigris sumatrae*). Molecular Ecology Notes 2, 110–112.
Wilson, G.J., Franz, A.C., Pope, L.C., Roper, T.J., Cheeseman, C.L., Delahay, R.J., 2003. Estimation of badger abundance using faecal DNA typing. Journal of Applied Ecology 40, 658–666.
Xu, Y.C., Li, B., Li, W.S., Bai, S.Y., Jin, Y., Li, X.P., Gu, M.B., Jing, S.Y., Zhang, W., 2005. Individualization of tiger by using microsatellites. Forensic Science International 151, 45–51.
Zhan, X., Li, M., Zhang, Z., Goossens, B., Chen, Y., Wang, H., Bruford, M., Wei, F., 2006. Molecular censusing doubles giant panda population estimate in a key nature reserve. Current Biology 16, R451–R452.
Zhang, Z.H., Zhang, W.P., Yue, B.S., Shen, F.J., Zhang, L., Hou, R., Zhu, M.Y., 2005. Twelve polymorphic microsatellite loci for the South China tiger *Panthera tigris amoyensis*. Molecular Ecology Notes 6, 24–26.

Appendix C

A hierarchical model for estimating density in camera-trap studies

Journal of Applied Ecology

British Ecological Society

Journal of Applied Ecology 2009, **46**, 118–127 doi: 10.1111/j.1365-2664.2008.01578.x

A hierarchical model for estimating density in camera-trap studies

J. Andrew Royle[1]*, James D. Nichols[1], K. Ullas Karanth[2] and Arjun M. Gopalaswamy[2]

[1]*US Geological Survey, Patuxent Wildlife Research Center, Laurel, Maryland 20708, USA; and* [2]*Wildlife Conservation Society-India Program, Centre for Wildlife Studies, Bangalore, Karnataka-560042, India*

Summary

1. Estimating animal density using capture–recapture data from arrays of detection devices such as camera traps has been problematic due to the movement of individuals and heterogeneity in capture probability among them induced by differential exposure to trapping.

2. We develop a spatial capture–recapture model for estimating density from camera-trapping data which contains explicit models for the spatial point process governing the distribution of individuals and their exposure to and detection by traps.

3. We adopt a Bayesian approach to analysis of the hierarchical model using the technique of data augmentation.

4. The model is applied to photographic capture–recapture data on tigers *Panthera tigris* in Nagarahole reserve, India. Using this model, we estimate the density of tigers to be 14·3 animals per 100 km^2 during 2004.

5. *Synthesis and applications.* Our modelling framework largely overcomes several weaknesses in conventional approaches to the estimation of animal density from trap arrays. It effectively deals with key problems such as individual heterogeneity in capture probabilities, movement of traps, presence of potential 'holes' in the array and *ad hoc* estimation of sample area. The formulation, thus, greatly enhances flexibility in the conduct of field surveys as well as in the analysis of data, from studies that may involve physical, photographic or DNA-based 'captures' of individual animals.

Key-words: Bayesian analysis, camera trapping, carnivore surveys, density estimation, hierarchical model, Markov chain Monte Carlo, point process, spatial capture–recapture, tigers, trapping grid

Introduction

Abundance is a parameter of central interest to animal ecologists because it affects virtually all other aspects of biology or conservation of the species of interest. Consequently, estimating abundance is a fundamental goal of many animal sampling problems, and it forms the basis of a vast body of literature on statistical methods in animal ecology as exemplified in the synthetic treatments by Seber (1982), Borchers, Buckland & Zucchini (2002), and Williams, Nichols & Conroy (2002). The main consideration in estimating abundance of most animal populations is that individuals cannot be observed or detected perfectly. That is, the probability of encountering or detecting an animal is less than 1. A myriad of techniques have been devised for dealing with imperfect detection so that estimates of N may be obtained from observations of individuals (Williams *et al*. 2002).

*Correspondence author. E-mail: aroyle@usgs.gov

Capture–recapture methods in particular are frequently applied to estimate abundance using encounter history data from arrays of traps or passive detection devices. These include classical animal trapping grids, general photography (Goswami, Madhusudan & Karanth 2007), camera traps and DNA collected from dung or 'hair snares' (Waits 2004). It is natural to view the resulting capture–recapture data as being generated by a closed population provided the sample occasions are close enough together in time to preclude mortality, recruitment and movement. However, lacking a physical barrier around the sampled area, there is almost always some movement of individuals onto and off the area during the survey – referred to as an absence of geographic closure (e.g. White *et al*. 1982). This phenomenon in the simplest case ('random temporary emigration', Kendall, Nichols & Hines 1997) biases $\hat{p}$ low, hence $\hat{N}$ high, with respect to the nominal area over which traps are physically situated. While we may be able to obtain an estimate of N that is relevant to a real population of individuals, we do not know the area from

© 2008 The Authors. Journal compilation © 2008 British Ecological Society

which animals were sampled. In addition to movement of individuals complicating the interpretation of abundance estimates, another problem that arises in capture–recapture studies based on spatially organized trap arrays is that the probabilistic models of the capture process (e.g. Otis *et al.* 1978) used to estimate N are no longer strictly appropriate. Specifically, heterogeneity in capture probability can arise as a result of individual locations relative to the array location. That is, some individuals experience more exposure to trapping than others. These problems have received considerable attention in the literature (Dice 1938, 1941; Stickel 1954; Smith *et al.* 1971; Otis *et al.* 1978; Wilson & Anderson 1985) and a large number of *ad hoc* inference procedures for spatial capture–recapture data have been proposed. Conversely, formal model-based inference procedures have only recently been developed (Efford 2004; Borchers & Efford 2008; Royle & Young 2008).

In this study, we consider the problem of estimating tiger density from camera-trap arrays in India. The use of camera traps to estimate tiger abundance was first proposed by Karanth (1995), and received further development by Karanth & Nichols (1998, 2000, 2002). It has since been widely used on several individually identifiable carnivore species (O'Brien, Kinnaird & Wibisono 2003; Trolle & Kéry 2003; Wallace *et al.* 2003; Karanth *et al.* 2004a,b, 2006 Silver *et al.* 2004; Wegge *et al.* 2004; Soisalo & Cavalcanti 2006). Density estimation using photographic capture–recapture methods has so far emphasized *ad hoc* approaches based on estimates of boundary strip width. This is viewed as the weak link in this widely used methodology (Karanth *et al.* 2006; Soisalo & Cavalcanti 2006). Soisalo & Cavalcanti (2006) noted that different estimation methods yielded very different density estimates for jaguars *Panthera onca*. Such differences have important conservation consequences. Because of the relationship between attainable tiger densities and prey densities (e.g. Karanth *et al.* 2004b), tiger management in India is largely a matter of managing prey species. Expected tiger densities can be predicted from estimates of available prey. Both the analysis of these relationships and the subsequent comparisons of predicted and realized tiger numbers depend on reasonable estimates of tiger densities.

We propose a spatial-explicit capture–recapture model for estimating density from trapping arrays that yield encounter histories of individuals in addition to auxiliary spatial information – the traps in which each individual is captured. Motivated by a similar problem (Royle & Young 2008), we formulate a hierarchical model for the observations (the locations of trapping) that is specified conditional on a set of latent variables (random effects). These latent variables are the hypothetical home range or activity centres of every individual in the population. The objective is to estimate the absolute density of home range centres in the region containing the trap array. Classical inference under such models is achieved by removing the latent variables from the conditional likelihood by integration (see Borchers & Efford 2008). Here, we adopt an approach based on Bayesian analysis of the hierarchical model using Markov chain Monte Carlo (MCMC). In particular, we make use of the method of data augmentation (Royle, Dorazio & Link 2007), which yields a simple treatment of this problem as a 'missing data' problem, where the home range centres are the missing data (Royle & Young 2008). Formally, data augmentation yields a parameterization of the model that can be implemented in the freely available software WinBUGS (Gilks, Thomas & Spiegelhalter 1994) with little more than a few lines of 'pseudo-code' to describe the model.

Model development

Suppose that an array of R traps is established, having locations $(\mathbf{x}_j; j = 1, 2, \ldots, R)$. We denote the set of trap locations as χ. Suppose that a population consists of N individuals potentially exposed to sampling, and that each individual in the population has a centre of activity, which we denote by $\boldsymbol{s}_i$ for individuals $i = 1, 2, \ldots, N$. The centre of activity is the two-dimensional coordinate $\boldsymbol{s}_i = (s_{1i}, s_{2i})$. This construction has been previously adopted in similar contexts (Efford 2004; Royle & Young 2008). We assume that the activity centres are fixed for the period over which the population is sampled. Thus, individuals may move around their activity centre, but their activity centre does not move.

The model is developed conditional on the individual activity centres $\boldsymbol{s}_i$ which are unknown. Following Efford (2004), Borchers & Efford (2008) and Royle & Young (2008), we regard these as random effects, and assume that they are uniformly distributed over some region S that contains the traps. We express this assumption by:

$$\mathbf{s}_i \sim \text{Uniform}(S)$$

Thus, the probability mass function for s_i is constant on S. The inference objective is estimation of the density of activity centres located within S. Alternatively, we may be interested in estimating population size or density over some polygon or formal subset of S. For example, a polygon might describe the boundaries of a National Park or a Reserve. Obviously, estimating the density with S or an arbitrary subset is equivalent (by a linear transformation) to estimating the *number* of activity centres within that polygon. In other words, if $\mathcal{P}$ is any polygon having area $A(\mathcal{P})$, and $N(\mathcal{P})$ is the number of activity centres in $\mathcal{P}$, then the density is $D(\mathcal{P}) = N(\mathcal{P})/A(\mathcal{P})$.

The observations are denoted by h_{it}, which are integers $h_{it} \in (0, 1, 2, \ldots, R)$, where $h_{it} = 0$ indicates no capture of individual i at sampling occasion t, and nonzero values indicate the trap location at which the animal is caught at occasion t. For example, if $T = 5$ and $R = 10$ traps, then the encounter history $\boldsymbol{h}_i = (0, 8, 9, 0, 8)$ indicates an individual not captured in occasions 1 and 4, and captured in trap number 8 (occasions 2 and 5) and number 9 (occasion 3). We assume that h_{it}; $t = 1, 2, \ldots, T$ are independent observations of a categorical random variable with probabilities $\boldsymbol{\pi}_i$ which depend on the latent variable $\boldsymbol{s}_i$ and $\mathbf{x}_j$. Note that is $\boldsymbol{\pi}_i$ is a vector of length $R + 1$, with one of the cells corresponding to the

© 2008 The Authors. Journal compilation © 2008 British Ecological Society,

event 'not captured.' The model can be thought of as describing the result of rolling an $R + 1$-sided die. We denote the observation model assumption by:

$$h_{it} \mid \boldsymbol{s}_i, \theta \sim \text{Categorical}(\boldsymbol{\pi}_i(\boldsymbol{s}_i, \theta)) \qquad \text{eqn 1}$$

As indicated here, the cell probabilities are assumed to depend on some parameter(s) θ, in a manner described subsequently. There is an equivalence between a categorical random variable and a multinomial trial. If we create a vector of 0s, of length $R + 1$, and insert a single 1 into that vector in position $h_{it} + 1$, then the resulting vector is a multinomial trial.

That we assume h_{it} to be categorical random variables that are independent among survey occasions for each individual *and* among individuals has several implications. In particular, it implies that multiple individuals may be captured in the same trap during any occasion. For camera traps this provides an accurate characterization of the method since cameras operate continuously (under normal circumstances). Also, the model implies that capture in traps for each individual are mutually exclusive events. This means: (i) individuals cannot be captured more than one time in the same trap for each period; and (ii) individuals cannot be captured in more than one trap. While the first is technically incorrect, since cameras (usually) operate continuously between checks, as an operational matter when the sampling interval is short (e.g. daily), data are processed into single trap-specific encounters of individuals for each sampling occasion (i.e., in Karanth & Nichols (1998, 2000, 2002)). Thus, the first implication is effectively satisfied under the current manner in which the data are processed. With regard to the second implication, in the data we analyse below, there are instances of individuals captured in multiple traps during the same daily sampling interval. Historically, (Karanth & Nichols 1998, 2000, 2002) multiple captures have been used to estimate the distance moved for calculation of the buffer around the trap array, but not in the definition of the encounter histories. In our analysis, we regard the multiple captures as *iid* multinomial trials with the belief that, for low capture rates, this should be reasonably efficient as the multiple captures are providing information on the movement process (see below). Moreover, given the scarcity of the data (low number of captures of each individual), this is preferred to discarding data.

Modelling exposure to trapping

Our development of the specific form of the cell probabilities $\boldsymbol{\pi}$ is based on thinking about exposure to trapping that results from animal movements about their activity centre $\boldsymbol{s}$. To motivate this 'movement/exposure process' in the context of trapping arrays, we draw on a related model described by Royle & Young (2008; henceforth 'RY'). RY described a survey of lizards based on repeated searches of quadrats using a team of observers. They described the encounter process explicitly in terms of individual movements. In fact, in camera-trapping problems, the linkage between encounter and movement/exposure is imprecise because exposure to trapping and encounter does not occur independently. Technical 'exposure' to a camera trap necessarily yields a detection (unless the camera malfunctions, which is not admitted explicitly in the encounter history data). Nevertheless, the underlying movement of individuals is fundamental to the encounter process, and thus, we motivate construction of the $\boldsymbol{\pi}_i$ probabilities in equation 1 using that model.

RY assumes that individuals move around their activity centre $\boldsymbol{s}_i$ according to a bivariate normal distribution. That is, locations on consecutive days ($\mathbf{x}_{it}$) are independent draws from a normal distribution having mean $\boldsymbol{s}_i$ and standard deviation σ:

$$\mathbf{x}_{it} \sim \text{Normal}(\mathbf{s}_i, \sigma^2)$$

If we conceptualize a 'trap' as having a physical area over which individuals are exposed, say χ, then the probability that individual i is exposed to capture during a survey is

$$\phi_i = \int_{x \in \chi} k(\mathbf{x}; \mathbf{s}_i, \sigma^2) d\mathbf{x} \qquad \text{eqn 2}$$

where $k()$ is the bivariate normal pdf.

This movement model provides a concise description of exposure to capture but does not provide a characterization of whether an individual is captured or not. Individual probability of capture p_i is naturally expressed conditional on exposure:

$$\begin{aligned} p_i &= \Pr(\text{capture individual } i) \\ &= \Pr(\text{capture}|\text{exposed})\ \Pr(\text{exposed}). \end{aligned}$$

This provides a partitioning of the capture process into information about detectability, Pr(capture|exposed), and spatial information, Pr(exposed), resulting from the juxtaposition of individuals with the trap array. RY assumed uniform capture intensity, so that

$$p_i = r\phi_i$$

Here there is individual heterogeneity in the probability of exposure due to the location of individuals relative to the trap array. RY provided a Bayesian analysis of this model by MCMC, and thus, the integral equation 1 is not computed explicitly. Rather, it is evaluated by Monte Carlo integration within the general framework for inference provided by a fully Bayesian analysis (described below).

The conceptual adoption of the movement/exposure model to trapping arrays leads to a specific formulation of the cell probabilities in equation 1. Let χ_j denote the region around trap j within which an animal may be exposed. The probability that individual i is exposed to trap j is the integral:

$$\phi_{ij} = \int_{\mathbf{x} \in \chi_j} k(\mathbf{x}; \mathbf{s}_i, \sigma) d\mathbf{x}.$$

To a reasonable approximation for small χ_j,

$$\phi_{ij} = \alpha \times k(\mathbf{x}_j; \mathbf{s}_i, \sigma),$$

where $\alpha = \text{area}(\chi_j)$. If exposure to traps is independent and mutually exclusive, then the total exposure is:

$$\phi_i = \sum_j \phi_{ij} = \alpha \sum_j k_{ij} = \alpha E_i \qquad \text{eqn 3}$$

where $k_{ij} = k(\mathbf{x}_j; \mathbf{s}_i, \sigma)$. We refer to $E_i = \sum_{j=1}^{R} k(\mathbf{x}_j; \mathbf{s}_i, \sigma)$ as the total exposure of individual i to the trap array, since it relates directly to ϕ_i. As such, we see that the total exposure of an individual to trapping, resulting from its movements, can be described as an additive function of trap-specific contributions involving the movement kernel $k()$.

The probability of capture for individual i is therefore

$$p_i = r\alpha E_i. \qquad \text{eqn 4}$$

This form suggests that the two constants r and α are not uniquely identifiable. However, we can uniquely estimate their product. Note that while $0 \le r \le 1$, the product may be greater than 1. Thus, to retain the interpretation of this parameter as a probability, we scale equation 4 appropriately. In particular, since $\phi_i \le 1$

$$\alpha \le \frac{1}{\max_s \sum_j k(\mathbf{x}_j; \mathbf{s}, \sigma)}.$$

Define $E_{max} = \max_s \sum_{j=1}^{R} k(\mathbf{x}_j; \mathbf{s}, \sigma)$. Therefore, multiply and divide equation 4 by E_{max}, and define

$$p_i = p_0 \frac{E_i}{E_{max}}. \qquad \text{eqn 5}$$

Then, $p_0 \in [0, 1]$, and p_0 has the interpretation as the probability of capture for the hypothetical 'most exposed' individual.

To complete the model specification, it remains to describe the cell probabilities $\boldsymbol{\pi}_i$ in equation 1. Equations 3 and 5 imply that

$$\pi_{ij} = p_0 \frac{k_{ij}}{E_{max}} \text{ for } j = 1, 2, \ldots, R$$
$$\pi_{i,R+1} = 1 - p_0 \frac{\sum_j k_{ij}}{E_{max}}. \qquad \text{eqn 6}$$

The movement-induced exposure to traps that motivated equation 6 describes the probability of capture in traps as a decreasing function of the distance between activity centre $\boldsymbol{s}$ and the trap locations. This basic phenomenon (detection decreasing with distance) is shared with the formulation of the model by Efford (2004). The main distinction is conceptual – here we have motivated the form as arising under explicit biological processes (movement-based exposure to sampling), whereas the construction by Efford (2004) was more phenomenological in origin. As a practical matter, this distinction may not be very important.

Effective sample area

An estimand that is a by-product of the model specification (Royle & Young 2008) is the effective sample area, say A_e. An intuitive way to describe effective sample area is to think of a discrete version of S, consisting of G pixels $s_1, \ldots, s_G$. Imagine that we were to sample S by flipping a coin for every pixel and then sampling that pixel if the coin comes up 'heads.' In that case, we will survey that pixel, thereby exposing individuals on that pixel to sampling. If the coin has probability π_i of coming up heads for pixel i then we would expect to sample an area of size

$$A_e = \sum_{i=1}^{G} \pi_i.$$

In the context of our spatial capture–recapture model, individuals are flipping the coins, with probabilities

$$\pi_i = 1 - (1 - \phi_i)^T,$$

which we can only evaluate under certain conditions (since we have absorbed the constant α into p_0 in equation 4). It might be reasonable to assert that the most exposed individual has $\phi_i = 1$ (i.e., it is always susceptible to capture). This is setting (from equation 3) $\alpha = 1/E_{max}$. Whether or not this is reasonable depends on the typical movements of individuals relative to the extent of the trap array. The assumption that $\phi = 1$ when s coincides with a trap location seems reasonable in some cases.

Example

A simulated realization under the model is shown in Fig. 1 for $N = 120$ individuals, subjected to capture $T = 6$ times by a 10×10 array of traps having unit spacing, using the normal

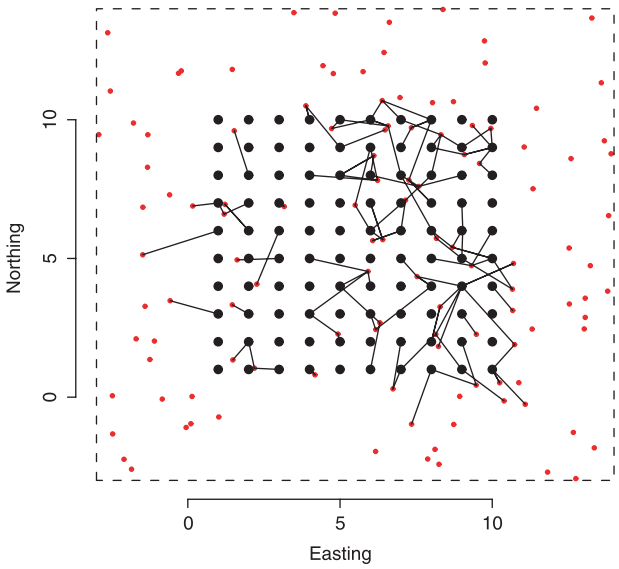


Fig. 1. Simulated realization of observations of $N = 120$ individuals (red dots) exposed to a grid of 10×10 traps (black dots). The trap(s) in which each individual was captured are indicated with black lines.

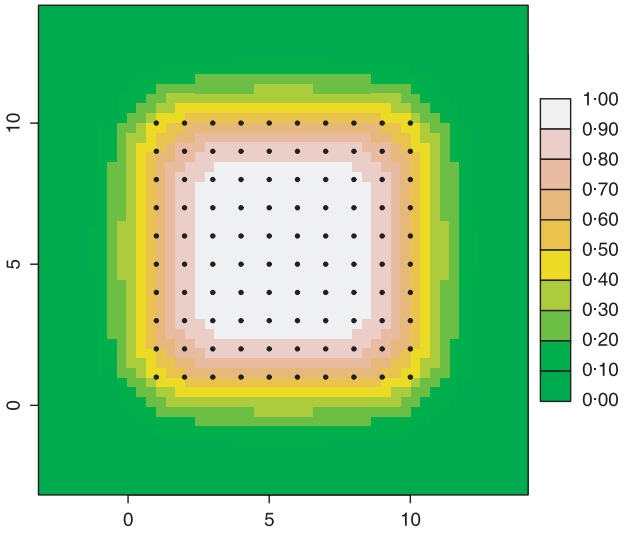


Fig. 2. Probability of exposure of individuals to a 10×10 array of traps under a normal exposure function with $\sigma^2 = 1$. The shading represents the exposure of a hypothetical individual having its activity centre at each point.

exposure kernel with $\sigma^2 = 1$ and with $p_0 = 0{\cdot}50$. The individual activity centres were uniformly distributed over a larger square of 17×17 units. Individuals are connected to traps where they were captured by black lines. The probability of exposure, ϕ, of an individual as a function of $\boldsymbol{s}$ is shown in Fig. 2. We see that individuals within the array have very high ϕ, which decreases rapidly as distance from the outer band of traps increases.

The summary statistics for p_i for the 120 individuals were (0·003, 0·063, 0·181, 0·3981) for first quartile, median, mean, and third quartile, respectively. Evidently, there is extreme heterogeneity in detection probabilities under this model, with one-half of the population having $p_i < 0{\cdot}063$ while the mean is 0·181. For the simulated $T = 6$ study, 55 unique individuals were captured a total of 118 times, the detection frequencies being 16, 21, 12, 6 individuals captured 1, 2, 3, 4 times, respectively (none captured 5 or 6 times). An analysis of this simulated data set can be found in Supporting Material Appendix S1.

Conditional formulation

The model is a spatial version of the standard null model referred to as Model M_0. A reformulation of the model is convenient for extending the model to, for example, allow for modelling covariates that influence detection probability, p_0. In particular, we express the observation model as the product of two components: (i) the probability that an individual is captured ('at all'), and (ii) the probability of capture in each trap conditional on capture. That is, given that an individual is captured, what are the trap-specific probabilities of capture? Let y_{it} be a binary indicator of capture for individual i. Thus, $(y_{i1}, y_{i2}, \ldots, y_{iT})$ is a traditional encounter history. Then,

$$y_{it} \sim \text{Bern}(p_i)$$

where p_i is the individual probability of capture, equation 4. Then, let γ_{ij} be the probability of capture in trap j given that $y_{it} = 1$. From equation 6, we obtain

$$\gamma_{ij} = \frac{\pi_{ij}}{\sum_{j=1}^{R} \pi_{ij}} = \frac{k_{ij}}{\sum_j k_{ij}}.$$

where k_{ij} is the bivariate normal kernel.

Thus, conditional on capture, the trap-of-capture is a categorical random variable having probabilities γ_{ij} for $j = 1, 2, \ldots, R$. This is convenient for modelling detection covariates because it separates (in the model) the parameter p_0 from the spatial information contained in the trap identities. Thus, the model for the traditional encounter histories is:

$$y_{it} \sim \text{Bern}(p_{it})$$

where

$$p_{it} = p_{0t} \frac{\sum_{j=1}^{R} k(\mathbf{x}_j; \mathbf{s}_i, \sigma)}{E_{max}},$$

It is possible to model temporal effects in p_{0t} such as time effects, behavioural effects, etc.

Bayesian analysis by data augmentation

We note that this model is conceptually similar to a broad class of capture–recapture models known as individual covariate models (Williams *et al.* 2002; p. 301). In the present situation, the activity centres s_i represent an individual covariate that is observed with error. Indeed, the model is most structurally similar to distance sampling with distance measurement error (Royle & Dorazio 2008, Ch. 7). Bayesian analysis of classical individual covariate models that motivates the following analysis can be found in Royle (2008). We adopt that basic strategy here.

To conduct a Bayesian analysis of the model requires that we describe prior distributions for the parameters N, p_0, and σ. A natural choice of priors that reflect the absence of information about these parameters is to assume a discrete uniform prior for N on the integers $0, \ldots, M$ for some large value of M [i.e., $N \sim \text{Du}(0, M)$]. A customary vague prior for p_0, which, recall has support on [0, 1], is the uniform (0, 1) prior. Choice of M is not a critical consideration except that it must be chosen large enough so as to not truncate the posterior distribution of N which can be checked after a trial analysis. For σ, we use a proper uniform prior distribution with a large upper-bound.

In principle, the model could be analysed under this prior specification by conventional MCMC methods for sampling from the posterior distribution (Link *et al.* 2002). However, we note that, because N is unknown, the dimension of the parameter space (i.e., the number of random effects $\boldsymbol{s}_i$) is also unknown. As such, each time that a new draw of N is made

from the posterior distribution, the number of random effects (activity centres) changes. Properly updating parameters in this setting has proved to be a challenging problem of some technical complexity. This technical problem motivated an approach to the analysis of such models using the method of data augmentation in Royle *et al.* (2007). We provide a brief description of data augmentation here. See Royle *et al.* (2007) and Royle & Dorazio (2008) for technical details and applications.

A heuristic description of data augmentation is that it arises by simply adding excess 'all zero' encounter histories to the data set. That is, for M sufficiently large, then we can augment the data set with $M - n$ all-zero encounter histories. We then recognize that the resulting model for the augmented data is a zero-inflated version of the model for the complete data set (i.e. as if N were known). In models with individual effects, data augmentation is a convenient framework because it allows us to retain a maximal set of random effects in the (augmented) data set, and their values are updated at each iteration of the MCMC algorithm.

Formally, data augmentation is justified as a re-parameterization of the model that arises under the discrete uniform prior for N. In particular, note that the discrete uniform prior for N can be constructed by specifying a binomial prior for N: $N \sim \text{Bin}(M, \psi)$, and then placing a uniform(0, 1) prior on ψ. When ψ is removed from the binomial component by integration, the result is $N \sim Du(0, M)$. While this may seem only a mathematical curiosity, it suggests a convenient implementation of Bayesian analysis for such models. Namely, we can think of the discrete uniform prior as suggesting a superpopulation consisting of M individuals, where $M - n$ of them have corresponding 'all-zero' encounter histories (because they were not captured). We recognize that some of the M individuals are fixed zeros, whereas some of them are sampling zeros – they correspond to individuals in the population that were not captured. This can be formalized by the introduction of a set of latent indicator variables $z_1, z_2, \ldots, z_M$ such that $z_i = 1$ if individual i is a member of the population and $z_i = 0$ if individual i is a fixed zero. We assume $z_i \sim \text{Bernoulli}(\psi)$. To implement data augmentation, we augment the n observed encounter histories with $M - n$ 'all-zero' histories, and then specify the model for the augmented data set in terms of the zero-inflated version of the 'known-N,' model. For most capture–recapture models (including the present), the model for the total number of detections, y_i (see Conditional formulation), is

$$y_i \sim \text{Bin}(p_i) \quad \text{if } z_i = 1,$$

$$y_i = 0 \quad \text{if } z_i = 0.$$

Under data augmentation, the parameter ψ formally replaces the parameter N, the two being related by the prior specification $N \sim \text{Bin}(M, \psi)$.

Because the dimension of the parameter space is fixed, this facilitates a formal analysis by standard methods of MCMC. While developing the MCMC algorithm for analysis of the augmented data is straightforward under this model, we avoid those technical details because the model can also be implemented in the freely available software package WinBUGS (Gilks *et al.* 1994) as demonstrated in Supporting Material Appendix S1 (see also Gardner, Royle & Wegan 2008).

Analysis of the tiger camera-trapping data

The tiger population of Nagarahole reserve in the state of Karnataka, southwestern India, has been studied via camera-trap methods from 1991 until the present (e.g. Karanth 1995; Karanth & Nichols 1998; Karanth *et al.* 2006). Here we present analyses of data collected during 24 January–16 March 2006 from 120 trap stations (Fig. 3). Two camera traps (unambiguous identification requires photographs of both flanks) were placed at each station. Not all 120 trap stations were operated simultaneously. Instead, the reserve was subdivided into four blocks of approximately 30 trapping stations each, and each block was run for 12 consecutive days. Then, cameras were moved to the next block for another 12 days and the process repeated until all four blocks were sampled. This design follows sample design 4 of Karanth and Nichols (2002, p. 133).

This study resulted in encounter history data on 45 individuals. Each encounter history was a vector of length $T = 48$, with a 0 recorded for each occasion where the animal was not detected and the location (station number) recorded for each capture. For instances where an individual was captured > 1 time in the same survey period, the additional captures were regarded as distinct recaptures. While our motivation was purely practical for doing this, we believe it is statistically efficient when the capture rate is low.

We refer to the 120 trap locations across the four blocks as 'traps', recognizing that the physical cameras were absent from most of the locations at any particular time. As a result of this trap movement, individuals could not be detected at approximately three-quarters of the traps during any particular interval. This is simple to handle in the model by defining a binary indicator of trap operation during each day, say o_{tj} for survey t and trap j, and then redefining exposure accordingly. For example, equation 4 becomes

$$p_{it} = p_0 \frac{\sum_{j=1}^{R} o_{tj} k(\mathbf{x}_j; \mathbf{s}_i, \sigma)}{E_{max}}.$$

Simply put, this accumulates exposure to trapping based only on traps that are in operation at time t.

For the analysis of these data, we excluded areas that were judged to be non-habitat within a 15-km buffer area containing the trap array (Fig. 3). This renders the definition of a uniform prior for the activity centres difficult. As such, we described this region of suitable habitat by a fine grid of 9961 equally spaced points, each representing approximately 0·336 km^2 over the buffered region. Of these, 4898 (1645·7 km^2) were judged to represent suitable habitat. The activity centres $\boldsymbol{s}_i$ were assumed to be uniformly distributed over these 4898 points. We developed an implementation of the model for this

© 2008 The Authors. Journal compilation © 2008 British Ecological Society,

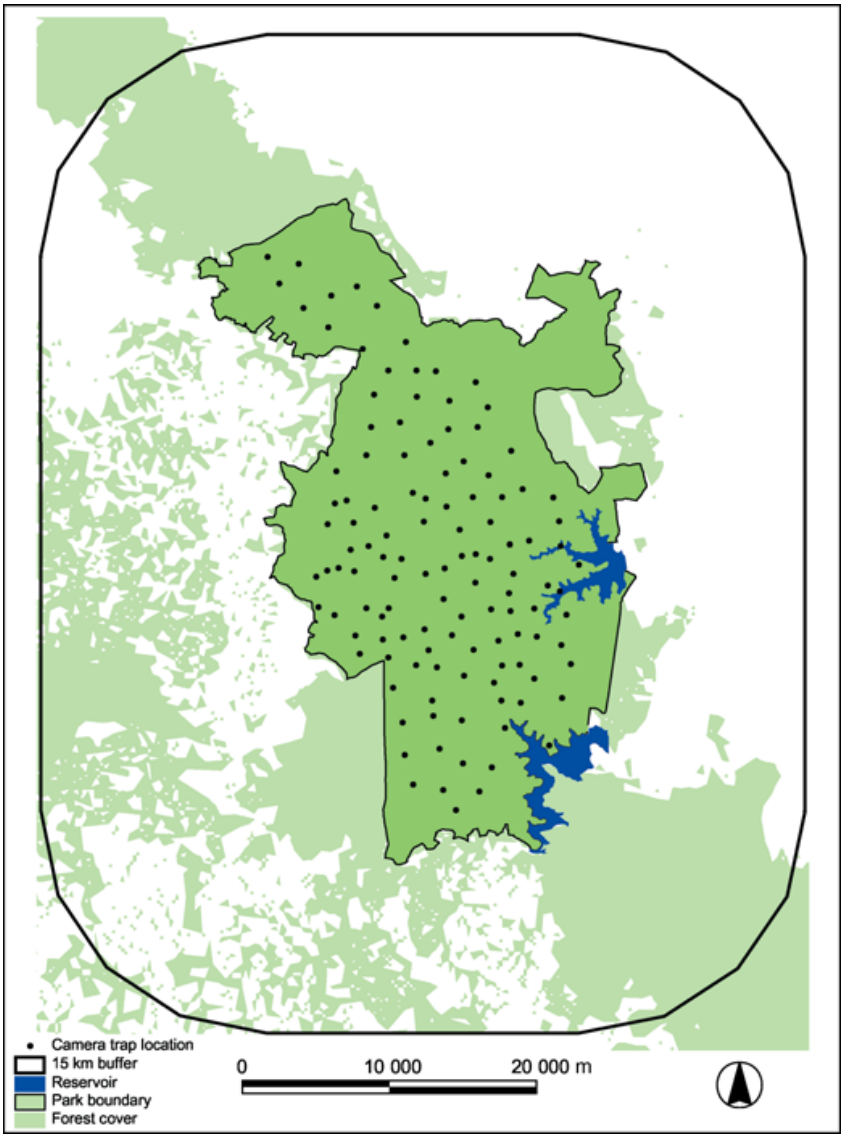


Fig. 3. Nagarahole tiger camera-trapping study area with 15-km buffer and non-habitat clipped out.

discrete-*S* situation in the R software (Ihaka & Gentleman 1996). Bayesian analysis of the model was conducted using data augmentation, augmenting the data set with 800 'all zero' encounter histories. The MCMC algorithm was run for 52 000 iterations, the first 2000 were discarded, and posterior summaries were computed from the remaining 50 000 iterations. The coordinate system was scaled so that a standard unit was 5 km, and thus also are the units of σ. Posterior summaries are provided in Table 1.

The posterior median of N, the population size for the 1645·7 km^2 area, is 229. This is the number of individuals existing within the region over which $\boldsymbol{s}$ was assumed uniformly distributed (i.e., suitable habitat). We see that the posterior mean of ψ (the data augmentation parameter) is approximately 0·28 and the posterior mass is concentrated well away from the boundary $\psi = 1$. This indicates that the posterior of N was not truncated as a result of too few all-zero encounter histories. In Table 1, D is the density per 100 km^2 of suitable habitat, and the posterior mean is 14·3 with a 95% posterior interval of (9·6, 20·5). Based on earlier work on tiger ecology and knowledge of the relationship of tiger density to prey abundance (Karanth 1995; Karanth *et al.* 2004b), as well as from estimates of tiger densities at this site over multiple years (Karanth *et al.*

Table 1 Posterior summaries of model parameters for the tiger camera-trapping data based on $n = 45$ observed individuals. Here, A_e is the area exposed to trapping or the effective sample area of the 120-trap array. It is a derived parameter under the model. N is the number of tiger activity centres in the population exposed to sampling and D is the density per 100 km^2, ψ is the data augmentation parameters, p_0 is detection probability, and σ is the parameter in the bivariate normal pdf

Parameter	Mean	SD	2·5%	Median	97·5%
ψ	0·278	0·056	0·182	0·272	0·404
σ	0·407	0·050	0·325	0·402	0·517
p_0	0·083	0·018	0·052	0·081	0·123
N	233·93	45·20	158·00	229·00	337·00
D	14·300	2·800	9·600	14·000	20·500
A_e	725·62	34·53	666·88	722·91	800·36

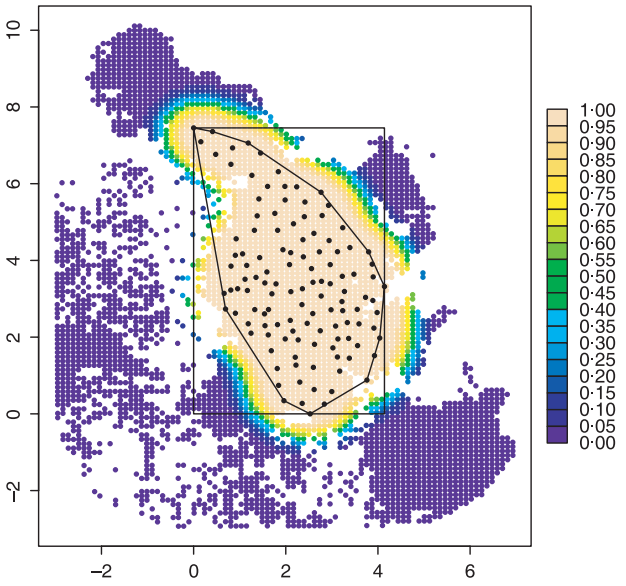


Fig. 4. Probability of exposure to trapping in a $T = 48$ occasion camera-trapping study. The colour of each pixel is the Pr(exposure to sampling) of a tiger with activity centre at that pixel. Only pixels judged to be suitable habitat are included. The sum of all pixels is the effective sample area of the trapping array.

2006), we believe the density estimate generated in this study is reasonable. Finally, we note that the estimate $p_0 = 0{\cdot}083$ (posterior mean) corresponds to the detectability of some hypothetical individual having maximal exposure to the trap array.

Putting the abundance estimate in context, fitting a classical closed-population model with heterogeneous detection probabilities (so-called 'model M_h'), we obtain $\hat{N} = 111.7$. It is impossible to compare this directly with the posterior mean of N produced under our model because the latter corresponds to a precise area over which the activity centres $\boldsymbol{s}$ were assumed to be uniformly distributed. Conversely, the estimate of N obtained under M_h is, at best, an estimate of the number of individuals exposed to sampling, for which we do not know the precise area. However, our model produces an estimate of the effective sample area (the posterior mean is *c*. 725 km^2, see Table 1). The probability of exposure as a function of location is shown in Fig. 4. We might hope that the estimate of N under model M_h is relevant to this area. If so, the comparable estimate of density would be approximately 15·4 individuals per 100 km^2.

Discussion

Capture–recapture studies of many species, based on camera traps, DNA sampling, and traditional traps that physically capture individuals, produce individual encounter history data with auxiliary spatial information related to location of capture. It has long been recognized that this information could be used to produce unbiased estimates of density when the population is not geographically closed due to movement of individuals (Dice 1938; Smith *et al.* 1971; Wilson & Anderson 1985; Efford 2004).

For camera-trap studies, the standard method for density estimation, as exemplified by Karanth & Nichols (1998, 2000, 2002), Trolle & Kéry (2003), and other recent studies, has been based on the use of conventional closed-population estimators with a buffer strip estimated from observed movements of individuals. In this study, we developed a hierarchical model for spatial capture–recapture data which integrates the detection information provided by the encounter history data as well as the auxiliary spatial information contained in the trap locations of capture. The model describes both the distribution of individuals in space (i.e. their activity centres), which are latent parameters regarded as random effects in the model, as well as the encounters of individuals that result from their movement-induced exposure to trapping (Royle & Young 2008). We adopt a Bayesian analysis of the hierarchical model, facilitated by the use of data augmentation (Royle *et al.* 2007).

Observations from camera-trap studies have a number of distinctive characteristics compared to conventional 'traps' that physically capture or kill individuals. Importantly, camera traps can detect multiple individuals during a single survey occasion (a night in our example). Our model accommodates this by the assumption that capture of individuals are independent multinomial trials (equation 1). Secondly, a camera trap can, conceivably, detect the same individual multiple times during a sample occasion. However, current data processing protocols (in our study) render the data into unique binary detection events. This is probably desirable when the sampling interval is short (daily) since within-day movements are likely to be highly correlated. Finally, an individual can be encountered in multiple traps during a sampling occasion. Historically, multiple captures have been used to estimate the distance moved for calculation of the buffer around the trap array, but not in the definition of the encounter histories. Our model does not properly admit such multiple captures. However, in our analysis of the tiger data, we regarded such multiple captures as ordinary recaptures (*iid* multinomial trials). This is more efficient than discarding such data and, we believe, at low capture rates, there is little loss of efficiency treating the data in this manner.

Despite this shortcoming, our formulation offers several practical advantages to studies that involve non-invasive animal 'captures' from photographs, DNA or other types of individual identification. First, our model explicitly includes a biological explanation for detection heterogeneity in terms of proximity of individuals to the trap array. Thus, it possibly addresses some concerns (Link 2003) regarding identifiability of abundance using standard models. More importantly, because the point process is parameterized explicitly in the hierarchical model, we believe that this component of the model can, in the future, be generalized considerably and in substantive, meaningful directions. For example, in carnivore populations, it is likely that the 'activity centres' are not independent and uniformly distributed over S. Individuals probably interact with neighbours and their density varies in response to habitat conditions. Both of these features can be integrated into the framework which we have described. In

particular, in the discrete-*S* formulation of the model used in the analysis of the tiger data, one may associate covariates with each pixel, and model density as a function of those covariates. To model interactions among individuals, it seems natural to formulate the point process in terms of Markovian dependence, wherein successive points are conditioned on the nearness and even 'type' (male, female; social status; age) of neighbours. Finally, the possibility exists of formulating point process models that admit explicit spatio-temporal structure, representing the survival and recruitment of individuals. As such, we can make efficient use of available data to potentially yield great improvements in the precision of density estimates. We are pursuing a number of extensions of the model in the context of camera-trapping studies.

We noted that our model is structurally similar to the model described by Efford (2004) for which he devised a simulation-based method of fitting based on inverse prediction. We regard our development here as a formalization of the inference framework, using a Bayesian hierarchical model. We note that formalization of the inference framework for such models was also recently addressed by Borchers & Efford (2008), who developed an integrated likelihood framework for inference. Our formulation is distinct with regard to formulation of the model, the manner in which inference is carried out, and implementation of the model. With regard to formulation, our development was motivated by exposure of individuals induced by movement about their activity centres. While it is more heuristic in this case (in contrast to Royle & Young 2008), since the processes are not separately identifiable, the explicit movement model provides a conceptual linkage between detection probability and 'temporary emigration' (Kendall *et al*. 1997) which we find appealing. One technical distinction with the Borchers & Efford (2008) formulation is that we condition on the population size *N*, and then analyse what is often referred to (Borchers *et al*. 2002) as the 'joint likelihood', using data augmentation. Conversely, Borchers & Efford (2008) adopt the conditional likelihood approach under which the likelihood is formulated conditional on *n* (Sanathanan 1972). We believe that there are important advantages of the explicit conditioning on *N*, and the hierarchical formulation of the model that this allows. In particular, as we demonstrated (Supporting Material Appendix S1), analysis of the Bayesian hierarchical formulation based on data augmentation can be achieved in the flexible and freely available software WinBUGS requiring little or no specialized knowledge. As such, implementation and extension are widely accessible to practitioners. More importantly, as analysis of the model is conditional on the individual point locations, this component of the model can be the focus of additional model structure describing the distribution of point locations, or relationships among point locations (e.g. as described in the previous paragraph). We believe that it would be generally difficult to carry out the integration necessary to estimate parameters for more complex point-process models based on integrated likelihood as developed by Borchers & Efford (2008). Indeed, parametric inference under such models is largely simulation-based at the present time for this reason (e.g. see Møller & Waagepetersen 2004). We expect such extensions to be the focus of considerable research in the future, especially as passive methods of DNA sampling and technological advances render spatial capture–recapture data easier and cheaper to obtain.

Acknowledgments

We are grateful to the Wildlife Conservation Society, New York for supporting this study. We are indebted to N. Samba Kumar for contributing substantially to the field work. We thank two anonymous referees for helpful reviews of a draft of this manuscript and Murray Efford for his extensive and forthright commentary.

References

Borchers, D.L. & Efford, M.G. (2008) Spatially explicit maximum likelihood methods for capture–recapture studies. *Biometrics* **64**, 377–385.

Borchers, D.L., Buckland, S.T. & Zucchini, W. (2002) *Estimating Animal Abundance: Closed Populations*. Springer-Verlag, London.

Dice, L. (1938) Some census methods for mammals. *Journal of Wildlife Management*, **2**, 119–130.

Dice, L. (1941) Methods for estimating populations of mammals. *Journal of Wildlife Management*, **5**, 398–407.

Efford, M. (2004) Density estimation in live-trapping studies. *Oikos*, **106**, 598–610.

Gardner, B., Royle, J.A. & Wegan, M.T. (2008) Hierarchical models for estimating density from DNA mark–recapture studies. *Ecology* (in press).

Gilks, W.R., Thomas, A. & Spiegelhalter, D.J. (1994) A language and program for complex Bayesian modelling. *The Statistician*, **43**, 169–178.

Goswami, V.R., Madhusudan, M.D. & Karanth, K.U. (2007) Application of photographic capture–recapture modelling to estimate demographic parameters for male Asian elephants. *Animal Conservation*, **10**, 391–399.

Ihaka, R. & Gentleman, R. (1996) R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, **5**, 299–314.

Karanth, K.U., Nichols, J.D., Kumar, N.S., Link, W. & Hines, J. (2004b) Tigers and their prey: predicting carnivore densities from prey abundance. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 4854–4858.

Karanth, K.U. & Nichols, J.D. (1998) Estimation of tiger densities in India using photographic captures and recaptures. *Ecology*, **79**, 2852–2862.

Karanth, K.U. & Nichols, J.D. (2000) Ecological status and conservation of tigers in India. WCS, US Fish and Wildlife Service. Centre for Wildlife Studies, Bangalore, India, 123.

Karanth, K.U. & Nichols, J.D. (2002) Monitoring tigers and their prey: a manual for researchers, managers and conservationists in tropical Asia. Centre for Wildlife Studies, Bangalore, India.

Karanth, K.U. (1995) Estimating tiger (*Panthera tigris*) populations from camera-trap data using capture–recapture models. *Biological Conservation*, **71**, 333–338.

Karanth, K.U., Chundawat, R., Nichols, J.D. & Kumar, N.S. (2004a) Estimation of tiger densities in the tropical dry forests of Panna, Central India, using photographic capture–recapture sampling. *Animal Conservation*, **7**, 285–290.

Karanth, K.U., Nichols, J.D., Kumar, N.S. & Hines, J. (2006) Assessing tiger population dynamics using photographic capture–recapture sampling. *Ecology*, **87**, 2925–2937.

Kendall, W., Nichols, J.D. & Hines, J. (1997) Estimating temporary emigration using capture–recapture data with Pollock's robust design. *Ecology*, **78**, 563–578.

Link, W.A. (2003) Nonidentifiability of population size from capture–recapture data with heterogeneous capture probabilities. *Biometrics*, **59**, 1123–1130.

Link, W.A., Cam, E., Nichols, J.D. & Cooch, E.G. (2002) Of BUGS and birds: Markov chain Monte Carlo for hierarchical modelling in wildlife research. *Journal of Wildlife Management*, **66**, 277–291.

Møller, J. & Waagepetersen, R.P. (2004) *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall/CRC, New York.

O'Brien, T., Kinnaird, M. & Wibisono, H. (2003) Crouching tigers, hidden prey: Sumatran tiger and prey populations in a tropical forest landscape. *Animal Conservation*, **6**, 131–139.

Otis, D.L., Burnham, K.P., White, G.C. & Anderson, D.R. (1978) Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, **62**, 1–135.

Royle, J.A. & Dorazio, R.M. (2008) *Hierarchical Models and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities*. Academic Press, 444 pp. Bullington, MA, USA.
Royle, J.A. & Young, K.V. (2008) A hierarchical model for spatial capture–recapture data. *Ecology* **89**, 2281–2289.
Royle, J.A. (2008) Analysis of capture–recapture models with individual covariates using data augmentation. *Biometrics* (in press).
Royle, J.A., Dorazio, R.M. & Link, W.A. (2007) Analysis of multinomial models with unknown index using data augmentation. *Journal of Computational and Graphical Statistics*, **16**, 67–85.
Sanathanan, L. (1972) Estimating the size of a multinomial population. *Annals of Mathematical Statistics*, **43**, 142–152.
Seber, G.A.F. (1982) *The Estimation of Animal Abundance and Related Parameters*, 2nd edn. Charles Griffin, London.
Silver, S., Ostro, L., Marsh, L., Maffei, L., Noss, A., Kelly, M., Wallace, R., Gomez, H. & Ayala, G. (2004) The use of camera traps for estimating jaguar *Panthera onca* abundance and density using capture/recapture analysis. *Oryx*, **38**, 148–154.
Smith, M., Blessing, R., Chelton, J., Gentry, J., Golley, F. & McGinnis, J. (1971) Determining density for small mammal populations using a grid and assessment lines. *Acta Theriologica*, **16**, 105–125.
Soisalo, M. & Cavalcanti, S. (2006) Estimating the density of a jaguar population in the Brazilian Pantanal using camera-traps and capture–recapture sampling in combination with GPS radio-telemetry. *Biological Conservation*, **129**, 487–496.
Stickel, L. (1954) A comparison of certain methods of measuring ranges of small mammals. *Journal of Mammalogy*, **35**, 1–15.
Trolle, M. & Kéry, M. (2003) Estimation of ocelot density in the Pantanal using capture–recapture analysis of camera-trapping data. *Journal of Mammalogy*, **84**, 607–614.
Waits, L.P. (2004) Using non-invasive genetic sampling to detect and estimate abundance of rare wildlife species. *Sampling Rare or Elusive Species: Concepts, Designs, and Techniques for Estimating Population Parameters* (eds W.L. Thompson), pp. 211–228. Island Press, Washington/Covelo/London.
Wallace, R., Gomez, H., Ayala, G. & Espinoza, F. (2003) Camera trapping for jaguar (*Panthera onca*) in the Tuichi Valley, Bolivia. *Journal of Neotropical Mammalogy*, **10**, 133–139.
Wegge, P., Pokheral, C.Pd. & Jnawali, S.R. (2004) Effects of trapping effort and trap shyness on estimates of tiger abundance from camera trap studies. *Animal Conservation* **7**, 251–256.
White, G.C., Anderson, D.R., Burnham, K.P. & Otis, D.L. (1982) *Capture–Recapture and Removal Methods for Sampling Closed Populations*. Los Alamos National Laboratory Report LA-8787-NERP, Los Alamos, NM, USA.
Williams, B.K., Nichols, J.D. & Conroy, M.J. (2002) *Analysis and Management of Animal Populations*. Academic Press, San Diego, CA. USA.
Wilson, K.R. & Anderson, D.R. (1985) Evaluation of two density estimators of small mammal population size. *Journal of Mammalogy* **66**, 13–21.

Received 23 June 2008; accepted 24 September 2008
Handling Editor: Paul Lukacs

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Analysis of simulated data and WinBUGS model specification.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Appendix D

Bayesian inference in camera trapping studies for a class of spatial capture-recapture models

Ecology, 90(11), 2009, pp. 3233–3244
© 2009 by the Ecological Society of America

Bayesian inference in camera trapping studies for a class of spatial capture–recapture models

J. Andrew Royle,[1,3] K. Ullas Karanth,[2] Arjun M. Gopalaswamy,[2] and N. Samba Kumar[2]

[1]*U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel, Maryland 20708 USA*
[2]*Wildlife Conservation Society-India Program, Centre for Wildlife Studies, Bangalore, Karnataka-560042, India*

Abstract. We develop a class of models for inference about abundance or density using spatial capture–recapture data from studies based on camera trapping and related methods. The model is a hierarchical model composed of two components: a point process model describing the distribution of individuals in space (or their home range centers) and a model describing the observation of individuals in traps. We suppose that trap- and individual-specific capture probabilities are a function of distance between individual home range centers and trap locations. We show that the models can be regarded as generalized linear mixed models, where the individual home range centers are random effects. We adopt a Bayesian framework for inference under these models using a formulation based on data augmentation. We apply the models to camera trapping data on tigers from the Nagarahole Reserve, India, collected over 48 nights in 2006. For this study, 120 camera locations were used, but cameras were only operational at 30 locations during any given sample occasion. Movement of traps is common in many camera-trapping studies and represents an important feature of the observation model that we address explicitly in our application.

Key words: abundance; Bayesian analysis; binomial point process; camera trapping; carnivore surveys; data augmentation; density estimation; hierarchical model; Markov chain Monte Carlo; spatial capture–recapture; tigers; trapping array; trapping grid.

Introduction

Much of the theory and methodology underlying inference about population size from closed capture–recapture models is concerned with animal populations that are well defined in the sense that one can randomly sample individuals that are associated with some location or area and, usually, uniquely identify them. However, for many populations, the spatial area over which individuals occur (and are exposed to capture) cannot be precisely delineated. Movement of individuals onto and off of a putative sample unit induces a violation of the key assumption of "geographic closure," which impacts our ability to interpret the estimates of population size, N, derived from closed population capture–recapture models.

One area of application where this problem is profoundly important is the use of "camera traps" to study secretive animal populations. In particular, the use of arrays of camera traps for estimating abundances of large felids is widespread. They have been used in studies of tigers (Karanth 1995, Karanth and Nichols 1998, Karanth et al. 2004), ocelots (Trolle and Kéry 2003, 2005), jaguars (Wallace et al. 2003, Maffei et al. 2004), and other species which are individually identifiable from their spot or stripe patterns. In the typical situation, an array of cameras is distributed over the landscape. Over time, the cameras provide encounter history data indicating the occasions of captures of individuals, as well as auxiliary data on spatial location of captures.

The conventional approach to the analysis of density from these systems is to apply closed population models, and then attempt to convert those estimates to densities using a wide range of heuristically motivated but essentially ad hoc methods. For example, ecologists have used auxiliary location information to estimate the mean or maximum distance moved to adjust the effective sample area, or used various other heuristic "adjustments." (Wilson and Anderson 1985*a*, *b*, Karanth and Nichols 1998, Parmenter et al. 2003, Trolle and Kéry 2003). The standard approach (Karanth and Nichols 1998) places a buffer strip around the trap array (or a convex hull containing the array) that is equal to half the mean maximum distance moved by individuals captured in more than one trap. While these procedures seem to work adequately in practice, the model or range of conditions for which they might work is poorly understood and difficult to characterize theoretically, and there is no basis for their extension. Formalization of the use of auxiliary spatial information requires the precise definition of a model: the linkage of encounter location to some notion of territory or home range (and perhaps movements, see Royle and Young 2008).

Manuscript received 14 August 2008; revised 27 January 2009; accepted 2 February 2009. Corresponding Editor: K. Newman.

[3] E-mail: aroyle@usgs.gov

In this paper, we describe a general class of hierarchical models for encounter data from studies that use camera traps and related methods. By considering an array of traps in which each trap functions independently of all other traps, individuals in the population may be captured in multiple traps during any capture occasion and even multiple times in the same trap. This is realistic in studies based on camera trapping and other sampling methods which do not physically capture individuals. We describe a hierarchical model for the observed spatial encounter history data which recognizes that individual trap encounter histories are the outcome of two processes: distribution of individuals across the landscape and an encounter process that describes whether or not individuals are encountered by traps as a function of their location. We specify a model for the distribution of individuals in space in terms of a simple binomial point process, where $\mathbf{s}_i$, $i = 1, 2, \ldots, N$, is the realization of such a process for the N individuals in the population. We interpret the point locations as individual activity centers (or home range centers, or centroids; Borchers and Efford 2008, Royle and Young 2008, Gardner et al. 2009, Royle et al. 2009). We consider two distinct observation models that are defined conditional on the underlying point process realization: In the most general model, we allow that an individual can be caught an arbitrary number of times in each of an arbitrary number of traps. In principle, this is a reasonable model for camera trapping and other situations (e.g., hair or dung sampling to obtain DNA data), except that data, for processing reasons, are often reduced to single encounters for each trap. Moreover, when sample intervals are short (e.g., nights), it seems likely that multi-trap encounters should be highly correlated with one another, and so there may be little loss of efficiency in discarding such data. Therefore, we consider a second model in which individuals can be caught at most one time per trap, but in an arbitrary number of traps. This model is a formal restriction of the more general model, which can be obtained by imposing an explicit restriction on the observed random variables.

We demonstrate that these spatial capture–recapture models have simple formulations as classical generalized linear models with random effects (i.e., generalized linear mixed models; GLMMs). Analysis of the models is technically challenging because there are many latent variables (random effects) in the model. In particular, the activity centers of each individual are unknown. Moreover, the number of such activity centers (i.e., the population size N) is also unknown. To attack inference under these models, we adopt a Bayesian analysis of the model based on data augmentation (Royle et al. 2007), which has been applied to a number of related models (e.g., Royle and Dorazio 2008, Gardner et al. 2009, Royle 2009). Using data augmentation, the spatial capture–recapture models are formulated as zero-inflated Poisson or binomial regression models with random effects. The models may be implemented in the freely available software WinBUGS (Gilks et al. 1994), which we believe yields an accessible platform for extension of the models described in our paper.

Camera Trapping: The Nagarahole Data

Photographic captures of felid species that possess individually unique spot or stripe patterns, permit estimation of their abundance and density using closed capture recapture models. Such surveys have been conducted for tigers (Karanth and Nichols 1998, Karanth et al. 2004), leopards (Henschel and Ray 2003), jaguars (Wallace et al. 2003, Maffei et al. 2004, Soisalo and Cavalcanti 2006), and ocelots (Trolle and Kéry 2003), for example. These surveys are typically conducted over a short period of 30–60 days to ensure demographic closure, and involve photo-capturing both flanks of the cat in order to assign a unique identity to the individual. To increase capture probabilities, the traps are placed at ecologically optimal sites chosen based on the expertise of biologists, resulting in an irregular pattern of trap locations across the sampled area.

The tiger population in Nagarahole Reserve in the state of Karnataka, southwestern India, has been studied via camera trap methods by Karanth and associates from 1991 until the present (e.g., Karanth 1995, Karanth and Nichols 1998, Karanth et al. 2006; see Plate 1). The specific data set examined here was obtained in 2006 from sampling at 120 trap stations, each referenced by latitude and longitude and by UTM coordinates (Fig. 1). Two camera traps (unambiguous identification requires photographs of both flanks of each detected animal) were placed at each location. The sampling took place over 48 nightly intervals between 24 January and 16 March 2006.

All photographic captures of tigers obtained were labeled with auxiliary data such as date, time and location of capture, and the two corresponding flank images were linked using these data to establish the identity of each tiger. Multiple independent captures of individuals in the same night and trap is generally rare (fewer than four such events). As a result, in earlier published analyses of these data based on classical (nonspatial) closed population capture–recapture models (e.g., Karanth et al. 2004), multiple captures of the same individual at different trap locations or multiple captures at one location, during the same night, were combined into single binary "capture" events. We consider models that could accommodate capture both in multiple traps and also multiple captures in the same trap. However, we only address the former in our analysis of the data because our data have been processed into binary encounter history data. In addition, as we noted previously, we believe that within-trap recaptures in the same night are liable to be highly correlated (e.g., due to individuals moving past a camera multiple times in quick succession), and this

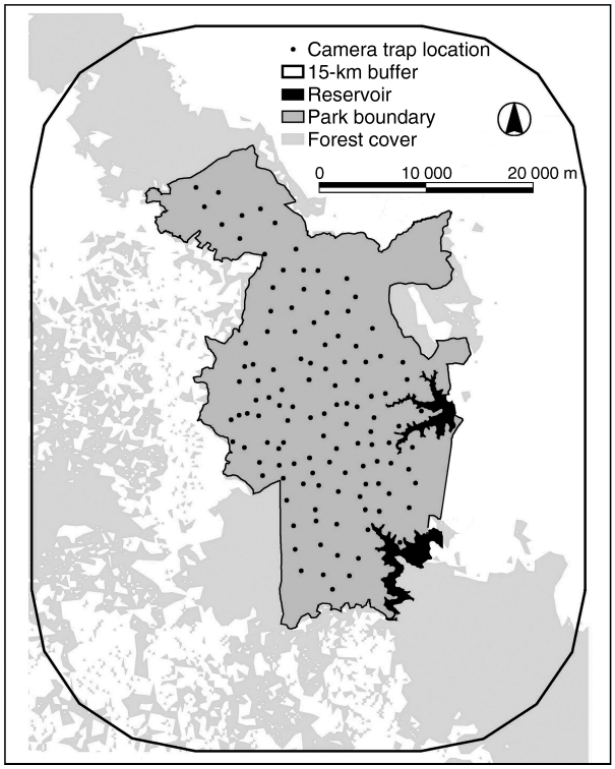


FIG. 1. Nagarahole Reserve tiger camera-trapping study area, Karnataka, southwestern India, with 15-km buffer outlined and non-habitat whited out.

may introduce an additional model selection/sensitivity issue while providing information that is less informative about density than captures among traps (see *Discussion*).

Sampling over this period resulted in encounter history data on 44 individuals during the 48 nightly sampling occasions. Thus, each encounter history was a $J \times K$ (120 traps $\times$ 48 nights) matrix, where a 0 was recorded for each occasion and trap location where the animal was not detected and a 1 was recorded for each occasion and trap location where a detection occurred. The number of distinct traps and capture occasions in which each of the 44 individuals was captured is shown in Table 1. We see that 15 individuals were captured in more than one trap. There were 65 unique individual/occasions and a total of 68 encounter events. Thus there were only three "extra-trap captures" during the study, that is, individuals captured in >1 trap on the same sampling occasion (night). In both instances, this corresponds to individuals captured in a single occasion: one individual captured at two traps and another individual captured at three traps.

An important feature of the study design is that not all 120 trap stations were operated simultaneously. Instead, the reserve was subdivided into four blocks of approximately 30 trapping stations each, and each block was run for 12 consecutive days. Then, cameras were moved to the next block for another 12 days and the process repeated until all four blocks were sampled. This design follows sample design four of Karanth and Nichols (2002:133). Thus, some of the zeros are structural zeros (as opposed to sampling zeros), which we accommodate in the analysis of these data as described in *Spatial capture–recapture as a GLM with individual effects*.

In a previous analysis of these data (Royle et al. 2009) a multinomial capture model was applied, assuming that individuals could be captured in at most one trap per survey occasion. However, this model mis-specifies the true observation process. Because camera traps do not physically retain individuals, an individual tiger can be captured in multiple traps during any night. Failure to accommodate this information in the model induces some loss of efficiency in estimating density (see *Analysis of Nagarahole data*). We also note that in the analysis of Royle et al. (2009) there were $n = 45$ individuals. Subsequent reanalysis of the photographs revealed that one of these individuals represented only a recapture. Thus, the data analyzed here has one fewer individual, but one additional recapture compared to the analysis in Royle et al. (2009).

Model Formulation

The basic deficiency with the application of closed population models to data from trapping arrays is that space and movement have no explicit manifestation in such models. i.e., the models are not "spatial." Under these traditional models, N is just an integer-valued parameter that has no spatial context whatsoever. Thus, we seek to formalize the manner in which spatial organization of individuals is relevant to how they are observed.

State process model

A natural framework for developing spatial models of abundance is based on point process models, and point processes have been considered as the basis for spatial capture–recapture models in a number of recent efforts (Efford 2004, Borchers and Efford 2008, Royle and Young 2008). To develop this notion, we suppose that each individual in the population has a fixed point associated with it, its center of activity, $\mathbf{s}_i = (s_{1i}, s_{2i})$, a two-dimensional coordinate representing a point in space about which the movements of individual i are

TABLE 1. Summary encounter frequency data on tigers in the Nagarahole Reserve, India, camera trapping study during 48 nightly sampling occasions using 120 camera trap locations (e.g., nine tigers were captured on two nights and at two distinct camera trap locations).

Number of distinct traps	Number of occasions				
	1	2	3	4	5
1	28	1	0	0	0
2	1	9	1	0	0
3	1	0	1	0	1
4	0	0	0	1	0

concentrated. We suppose that these point locations represent the realization of a binomial point process. That is, we suppose there exists a population of N independent centers $\mathbf{s}_i$, $i = 1, 2, \ldots, N$, distributed uniformly over some region, say S, the state-space of the point process. We denote this assumption by

$$\mathbf{s}_i \overset{\text{ind}}{\sim} \text{Uniform}(\text{S}).$$

In practice, S will be prescribed (e.g., by specifying coordinates of some polygon that contains a trapping array).

The basic inference problem is to obtain an estimate of density, the number of activity centers per unit area of S, that is equivalent to an estimate of N under the point process model. We may also wish to estimate the number of activity centers in specific subsets of S, say some polygon $\text{P} \in \text{S}$. For example, P might be a national park, a reserve, or some block of contiguous habitat. We describe the model in some detail and then address this inference problem explicitly in *Estimating derived parameters*.

This uniform point process model represents a prior distribution for individual activity centers. While the assumptions of independence and uniformity are biologically untenable in many cases, we will see in our analysis of the tiger data that independence and uniformity of point locations in the prior does not preclude clustering or patchiness of point locations in the estimated posterior density of $\mathbf{s}$.

Observation models

Next we describe the juxtaposition of individual activity centers with the camera trapping array. We suppose that sampling is carried out by a network of J camera traps, having locations $\{\mathbf{x}_j, j = 1, 2, \ldots, J\}$. Further, we suppose that cameras function independently of one another. In particular, encounter by cameras is not mutually exclusive so that individuals can be encountered by multiple cameras. In the subsequent development of the observation models we will suppose that the probability of an individual being encountered by some camera j is a function of the distance from the camera to its activity center, and one or more parameters that will be estimated. Let $d_{ij} = ||\mathbf{s}_i - \mathbf{x}_j||$ be the distance between an individual i's activity center and camera j where $||\cdot||$ is the normal Euclidean distance.

In an ideal situation where cameras are operational continuously and individual encounter events are independent in time, an individual may be captured an arbitrary number of times yielding encounter frequencies y_{ijk} for individual i, in trap j, during interval k. One possible model for such encounter frequency data is the Poisson model:

$$y_{ijk} \sim \text{Poisson}(\lambda_0 g_{ij}) \tag{1}$$

where λ_0 is the baseline encounter intensity and g_{ij} is some function of distance between individual i and camera trap j which we will suppose equals 1 for $d_{ij} = 0$. Thus, for a trap that is located precisely at an individual's activity center, λ_0 is the expected number of captures in that trap. For example,

$$g_{ij} = \exp(-d_{ij}^2/\sigma) \tag{2}$$

which is a common "detection function" used in distance sampling and in the model described by Efford (2004). It can also be related explicitly to movement in some situations (Royle and Young 2008). In our analysis (below), we will also consider an exponential form for g_{ij}, i.e., $g_{ij} = \exp(-d_{ij}/\sigma)$ in order to assess the sensitivity to choice of model relating encounter to distance.

For the case where individuals can be captured at most once per trap, the observations are binary: $y_{ijk} = 1$ if individual i is captured in trap j during sample occasion k, and $y_{ijk} = 0$ otherwise. We will view the binary observations conceptually as reductions of the counts that we could have observed in the more general case. This might be most realistic for bear hair snare studies (and other DNA-based sampling) where an individual might be encountered a number of times during any period, and the biological material (hair, and so on) accumulates but cannot be partitioned into distinct visits after it is collected. In addition, as in our case, while camera traps may yield multiple captures during each occasion, it is difficult to imagine that such multiple captures are independent. Instead of either applying an arbitrary and subjective rule to determine how to partition events into independent recaptures, or devising a model of within-trap dependence, it is natural to reduce such data into binary encounter events. To formalize this, suppose we obtain binary observations y_{ijk} which are Bernoulli outcomes,

$$y_{ijk} \sim \text{Bernoulli}(\pi_{ij}) \tag{3}$$

with success probability that arises as the positive mass of the Poisson model described above:

$$\Pr(y_{ijk} = 1) = 1 - \exp(-\lambda_0 g_{ij}). \tag{4}$$

We will refer to the two models in Eqs. 1 and 3 as the Poisson and Bernoulli encounter models, respectively. Note that the parameters of the two distinct models are fundamentally equivalent, but the observable data under the second model is a reduced-information summary of what we would prefer to observe, the actual trap frequencies (see Royle and Nichols [2003] and Royle [2004] for a similar pairing of models).

Spatial Capture–Recapture as a GLM with Individual Effects

Note that the Bernoulli observation model, with $y_{ijk} \sim \text{Bernoulli}(\pi_{ij})$ and $\pi_{ij} = \Pr(y_{ij} = 1) = 1 - \exp(-\lambda_0 g_{ij})$, can be viewed as a logistic-regression type of model. Specifically, with the choice of g from Eq. 2 we have

$$\text{cloglog}(\pi_{ij}) = \log(\lambda_0) - (1/\sigma) d_{ij}^2$$

where cloglog(argument) = log[–log(1 − argument)] is the complementary log–log transform. For Poisson observations, we have a standard Poisson regression formulation:

$$\log(\lambda_{ij}) = \log(\lambda_0) + (1/\sigma)d_{ij}^2.$$

This specification reveals the essential simplicity of spatial capture–recapture models as GLM type models, and it elucidates a number of important aspects of this model.

(1) The model is closely related to both "individual heterogeneity" (Link 2003, Dorazio and Royle 2003) models and also individual covariate models (Royle 2009) with individual covariate "distance" being a deterministic function of the latent variable $\mathbf{s}_i$. The model is not precisely an individual covariate model because the variables $\mathbf{s}_i$ are unobserved even for the captured individuals. However the model is not entirely an individual heterogeneity model because they are partially observed for the captured individuals. In a sense, spatial capture–recapture models represent a conceptual intermediate between the two classes of models.

(2) We see that factors that influence λ_0 are linear effects on the cloglog π_{ij} scale. Thus, very general models can be represented as simple logistic-regression type models.

(3) We see that the (squared) distance between each trap and $\mathbf{s}_i$ is also a linear effect on the complementary log–log scale. Changing the distance function to be exponential produces a linear distance effect. Thus, in general, choice of distance function determines a transformation applied to the individual covariate.

Covariates that may influence detectability among individuals, or traps or through time may be modeled directly on the parameter λ_0 (regardless of the observation model under consideration). One situation that we will consider is the possibility of a behavioral response. In this case, we allow λ_0 to vary by individual and occasion so that

$$\log(\lambda_{0,ik}) = \alpha + \beta c_{ik}$$

where c_{ik} is an indicator covariate of previous encounter. Under this model, an individual's encounter probability may increase or decrease after initial encounter. The general model that we consider in our application (see *Analysis of Nagarahole data*) is of the form:

$$\text{cloglog}(\pi_{ijk}) = \alpha + \beta c_{ik} - (1/\sigma)d_{ij}^2.$$

A special kind of covariate is effort. In the context of camera trapping studies, this might be the time that a camera was operational, which could be modeled as described above (as a covariate on λ_0). In the Nagarahole study, traps were moved around such that only 30 of the 120 locations contained a trap on any particular night. This is a standard design both for camera trap studies, as well as so-called "hair snares" used for obtaining DNA, and other methods of detecting individuals. In this case, the parameter λ_0 has to be forced to 0 for such instances so that $\Pr(y = 1) = 0$. This is simple to handle in the analysis by defining

$$\Pr(y_{ijk} = 1) = 1 - \exp(-\lambda_0 m_{jk} g_{ij})$$

where $m_{jk} = 1$ if trap j is operational during occasion k and $m_{jk} = 0$, otherwise. Thus, whenever a trap is not operational, $\Pr(y_{ijk} = 1) = 0$, as it should.

Analysis of the Model with Known N

If we knew $\mathbf{s}$ and N then the models are simple Poisson or logistic regression models and inference would be no more difficult than under those conventional models. To introduce readers to analysis of the models, we provide the basic specification of the models in WinBUGS for this situation. This "conditional-on-$\mathbf{s}$" formulation of the model reveals the simplicity of the hierarchical model for camera trap array data. Further, the extension of the model to allow for $\mathbf{s}$ to be unknown is technically and conceptually straightforward. We formally address that situation in the following section.

As our analysis of the Nagarahole data makes use of the Bernoulli model, we consider that model here. The WinBUGS model specification for the Poisson model is provided in the Appendix. The implementation of this model in WinBUGS is given in Fig. 2 where the model is described in the standard WinBUGS pseudo-code. We can improve the efficiency of fitting the model in some restricted cases (e.g., when λ_0 is not time varying) by recognizing that the total number of captures of each individual in trap j is a binomial random variable based on a sample of size K. Then, the data can be reduced to the $n \times J$ matrix of capture frequencies (number of captures out of K samples).

Inference when N is Unknown

While we have described the models conditional on the variables $\mathbf{s}_i$, $i = 1, 2, \ldots, N$, they are unobservable quantities. Conceptually, these can be thought of as random effects in the usual sense of the concept as it is used in classical statistics. For analysis of random effects models, we adopt a prior distribution ("random effects distribution") for $\mathbf{s}$ and proceed with standard methods for analyzing such models. Precisely how we proceed depends in large part on whether we adopt a classical approach to the analysis of random effects or a Bayesian approach. In the classical treatment of random effects, we would remove them from the likelihood by integration. This was the strategy recently adopted by Borchers and Efford (2008) in a similar class of spatial capture–recapture models. Alternatively, Bayesian analysis of the random effects model is relatively straightforward. We have previously specified this random effects distribution under the binomial point process as

$$\mathbf{s} \sim \text{Uniform}(\mathrm{S}).$$

```
model {
   sigma~dunif(0,5)
   lam0~dgamma(.1,.1)

   for(i in 1:nind){
      for(j in 1:ntrap){
         dist2[i,j]<- ( pow(sx[i]-trap[j,1],2)  +  pow(sy[i]-trap[j,2],2) )
         mu[i,j]<- 1-exp( -lam0*exp(-dist2[i,j]/sigma) )
         for(k in 1:K){
            y[i,k,j] ~ dbern(mu[i,j])
         }
       }
    }
}
```

FIG. 2. WinBUGS model specification for the Bernoulli encounter model when $\mathbf{s}_i$ are known for $i = 1, 2, \ldots, N$ independent centers (where $\mathbf{s}_i$ is a two-dimensional coordinate representing a point in space about which the movements of individual i are concentrated). In this model description, sx and sy are the x- and y-coordinates of each individual activity center, respectively.

Whereas when N is known, the model is just a form of generalized linear mixed model (with random effects $\mathbf{s}_i$), the difficulty in analyzing the model with unknown N is that the dimension of the parameter space (the number of "random effects") is itself an unknown quantity. It is this problem that motivated the analysis of similar models using the method of data augmentation (Royle et al. 2007, Royle and Dorazio 2008). In effect, data augmentation allows us to analyze a version of the "complete data" model: the model with a fixed number of activity centers.

To implement data augmentation, we physically augment the n observed encounter histories with some large number of "all-zero" histories, say $M - n$ such histories. In the present case, the encounter history for each augmented individual is the two-dimensional array of $J \times K$ zeros. We assume that this list of M pseudo-individuals includes the actual N individuals in the population as a subset. We must choose M sufficiently large so as that the posterior of N is not truncated. This can be achieved by trial and error with no philosophical or practical consequence. Given the augmented data set, the key result (Royle et al. 2007) is that the model for the augmented data is a zero-inflated version of the "known-N" model, i.e., that corresponding to the case where N is known. There are no additional parameters to estimate, but the parameter N is replaced by a zero-inflation parameter, say $1 - \psi$. The parameter ψ is the probability that an individual on the list of size M is a member of the population of size N that was exposed to sampling by the trap array.

Analysis by data augmentation has a formal Bayesian development, which can be motivated by the assumption of a discrete uniform prior for N on the integers $0, 1, \ldots, M$. This prior can be specified hierarchically in the form of a Binomial prior for N: $N \sim \text{Bin}(M, \psi)$ and a uniform prior for ψ: $\psi \sim \text{Unif}(0, 1)$. Integration of the binomial prior for N over the uniform prior for ψ yields the discrete uniform prior for N. This hierarchical specification is useful because it yields a convenient Bayesian implementation. Namely, we introduce $M - n$ observations of $y_i = 0$ for $i = n + 1, n + 2, \ldots, M$. For these individuals, there is no trap information and thus the encounter history record is $y_{ijk} = 0$ for all j and k as well. Given the augmented data set, we now introduce a set of latent indicator variable w_i, $i = 1, 2, \ldots, M$, such that $w_i = 1$ if the ith element of the augmented list is a member of the population of size N, and $w_i = 0$ otherwise (an "excess zero"). We impose the model $w_i \sim \text{Bern}(\psi)$. With a uniform(0, 1) prior on ψ, the induced prior distribution on $N = \Sigma_i w_i$ is uniform on the integers $0, 1, \ldots, M$, as noted above.

Estimating derived parameters

In some instances, there will be interest in derived parameters. That is, parameters that are not canonical or structural parameters of the model (i.e., λ_0, σ, and ψ introduced by data augmentation). For example,

$$N(S) = \sum_{i=1}^{M} w_i$$

is the population size of individuals on the set S and $D(S) = N(S)/A(S)$ is the density on S (here, $A(S)$ denotes the area of S). Similarly, the number of activity centers in any prescribed polygon, say P, is the quantity denoted by $N(P)$. This is calculated by simply tallying up the number of $\mathbf{s}_i$ contained in P, and for which $w_i = 1$ (the data augmentation indicator variable), at each iteration of the Markov chain Monte Carlo (MCMC) algorithm. The resulting sequence of $N(P)$ (one value for each iteration of the MCMC algorithm) constitutes a sample from the desired posterior distribution. In addition, density, say $D(P) = N(P)/A(P)$, is also a derived parameter.

```
model {
    sigma~dunif(0,20)
    lam0~dgamma(.1,.1)
    psi ~ dunif(0,1)
    for(i in 1:nind){
       w[i]~dbern(psi)
       sx[i]~dunif(Xl,Xu)
       sy[i]~dunif(Yl,Yu)

       for(j in 1:ntrap){
          dist2[i,j]<- ( pow(sx[i]-trap[j,1],2)  +  pow(sy[i]-trap[j,2],2) )
          mu[i,j]<-w[i]*(1-exp(-lam0*exp(-dist2[i,j]/sigma)))
          y[i,j] ~ dbin(mu[i,j],K)
       }
    }
   N<-sum(w[1:nind])
 }
```

FIG. 3. WinBUGS model specification for the Bernoulli encounter model for unknown $\mathbf{s}_i$ and N. In this model description, sx and sy are the x- and y-coordinates of each individual activity center, respectively.

Implementation

While developing the MCMC algorithm for analysis of the augmented data under the models is straightforward, we avoid those technical details because the model can also be implemented in WinBUGS. Applications of data augmentation in similar models can be found in Royle and Young (2008), Royle (2009), and Gardner et al. (2009). One detail that we have avoided discussion of is that Bayesian analysis requires that we specify prior distributions for model parameters. In the analyses below, we used a uniform(0, 1) prior for the data augmentation parameter ψ, and flat normal priors for any regression coefficients in the model, including the intercept, behavioral response and $\log(1/\sigma)$ (as this is essentially a regression coefficient on the distance covariate). The WinBUGS specification of the Bernoulli model for unknown N is shown in Fig. 3. As shown in Fig. 3, the encounter histories have been aggregated over the number of sampling occasions. In that case, the total number of encounters of each individual, and in each trap, is a binomial random variable with sample size K, as indicated in Fig. 3. Additional models for the Poisson encounter model are given in the Appendix.

ANALYSIS OF NAGARAHOLE DATA

For the analysis of the Nagarahole data, we excluded areas that were judged to be non-habitat within a 15 km buffer area containing the trap array (Fig. 1). This renders the definition of a uniform prior for the activity centers difficult because the polygon is highly irregular. As such, we described this region of suitable habitat by a grid of 9961 equally spaced points, each representing approximately 0.336 km^2 over the buffered region. Of these, 4898 (1645.7 km^2) were judged to represent suitable habitat. The activity centers $\mathbf{s}_i$ were therefore assumed to be uniformly distributed over this discrete space of 4898 points, an area of approximately 1645.7 km^2. We developed an implementation of the model for this discrete state-space situation in the R programming language. The coordinate system was scaled so that a standard unit was 5 km, and thus also are the units of σ.

As described in *Spatial capture–recapture as a GLM with individual effects*, we considered the Bernoulli encounter model which allows a single capture per trap in each occasion. The general model considered was of the form

$$\text{cloglog}(\pi_{ijk}) = \alpha(1 - c_{ik}) + \beta c_{ik} - (1/\sigma)d_{ij}^2 \qquad (5)$$

where c_{ik} is an indicator covariate that takes on the value $c_{ik} = 1$ if individual i was encountered in a sample occasion prior to k. Here, we have reparameterized the intercept α in order to interpret α and β as the log-encounter intensity parameters for individuals pre- and post-encounter, respectively. Bayesian analysis of this model was carried out using flat priors for the regression coefficients α and β and for $\log(1/\sigma)$. We considered also a model with a linear distance term, which corresponds to an exponential detection function as noted previously.

Given the simple formulation of the model for encounter probability as a generalized linear mixed model, we could conceivably extend the model to arbitrary levels of complexity. For example, time effects or individual heterogeneity (Dorazio and Royle 2003) could also be considered. However, because of the sparsity of our data set and low encounter rate (24 total recaptures), we made several intentional decisions to limit the complexity of the model. In particular, we have more sampling occasions than individuals (48 vs. 44), and they are short intervals (nightly), and so we opted

TABLE 2. Posterior summaries of model parameters for the tiger camera trapping data.

Parameter	Mean	SD	2.5%	Median	97.5%
Half-normal model					
σ	0.3130	0.0761	0.2010	0.3006	0.4957
α	0.0159	0.0043	0.0083	0.0150	0.0246
β	0.0297	0.0088	0.0149	0.0289	0.0478
$1/\sigma$	3.3734	0.7710	2.0172	3.3261	4.9742
ψ	0.4970	0.1008	0.3364	0.4839	0.7335
p_1	0.0157	0.0042	0.0082	0.0149	0.0243
p_2	0.0292	0.0086	0.0148	0.0285	0.0467
D	13.4132	2.6484	9.2968	13.0641	19.5658
Exponential model					
σ	0.2348	0.0328	0.1808	0.2317	0.3093
α	0.0435	0.0130	0.0224	0.0419	0.0688
β	0.0451	0.0187	0.0245	0.0379	0.0939
$1/\sigma$	4.3407	0.5925	3.2333	4.3159	5.5299
ψ	0.5078	0.1041	0.3367	0.4965	0.7521
p_1	0.0425	0.0124	0.0221	0.0411	0.0665
p_2	0.0440	0.0176	0.0242	0.0372	0.0897
D	13.7066	2.7508	9.3576	13.3679	20.2340

Notes: Summaries in the top half of the table correspond to the half-normal detection function, and summaries in the bottom half of the table correspond to an exponential distance function; 2.5% and 97.5% are posterior percentiles. The parameters p_1, p_2, and density, D, are derived parameters. Density units are individuals per 100 km^2, p_1 is the encounter probability for individuals that have not previously been encountered, and p_2 is the encounter probability for individuals subsequent to their initial encounter.

not to consider the additional complexity of temporal variation in detectability. We also did not consider individual heterogeneity in detection probability. Such models would be poorly identifiable with only 44 individuals (Dorazio and Royle 2003), and exhibit extreme sensitivity to model choice (Link 2003). We believe that spatial proximity of individuals to traps should be the predominant mechanism responsible for heterogeneity and seek to effectively model that phenomenon using the spatial covariate. In light of Link (2003), we believe that researchers should make their own judgment as to whether individual heterogeneity should be fit in addition to a spatial covariate because this is a decision that cannot be made objectively in small samples.

Results

Posterior summaries from fitting the model with the half-normal detection function (i.e., corresponding to a quadratic distance effect) and "behavioral response" to capture are given in Table 2. Recall that the number of unique individuals observed was 44. The estimate (posterior mean) of tiger density over S is 13.4 tigers/100 km^2, with a 95% posterior interval of (9.3, 19.6). We contrast these estimates with those reported for a misspecified multinomial observation model (Royle et al. 2009) who reported an estimated density over S of approximately 14.30 with a 95% posterior interval of (2.8, 20.5). We observe a similar estimate of density but a substantial increase in precision. The result suggests that the "non-encounters," which are independent observations under the Bernoulli model considered here, provide considerable information about model parameters. Note that the data set analyzed here has one fewer individual, and so this contributes to the lower estimated density.

Fig. 4 shows the posterior density of the point process $\mathbf{s}$. Each pixel is marked with $\log(E[N(\mathbf{s}) \mid \text{data}])$ where $N(\mathbf{s})$ is the number of activity centers located in pixel $\mathbf{s}$. We note the extreme spatial variation and the high density in the western corner of the trap array which is an area known to have high prey densities (S. Kumar, *unpublished data*). The scale in Fig. 4, roughly -4 to -2, equates to a density range from approximately 0.14 to 0.018 (tigers per 0.336-km^2 pixel), respectively. Scaling these figures to 100 km^2 yields a density range between 5.35 and 42 tigers per 100 km^2. Thus the highest density areas have a density of eight times that of the lowest density areas.

The distance covariate appears to be highly important. The posterior mean of the coefficient, $1/\sigma$ is very positive and the posterior mass is concentrated away from 0 (95% interval: (2.017, 4.974)). There appears to be a moderate behavioral response to encounter in this model. The encounter probabilities (posterior means) for individuals pre- and post-initial encounter are $p_1 = 0.0157$ and $p_2 = 0.0292$, respectively. These are related to α and β according to $1 - \exp(-\alpha)$ and $1 - \exp(-\beta)$. Thus, once individuals are captured, there appears to be almost a doubling of encounter probability in subsequent occasions. The estimated difference is imprecise, however. The posterior probability of a positive response is $\Pr(\beta - \alpha > 0) = 0.92$. While a positive trap response ("trap-happiness") is indicated, this is not realistic biologically. We believe that this is likely a result of some non-independence among encounters. For example, if space usage (i.e., movement) is not random for individuals, we think this could appear as a

positive behavioral response. This might happen if individual tigers favor certain trails for moving about their territory.

We also fitted the exponential detection model (Table 2). We see that the summaries of N and density are little changed, and we don't have any intuition as to whether or not we would expect sensitivity in this regard. Interestingly, the behavioral response appears much less important (the posterior was nearly centered at 0).

Assessment of model adequacy

We evaluated the fit of each model using a Bayesian P value (Gelman et al. 1996). The basic idea is to describe a metric of model adequacy and then compare posterior draws of that metric to those obtained from data sets simulated from the posterior distribution. If we denote a posterior sample of this metric for the observed data set as $D_i^{(\text{obs})}$ for posterior sample i and for a data set simulated from the posterior distribution as $D_i^{(\text{new})}$, then the Bayesian P value is $\Pr(D_i^{(\text{obs})} > D_i^{(\text{new})})$. For a model that provides an adequate description of the data this should be near 0.50. Thus, extreme values (near 0 or 1) indicate a lack of model adequacy.

A practical problem with this approach in the context of the model under consideration is the lack of an obvious omnibus fit statistic. For our analysis we considered two distinct measures of model adequacy: One to assess the model's ability to describe the encounter frequency distribution and another statistic to evaluate the clustering of individual captures in space. For the former we considered a sum-of-squares between the observed trap and occasion encounter frequencies and their fitted values. That is, $\text{FIT}_1 = \Sigma_k\ (n_k - E[n_k])^2$ where n_k is the number of individuals captured k times (or in k traps). We aggregated the sum of squares of both components (traps and occasions). For the assessment of clustering (say FIT_2), we used recaptures of individuals to compute within-individual sum of squares of capture locations (this is analogous to the SSE of an ANOVA with individuals as blocks).

The spatial model with either half-normal or exponential detection functions showed about the same fit by both fit statistics. For the half-normal model the P values for FIT_1 and FIT_2 were 0.48 and 0.47, respectively. For the exponential model, the P values were 0.52 and 0.61. For comparison, we also computed the Bayesian P values for model M_0 which is the model described previously setting $1/\sigma$ and β to 0. The statistic based on encounter frequencies suggested a fit of that model as well ($P = 0.58$), which is perhaps not surprising given the sparsity of the observations (see Table 1). Conversely, the assessment of spatial clustering using FIT_2 indicated inadequacy of model M_0 ($P < 0.001$).

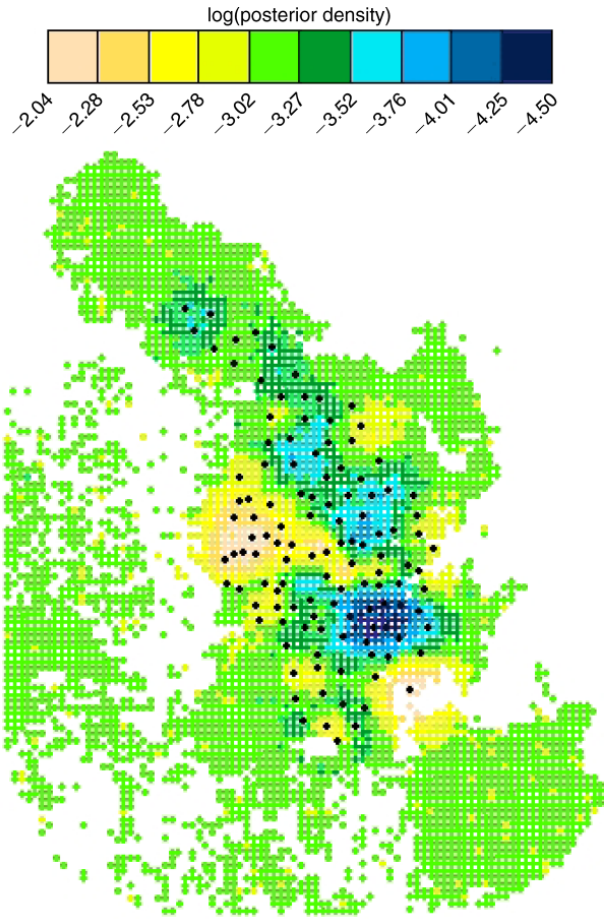


FIG. 4. Logarithm of the estimated posterior density of tigers in the Nagarahole reserve. Each pixel is marked with $\log(E[N(\mathbf{s}) \mid \text{data}])$ where $N(\mathbf{s})$ is the number of points at pixel $\mathbf{s}$. Lighter colors (yellow) indicate higher densities, while darker colors (blue) correspond to lower densities.

DISCUSSION

One of the fundamental objectives of many camera trapping studies is the estimation of abundance and density of the species under study. Historically, this estimation problem has been addressed using a large number of essentially ad hoc or heuristic methods based on closed population capture–recapture estimators of population size applied to individual encounter history data. The conceptual limitation of closed population estimators is that, while the estimate of N may be valid in the sense of estimating the size of a population exposed to sampling, the effective sample area of the trapping array is unknown. Conventional methods have sought to estimate effective sample area using methods not formally linked (by a statistical model) to the observed spatial encounter history data. For example, N might be estimated by a conventional estimator of population size for closed populations, and then a buffer applied to the trapping grid based on observed movements of individuals. Because the underlying models are not specified precisely, they are not sufficiently flexible or extensible. For example, the formal treatment of multiple captures has not been integrated directly into such analyses.

Here we described a hierarchical modeling framework for inference from spatial capture–recapture data for



PLATE 1. Camera trap data from tigers in Nagarahole Reserve, India, have facilitated the development of spatially explicit capture–recapture models. Photo credit: WCS/K. U. Karanth.

trapping methods wherein the traps function independently of one another such that multiple encounters of an individual may occur during each sampling occasion. This is typical of camera trapping studies as well as other methods including DNA sampling from arrays of "hair snares." The spatial capture–recapture model is developed by conditioning the encounter history data on the realization of an underlying point-process that describes the distribution of individuals in space. This concept was first adopted in the context of spatial capture–recapture models by Efford (2004), and also exploited recently by Royle and Young (2008), Borchers and Efford (2008), and Royle et. al. (2009) for estimating density in the context of a multinomial observation model wherein an individual can be captured in only a single trap.

We noted that models in which traps function independently of one another can be formulated as generalized linear models (GLMs) with random effects (i.e., generalized linear *mixed* models; GLMMs), similar to other classes of models in capture–recapture including individual heterogeneity models ("Model M_h") and individual covariate models. For models in which individuals can be captured >1 time in a single trap, a natural model for the count frequencies is a Poisson model but obviously other models for count frequencies could be considered (see discussion below). A binary encounter model can be derived as a reduction of the model for the frequency encounter model (i.e., as the event that $y > 0$). Technically, these spatial capture–recapture models are GLMMs only when N is known. In this case, the activity centers, $\mathbf{s}$, are the random effects. When we allow for N to be unknown, the resulting models (when reformulated using data augmentation) are essentially zero-inflated versions of the corresponding generalized linear mixed models. Analysis of these models can be achieved in WinBUGS, which we believe makes them generally accessible to practitioners. However, in our analysis of the Nagarahole camera trapping data, we developed an implementation of the models in R for the case where the point process has a discrete state-space, allowing us to distinguish between suitable and unsuitable habitat. In addition to the conceptual and technical relationships between spatial capture–recapture models and generalized linear models, there is one other class of models that are similar in form to spatial capture–recapture models. If we consider aggregating total detections by trap, the Poisson model has a structural similarity to the Poisson-Gamma convolution models described by Wolpert and Ickstadt (1998) for modeling spatially indexed counts. However, in this case we have a single intensity parameter and allow the "support points" $\mathbf{s}_i$ to be unknown. Thus, potentially, the model developed here could be employed in similar contexts to that of Wolpert and Ickstadt (1998), perhaps permitting a lower-dimensional set of support points that adapts to the data.

While the Poisson model (or other model for count frequencies) might be applicable for some camera trapping studies because individuals can potentially be encountered multiple times at the same camera location during any particular sampling occasion, the data are often reduced to binary encounter events either for convenience or because recaptures are seldom independent within short sampling periods (i.e., nights). Another reason to reduce within-occasion recaptures to binary events is that attention to modeling the within trap/occasion variability becomes necessary. Moreover, this information may not be directly informative about density. Heuristically, information in spatial capture–recapture data originates from spatial clustering of recaptures. Thus, an individual that is only ever captured in the same trap is providing no information about the parameter that governs detection as a function of distance (σ in our model). Another reason to be concerned with modeling this extra source of variability is related to the results of Link (2003) who argued that N is not an identifiable parameter in individual heterogeneity models when the mixture distribution is not known. Consider adding an extra source of noise to the linear predictor (Eq. 5) in order to allow for extra-Poisson variation in trap-specific encounter frequencies. This type of heterogeneity, while seemingly natural to try to model, leads immediately to a model that resembles the classical individual heterogeneity model of the type that Link (2003) addressed. Thus, for most practical situations involving a small number of individuals, a priori limiting the complexity of the model is probably advisable.

We have adopted a Bayesian formulation of spatial capture–recapture models in this paper. However, these models could also be analyzed by integrated likelihood which is the technical approach adopted by Borchers and Efford (2008). However, we believe the hierarchical formulation adopted in our analysis (and in Royle and Young 2008) will prove more flexible in the development of model extensions. For example, a fundamental component of the hierarchical model is the underlying point process model that governs the distribution of individual activity centers. The models have been developed here under the assumption that individuals are distributed uniformly in space. As the underlying point process model is made more complex (e.g., containing interactions and conditional dependencies) the integration required for inference by integrated likelihood may become computationally prohibitive. However, Bayesian analysis of the hierarchical formulation only requires (in principle) the capability to carry out conditional simulation of the activity center locations. That is, we need only be able to simulate from general point processes in order to carry-out inference within a Markov chain Monte Carlo framework.

While extensibility of the point process model is a potential virtue of the Bayesian formulation of spatial capture–recapture models, we feel that the biggest practical advantage apparent at this time has to do with the validity of inferences achieved by Bayesian analysis. In particular, classical inference procedures are asymptotic and as such their relevance to small sample situations is questionable. Conversely, Bayesian inferences do not rely on asymptotic arguments and are valid regardless of the sample size. There seems to be a prevailing view in statistical ecology that classical likelihood-based procedures are virtuous because of the availability of simple formulas and procedures for carrying out inference, such as calculating standard errors, doing model selection by AIC, and assessing goodness of fit. In large samples, this is an important practical benefit of classical likelihood-based inference. However, the practical validity of these procedures cannot be asserted in most situations involving small samples. In the study which motivated our analysis, there were a total of 68 encounter events on 44 individuals. Only 24 of these are recaptures, from which the information about the encounter process is obtained. Reliance on conventional asymptotic procedures seems difficult to justify in this context. We note that the size of our data set is similar (or even larger) than many spatial capture–recapture studies on carnivores that we are familiar with, including a study of ocelots (Trolle and Kéry 2003), black bears (Gardner et al. 2009), Pampas cats in Argentina (B. Gardner, J. Reppucci, M. Lucherini, and J. A. Royle, *in review*), European wildcats in Switzerland (M. Kéry, B. Gardner, T. Stoeckle, D. Weber, and J. A. Royle, *unpublished manuscript*) and wolverines in Alaska (Magoun et al. 2008). We also believe that some restraint should be exercised in developing many (or overly complex) models for small data sets. While our framework is quite general and flexible in terms of model development (as a generalized linear mixed model) the ability to fit models of arbitrary complexity should not be perceived as a requirement to do so.

Modern technology has greatly advanced our ability to obtain information about the demography and population dynamics of secretive animal populations. Photographic identification, DNA from dung, and hair snares, or identification of individuals from scent dogs are now widely used in studies of many taxa. Almost universally these methods generate spatial encounter histories and also data which deviate from the standard multinomial structure of one capture per sample occasion. For such data, generalizations of the capture–recapture modeling framework, such as those models we have presented here, are necessary to make efficient use of sparse data that is typically expensive of both time and effort to obtain.

Acknowledgments

We are grateful to the Wildlife Conservation Society, New York, for supporting this study. We are indebted to Jim Nichols for contributing substantially to the conceptual development of

this work. We thank two anonymous referees for helpful reviews of a draft of this paper.

Literature Cited

Borchers, D. L., and M. G. Efford. 2008. Spatially explicit maximum likelihood methods for capture–recapture studies. Biometrics 64:377–385.

Dorazio, R. M., and J. A. Royle. 2003. Mixture models for estimating the size of a closed population when capture rates vary among individuals. Biometrics 59:351–364.

Efford, M. 2004. Density estimation in live-trapping studies. Oikos 106:598–610.

Gardner, B., J. A. Royle, and M. T. Wegan. 2009. Hierarchical models for estimating density from DNA mark–recapture studies. Ecology 90:1106–1115.

Gelman, A., X. L. Meng, and H. Stern. 1996. Posterior predictive assessment of model fitness via realized discrepancies. Statistica Sinica 6:733–759.

Gilks, W. R., A. Thomas, and D. J. Spiegelhalter. 1994. A language and program for complex Bayesian modelling. Statistician 43:169–178.

Henschel, P., and J. Ray. 2003. Leopards in African rainforests: survey and monitoring techniques. WCS Global Carnivore Program. ⟨http://www.savingwildplaces.com/media/file/LeopardManual_English_lowres.pdf⟩

Karanth, K. U. 1995. Estimating tiger (*Panthera tigris*) populations from camera-trap data using capture–recapture models. Biological Conservation 71:333–338.

Karanth, K. U., and J. D. Nichols. 1998. Estimation of tiger densities in India using photographic captures and recaptures. Ecology 79:2852–2862.

Karanth, K. U., and J. D. Nichols. 2002. Monitoring tigers and their prey: a manual for researchers, managers and conservationists in tropical Asia. Centre for Wildlife Studies, Bangalore, India.

Karanth, K. U., J. D. Nichols, N. S. Kumar, and J. E. Hines. 2006. Assessing tiger population dynamics using photographic capture–recapture sampling. Ecology 87:2925–2937.

Karanth, K. U., J. D. Nichols, N. S. Kumar, W. A. Link, and J. E. Hines. 2004. Tigers and their prey: predicting carnivore densities from prey abundance. Proceedings of the National Academy of Sciences (USA) 101:4854–4858.

Link, W. A. 2003. Nonidentifiability of population size from capture–recapture data with heterogenous capture probabilities. Biometrics 59:1123–1130.

Maffei, L., E. Cuellar, and A. Noss. 2004. One thousand jaguars (*Panthera onca*) in Bolivia's Chaco? Camera trapping in the Kaa-Iya National Park. Journal of Zoology 262:295–304.

Magoun, A. J., P. Valkenburg, and R. E. Lowell. 2008. Habitat associations and movement patterns of reproductive female wolverines (*Gulo gulo luscus*) on the Southeast Alaska mainland. Wildlife Research Annual Progress Report. Alaska Department of Fish and Game, Petersburg, Alaska.

Parmenter, R., et al. 2003. Small-mammal density estimation: a field comparison of grid-based vs. web-based density estimators. Ecological Monographs 73:1–26.

Royle, J. A. 2004. N-mixture models for estimating population size from spatially replicated counts. Biometrics 60:108–115.

Royle, J. A. 2009. Analysis of capture–recapture models with individual covariates using data augmentation. Biometrics 65:267–274.

Royle, J. A., and R. M. Dorazio. 2008. Hierarchical modeling and inference in ecology. Academic Press, San Diego, California, USA.

Royle, J. A., R. M. Dorazio, and W. A. Link. 2007. Analysis of multinomial models with unknown index using data augmentation. Journal of Computational and Graphical Statistics 16:67–85.

Royle, J. A., and J. D. Nichols. 2003. Estimating abundance from repeated presence–absence data or point counts. Ecology 84:777–790.

Royle, J. A., J. D. Nichols, K. U. Karanth, and A. M. Gopalaswamy. 2009. A hierarchical model for estimating density in camera-trap studies. Journal of Applied Ecology 46:118–127.

Royle, J. A., and K. V. Young. 2008. A hierarchical model for spatial capture–recapture data. Ecology 80:2281–2289.

Soisalo, M., and S. Cavalcanti. 2006. Estimating the density of a jaguar population in the Brazilian Pantanal using camera-traps and capture–recapture sampling in combination with GPS radio-telemetry. Biological Conservation 129:487–496.

Trolle, M., and M. Kéry. 2003. Estimation of ocelot density in the Pantanal using capture–recapture analysis of camera-trapping data. Journal of Mammalogy 84:607–614.

Trolle, M., and M. Kéry. 2005. Camera-trap study of ocelot and other secretive mammals in the northern Pantanal. Mammalia 69:409–416.

Wallace, R., H. Gomez, G. Ayala, and F. Espinoza. 2003. Camera trapping for jaguar (*Panthera onca*) in the Tuichi Valley, Bolivia. Journal of Neotropical Mammalogy 10:133–139.

Wilson, K. R., and D. R. Anderson. 1985*a*. Evaluation of two density estimators of small mammal population size. Journal of Mammalogy 66:13–21.

Wilson, K. R., and D. R. Anderson. 1985*b*. Evaluation of a nested grid approach for estimating density. Journal of Wildlife Management 49:675–678.

Wolpert, R. L., and K. Ickstadt. 1998. Poisson/gamma random field models for spatial statistics. Biometrika 85:251–267.

APPENDIX

Analysis of the Poisson Encounter Model (*Ecological Archives* E090-230-A1).

Appendix E

Factors influencing densities of striped hyenas (Hyaena hyaena) in arid regions of India

Journal of Mammalogy, 91(5):1152–1159, 2010

Factors influencing densities of striped hyenas (*Hyaena hyaena*) in arid regions of India

PRIYA SINGH,* ARJUN M. GOPALASWAMY, AND K. ULLAS KARANTH

Postgraduate Program in Wildlife Biology and Conservation, Wildlife Conservation Society–India Program & Centre for Wildlife Studies, National Centre for Biological Sciences, GKVK Campus, Bangalore-560 065, India (PS)
Wildlife Conservation Society–India Program, Centre for Wildlife Studies, 1669, 31st Cross, 16th Main, Banashankari 2nd Stage, Bangalore-560 070, India (AMG)
Wildlife Conservation Society, Centre for Wildlife Studies, 26-2, Aga Abbas Ali Road (Apartment 403), Bangalore-560 042, India (KUK)
Present address of PS: Centre for Wildlife Studies, 1669, 31st Cross, 16th Main, Banashankari 2nd Stage, Bangalore-560 070, India
Present address of AMG: Wildlife Conservation Research Unit (WildCRU), The Recanati-Kaplan Centre, University of Oxford, Department of Zoology, Tubney House, Abingdon Road, Tubney, Abingdon OX13 5QL, United Kingdom

** Correspondent: karnisar@gmail.com*

The striped hyena (*Hyaena hyaena*), despite being a threatened species, frequently occurs in human-dominated landscapes of India's Rajasthan State. We estimated hyena densities using photographic capture–recapture sampling to identify key ecological factors influencing hyena abundances in such areas. Our 2 study sites (307 km^2 in Kumbhalgarh Wildlife Sanctuary and 218 km^2 in Esrana Forest Range) had different topographies and levels of human disturbances. We proposed explicit hypotheses regarding effects of livestock densities and topographic features on hyena abundances. We tested these hypotheses by examining the correspondence of estimated hyena densities to food availability in the form of livestock carcasses and potential refugia offered by hilly terrain. Sampling efforts of 548 and 538 camera-trap nights were invested in Kumbhalgarh and Esrana, respectively. Density estimates (hyenas/100 km^2) based on capture–recapture sampling were higher (6.5 ± 2.6 *SE*) for Kumbhalgarh than Esrana (3.67 ± 0.3 *SE*). Our results supported the prediction that denning refugia in hilly terrain sustain higher hyena densities, but the prediction that higher livestock densities maintain higher hyena densities was not supported. Because the striped hyena is a threatened species for which few data exist, our findings have major potential utility for range-wide conservation of the species. DOI: 10.1644/09-MAMM-A-159.1.

Key words: abundance estimation, capture–recapture, denning refugia, land use, livestock, semiarid zones

© 2010 American Society of Mammalogists

The striped hyena (*Hyaena hyaena*) is a large carnivore that prefers rocky and open landscapes (Hofer 1998; Reiger 1981) within semiarid and arid ecosystems in the tropics (Leakey et al. 1999; Mendelssohn and Yom-Tov 1999; Wagner 2006). In India, arid regions in Rajasthan State are important habitats for hyena persistence (Karanth et al. 2009, 2010). Rajasthan is among the most densely populated arid regions of the world for both humans (Baqri and Kankane 2001) and livestock (Rahmani and Soni 1997), with only 4.3% of land area held in legally protected nature reserves (Kankane 2009). Consequently, conservation efforts urgently require knowledge of hyena ecology in dry zones.

Population size is a key ecological parameter for understanding the biology and conservation status of a species (Williams et al. 2002). However, traditional methods such as pugmark censuses used by wildlife managers in India to count large carnivores are seriously flawed (Karanth et al. 2003) relative to modern approaches to sampling animal populations (Williams et al. 2002). Robust estimates of striped hyena densities in India have not been published from dry zones, although some studies are currently in progress (Y. V. Jhala, Wildlife Institute of India, pers. comm.). Data on environmental factors influencing hyena densities also are limited.

Because of the complex nature of ecological processes, interactions between several biotic and abiotic factors (Williams et al. 2002) potentially can influence animal

densities. However, measuring all such factors to identify those that significantly influence hyena densities using classical experiment-based hypothesis testing is impractical for field studies of large carnivores. Therefore, we could not use classical hypothesis-testing paradigms that often must satisfy assumptions such as normally distributed point estimates and equal variances in linear regression or analysis of variance approaches. Instead, we chose an evidence-based approach more practical (Burnham and Anderson 2002) for our observational study. We used a probabilistic modeling framework for explicitly comparing a priori ecological predictions to observed data (Bolker 2008; Burnham and Anderson 2002; Hilborn and Mangel 1997; Royle and Dorazio 2008; Williams et al. 2002).

Based on prior knowledge of striped hyena biology (Hofer 1998; Kruuk 1976; Leakey et al. 1999; Mendelssohn and Yom-Tov 1999; Reiger 1981; Wagner 2006) and nature of human impacts, we identified 2 ecological factors that appeared to be important in influencing hyena densities in our 2 study areas that differed in ecology and levels of human impacts. We observed that striped hyenas are predominantly scavengers on domestic ungulate carcasses in the region (see Appendices I and II). Most livestock in Rajasthan consists of cattle, goats, and sheep that exist at very high densities ranging from 42 animals/km^2 to 226 animals/km^2 (Rahmani and Soni 1997) reared under conditions of semistarvation by uncontrolled free-grazing on overloaded ranges (Robbins 1998; Sharma and Mehra 2009). In combination with inadequate veterinary care, this leads to high livestock mortality rates. A religiously rooted avoidance of meat in general, and beef specifically, by a large proportion of the population (Chhangani 2009; Karanth et al. 2009) increases availability of unexploited carcasses for hyenas. Therefore, we hypothesized that a facultative scavenger like the striped hyena might benefit more in this circumstance than obligate predators that need to hunt prey in such a setting (Bagchi et al. 2003; Carbone and Gittleman 2002; Karanth et al. 2004).

Despite abundant food availability, hyenas with cubs or at rest are vulnerable to harassment by humans and predation by feral dogs in daytime. Therefore, we additionally hypothesized that availability of secure den sites (Prater 1948) also could be critical to hyena survival. Steep and rocky terrain, unsuitable for agriculture or pasturage and hence unattractive to humans or guard dogs, offers such refugia in the study areas. Thus, our specific objectives were to test whether, comparatively, higher livestock densities would support higher hyena densities, and whether a site with higher proportion of hilly terrain would support higher hyena densities.

Although additional habitat-related covariates, such as targeted hunting by humans or presence of other large predators, could influence hyena densities in other ecological contexts, we did not consider these as critical because wildlife legislation in India strictly prohibits hunting of hyenas throughout the country, and social tolerance of hyenas is typical of the region (Karanth et al. 2009, 2010). Although leopards were present at both sites, the specialized scavenging niche occupied by hyenas did not appear to overlap substantially with the leopard niche as obligate hunters of wild and domestic prey animals. Furthermore, both study areas had lower numbers of leopards camera-trapped compared to hyenas, suggesting that leopard numbers could be much lower (P. Singh, pers. obs.). Because striped hyenas could be individually identified from differences in stripe patterns and other natural marks, we used a photographic capture–recapture sampling method (Karanth and Nichols 1998, 2002b) to estimate their densities to test our predictions.

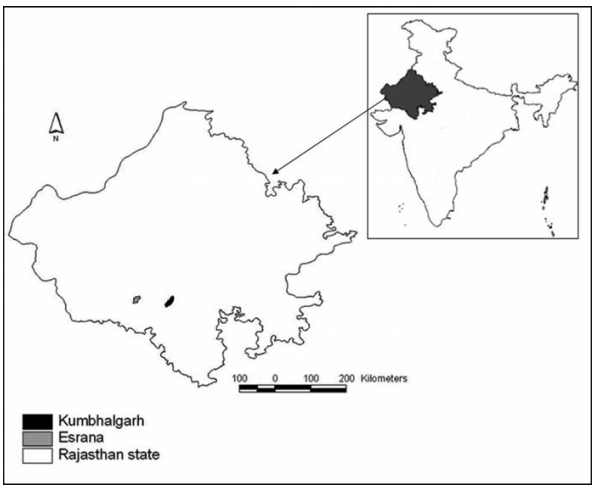
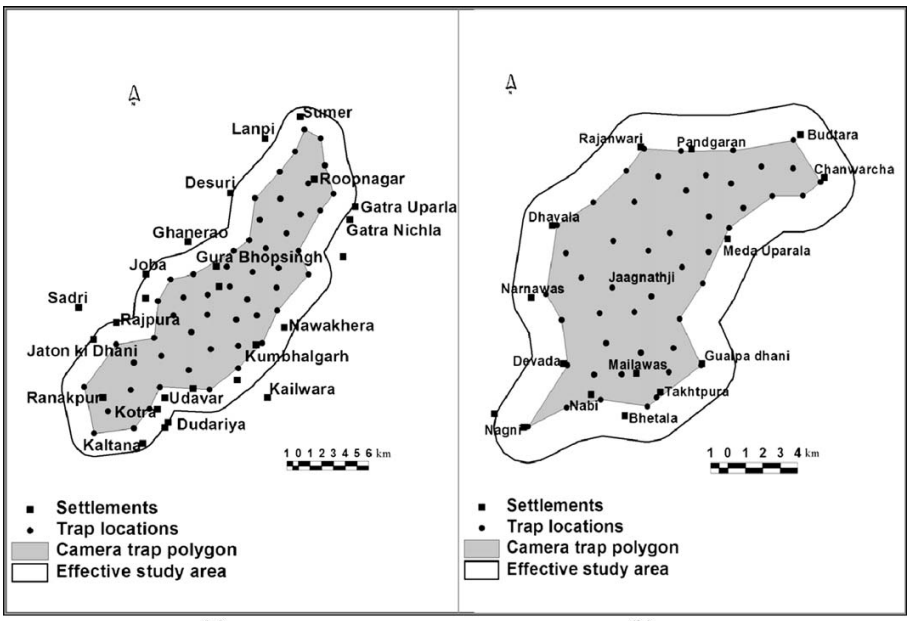


FIG. 1.—Rajasthan state map showing the 2 study locations, Kumbhalgarh and Esrana, with a map of India in the inset.

MATERIALS AND METHODS

Study areas.—This study was conducted between December 2007 and May 2008 at Kumbhalgarh and Esrana sites in the state of Rajasthan in northwestern India (Fig. 1). The selection of these 2 sites was based predominantly on the differences in proportion of hilly terrain, land-use regimes, and logistical feasibility of conducting camera-trap surveys. Unlike Kumbhalgarh, which is a legally designated wildlife reserve, the Esrana Forest Range is managed for multiple land uses. The State Forestry Department managed the entire Kumbhalgarh area but had authority only on a part of Esrana, with remaining hyena habitat being under other public authority or in private holdings.

Kumbhalgarh.—A study area of 307 km^2 was selected within Kumbhalgarh Wildlife Sanctuary (25°00′–25°30′N, 73°15′–73°45′E), which covers a total area of 610 km^2, in the Aravalli hill range of northwestern India. The region receives an average annual rainfall of 730 mm, and temperatures can be as low as 2°C in December–January to as high as 46°C in May–June. The altitudinal gradient is from 288 to 1,215 m. Such dry-deciduous forest tree species as *Anogeissus pendula*, *Anogeissus latifolia*, *Boswellia serrata*, *Butea monosperma*, and *Acacia senegal* are common. The area supports several other carnivores, including leopard (*Panthera pardus*), jungle cat (*Felis chaus*), wolf (*Canis lupus*), jackal (*Canis aureus*),



(a) (b)

Fig. 2.—Map of a) Kumbhalgarh study area and b) Esrana study area in western Rajasthan, India. Camera-trap locations are indicated by dots within the camera trap polygon area (gray). Human settlements are indicated within the camera-trap polygon area and outer strip (white), which represents the effective sampling area.

and sloth bear (*Melursus ursinus*), and the wild ungulate species nilgai (*Boselaphus tragocamelus*), chinkara (*Gazella bennettii*), four-horned antelope (*Tetracerus quadricornis*), and sambar (*Rusa unicolor*—Robbins et al. 2007). Although Kumbhalgarh is legally protected, to some extent human settlements on its peripheries do illegally impact the habitat by grazing livestock, lopping fodder or firewood, and collecting fruits of *Diospyros melanoxylon* and *Madhuca longifolia*.

Esrana.—The 2nd study area of 218 km^2 was located in Jalore District of southwestern Rajasthan around the Esrana Forest Range (25°′–25°30′N, 73°15′–73°45′E). The region receives an annual average rainfall of 300 mm and has an altitudinal gradient of 160–835 m. Temperature can be as low as 1°C in January to as high as 46°C in June. The Esrana site supports low vegetation cover composed chiefly of such xeromorphic species as *Euphorbia* sp., *Salvadora* sp., *Senna alexandrina*, *Zizyphus nummularia*, *Calligonum polygonoides*, and *Aristida* sp. (Bhandari 1990). With the exception of *M. ursinus*, all carnivore species found in Kumbhalgarh also are found in this region, together with the desert cat (*Felis silvestris ornata*) and desert fox (*Vulpes vulpes pusilla*). Among wild ungulates only *G. bennettii* and *B. tragocamelus* are present. Wild pig (*Sus scrofa*) also is common.

Field-survey design.—We conducted an initial field survey covering more than 500 km^2 at each of the sites, based on responses to a preliminary questionnaire survey of expert opinions (n = 30) on hyena status across the region. Using observed encounter rate of hyena tracks, scats, and dens, we identified 52 camera-trap locations in Kumbhalgarh and 48 in Esrana (Fig. 2).

Camera traps were spaced ~2.2 km apart in an irregular configuration to effect saturation sampling and maximize the probability of encountering all individuals (Karanth and Nichols 1998, 2002b). This trapping design was based on striped hyena movement and home-range data reported from sites in East Africa (Kruuk 1976; Wagner 2006). To increase photo-capture rates and keep hyenas in position to get clear photos we baited trap sites with putrid meat.

A total of 36 passive infrared camera traps (MC2-GV STEALTHCAM, Good Sportsman Marketing, LLC, Grand Prairie, Texas) were used. Each trap consisted of 2 cameras positioned 6–7 m apart to photograph both flanks of passing hyenas (Karanth and Nichols 1998). The study areas were divided into 3 trapping blocks that were successively trapped under ''survey design 4'' (Karanth and Nichols 2002b). The traps were stationed at each location for 12 successive nights before being moved to the next block, thus enabling the allocation of each photo-capture event to 1 of the 12 sampling periods (Karanth and Nichols 2002b). The total survey efforts were 548 trap nights in Kumbhalgarh and 538 trap nights in Esrana, respectively. All hyena photographs were assigned to specific trap locations and sampling occasions, based on date and time of capture.

We used pelage markings on hind limbs and forelimbs to identify individual hyenas from camera-trap photos (Fig. 3). Stripe patterns on hind limbs were most variable and useful, followed by patterns on the forelimb. In some cases we made additional use of such other conspicuous features as notches on the ear pinnae.

Capture–recapture analyses.—Because of high camera failure rates most photographs were obtained from only 1 of



FIG. 3.—Camera-trap photographs showing variability in hind-limb and forelimb stripe patterns of hyenas.

the 2 paired cameras, thereby yielding several single-flank photo-captures and leading to loss of some data. Therefore, capture-history matrices for individual hyenas could be constructed only by using photos from a flank that provided higher number of captures.

The survey duration was 36 days at each site, and thus kept short in relation to expected demographic turnover rates to provide reasonable assurance that the assumption of ''demographic closure'' was met. We used the closure test (Z) to test the null hypothesis of population closure (Karanth and Nichols 2002b; Otis et al. 1978; Williams et al. 2002).

A buffer width estimated at half the home-range diameter was added to the polygon formed by traps (Fig. 2) to estimate the effective area sampled by camera traps, after considering geographic closure (Karanth and Nichols 1998; Wilson and Anderson 1985). The mean maximum distance moved by hyenas photo-captured more than once (Wilson and Anderson 1985) was used as a surrogate for average home-range diameter.

Closed capture–recapture analytic models implemented in standard software CAPTURE 2.1 (Rexstad and Burnham 1991) were used to analyze capture history data. We assessed the goodness-of-fit tests (χ^2) and the overall discriminant function test to compare models M (o), M (h), and M (b) implemented in CAPTURE 2.1 (Otis et al. 1978) and selected the appropriate model to derive abundance estimates for hyenas in the 2 study areas. Model M (o) assumes an equal capture probability for all individuals. Model M (h) assumes a differing capture probability for each individual. Model M (b) allows for variations in capture probabilities due to behavioral responses (Karanth and Nichols 2002b; Otis et al. 1978; Williams et al. 2002).

Identification of hyena refugia from topographic features.—Availability of striped hyena refugia in the form of den sites was observed to be closely associated with hilly terrain, which is unattractive to grazing livestock and human uses. Thus, a slope-based landform classification map used to identify steep terrain was created. We used the extension Topographical Position Index in geographic information system program ArcView (Jenness 2006; Fig. 4) using elevation grids from an SRTM 3 hole-filled Digital Elevation Model (Jarvis et al. 2006) for the 2 study areas. The number of pixels in each topographic gradient category was used to calculate the areas under flat or hilly terrain.

Estimates of livestock abundance.—We obtained livestock census data for the year 2007–2008 from official local government records in Jalore, Ahore, Kailwara, and Desuri Tehsils to obtain numbers of bovids for each village within the effectively sampled areas at both study sites (Appendixes I and II).

RESULTS

Abundance of hyenas at Kumbhalgarh and Esrana.—We obtained a total of 33 hyena photographic captures at Kumbhalgarh and 16 captures at Esrana, from which we identified 15 and 8 individual hyenas (M_{t+1}), respectively, and constructed standard capture–recapture matrices (Otis et al. 1978; Williams et al. 2002). The demographic closure test ($Z = -1.289$, $P = 0.10$) in program CAPTURE 2.1 supported the assumption that the Kumbhalgarh hyena population was closed during the survey duration. The goodness-of-fit test showed that the null model M (o) provided a better fit to data relative to M (b) ($\chi^2_1 = 0.91$, $P = 0.34$). The goodness-of-fit test supported model M (h) versus any alternate model ($\chi^2_{11} = 4.81$, $P = 0.94$). Similarly, model M (b) was supported versus any alternate model ($\chi^2_{18} = 15.72$, $P = 0.61$). Thus, M (o), M (b), and M (h) emerged as 3 candidate models that fit these capture data well. Because of small sample sizes model M (o) versus model M (h) could not be tested.

During field surveys, on several occasions hyena track patterns indicated some degree of trap-avoidance behavior. Thus, field observations, together with model selection test results, showed that the null hypothesis of the trap response model, M (b), could not be rejected. Also the assumption of demographic closure was not violated ($Z = -1.289$, $P = 0.10$), which provided support for the M (b) model (Otis et al. 1978). Furthermore, considering that hyenas are carnivores with a well-defined territorial spatial organization (Wagner 2006) and with some behavioral responses to camera trapping, we did not consider the M (o) model because of its lack of robustness to violations of underlying assumptions (Karanth

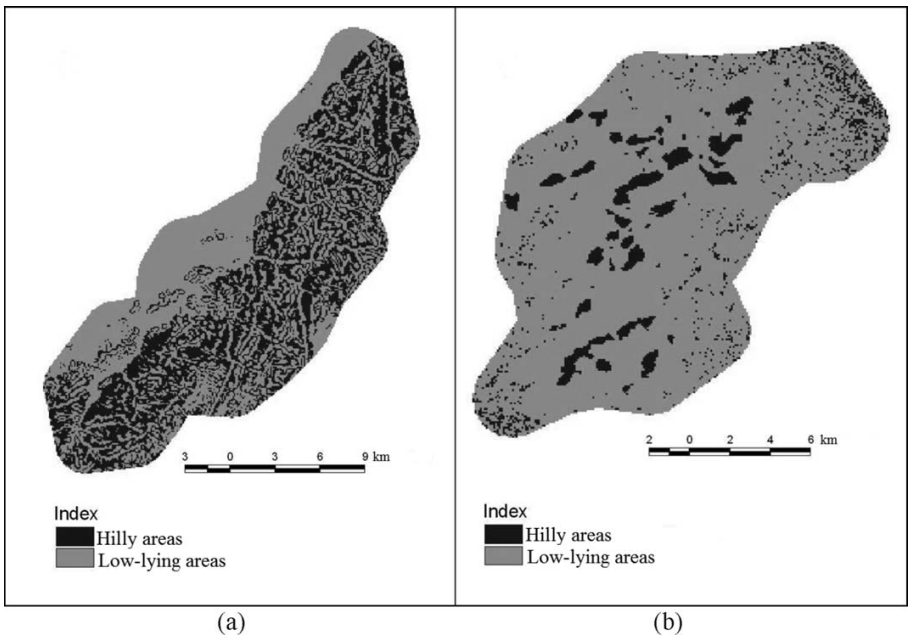


Fig. 4.—Slope-based landform classification map showing hilly areas and low-lying areas (flat land) at a) Kumbhalgarh and b) Esrana sites in India.

and Nichols 1998; Williams et al. 2002). Overall, based on ecological considerations, field observations, and model selection test scores above, the trap-response model M (b) was selected as a plausible model for estimating capture probabilities and abundance. Using the M (b) model, the estimated per sample capture probability ($\hat{p}$) was 0.10 with an estimated recapture probability ($\hat{c}$) of 0.03, leading to an estimated abundance ($\hat{N}$) of 20 hyenas ($SE = 7.85$) for this site.

The closure test in program CAPTURE 2.1 for Esrana ($Z = 6.00$, $P = 1.00$) could not converge in the absence of any recaptures (based on single-flank comparisons). Because of our sparse data set, with no recaptures, model selection tests based on the discriminant functional analysis scores generated by CAPTURE 2.1 could not be used. Instead, we could select the model based only on ecological and statistical considerations. Thus, the removal model M (b) was selected for estimation of capture probabilities and abundance (Karanth and Nichols 2002b). The M (b) model is a maximum-likelihood estimator that is valid even when no individuals are recaptured (Flickinger and Nichols 1990; Karanth and Nichols 2002a) because the estimation of the parameters N and p is independent of the estimation of c (Otis et al. 1978). Using the M (b) removal model, the estimate of per sample capture probability ($\hat{p}$) was 0.27, resulting in an estimate of abundance ($\hat{N}$) of 8 hyenas ($SE = 0.55$) for the Esrana site. The higher estimate of $\hat{p}$ in Esrana relative to Kumbhalgarh resulted in reducing the variance for parameter N.

Sampled area and hyena density estimation.—In Kumbhalgarh the buffer width was set at half the mean-maximum distance between 2 captures observed for each individual hyena. Because no recaptures were obtained in Esrana, we used the same buffer width as at Kumbhalgarh. We assumed that differences in average home-range sizes between the 2 areas were unlikely to result in major differences in estimated buffer width using the mean–maximum distance moved approach. The trap polygon measured 165 km^2 in Kumbhalgarh and 110 km^2 in Esrana. A buffer strip width ($\hat{W}$) of 1.95 km with an SE of 0.15 km was estimated for Kumbhalgarh using the half mean–maximum distance moved approach described in Karanth and Nichols (1998, 2002b). This area was added to the trap polygons for both study sites, deriving an effective sampled area ($\hat{A}$) of 307 km^2 for Kumbhalgarh and 218 km^2 for Esrana. The estimated population size ($\hat{N}$) was divided by the effective sampled area ($\hat{A}$) to derive hyena density estimates ($\hat{D}$). The estimated density for Kumbhalgarh was estimated at 6.5 hyenas/100 km^2 $\pm$ 2.6 SE, and that for Esrana 3.67 $\pm$ 0.3 hyenas/100 km^2.

Assessment of potential hyena refugia.—Kumbhalgarh had 85% of its area (261 km^2) in the hilly area category, and Esrana had only 35% such area (76 km^2), with the remainder being flat land based on the number of pixels in either of the topographic categories (Fig. 4). Thus, Kumbhalgarh offered more potential refugia for supporting hyenas.

Availability of livestock.—The total livestock population size in the Kumbhalgarh study site was 22,304 animals, with an estimated density of 73 animals/km^2 for the total sampled area of 307 km^2. The livestock numbers in Esrana were much higher at 67,842 animals, resulting in a density of 311 animals/km^2.

We further examined additional ecological factors that could influence carcass availability temporally across seasons. Most livestock in Esrana were owned by a pastoral community

known as Raika. This community and other nonspecialized livestock herders in Rajasthan migrate annually out of the area (Agarwal 1991; Robbins 1998) for 6–8 months of the year starting in October–November (Agarwal 1991). During this period only livestock holders with <25–30 animals remain (Agarwal 1991). Therefore, it is likely that the number of livestock carcasses available to hyenas in Esrana for a major part of the year decreases dramatically. Approximately 50% of cattle and almost the entire sheep population of Esrana had migrated by December 2007 when the study was conducted (P. Siana, Siana Safari and Camps, pers. comm.), but <4,000 sheep migrate out of the Kumbhalgarh study site annually (H. Singh, Lokhit Pashu-Palak Sansthan, pers comm.). After accounting for this migratory trend, we find that Esrana still continues to have much higher livestock density (140 animals/km^2) when compared to Kumbhalgarh (60 animals/km^2).

Discussion

We attempted to identify the key ecological determinants of striped hyena densities in an arid landscape of India. Unexpectedly, this large carnivore appears to survive at high densities even in these human-dominated landscapes compared to elsewhere across its range (Kruuk 1976; Wagner 2006). We expected hyena densities to vary positively with food availability in the form of livestock carcasses and with increased availability of refugia in the form of hilly, steep terrain. Hence, we hypothesized that hyena densities would be higher at the site with greater availability of livestock and having a greater proportion of area with hyena denning refugia.

Livestock densities from the 2 study sites differed remarkably, with Esrana having more than 3 times the total livestock density at Kumbhalgarh. Contrary to our prediction, hyena densities were higher at Kumbhalgarh. However, absolute livestock numbers were very high in both sites. Based on dietary studies on the spotted hyena that indicate that the species requires approximately 946 g of dry meat/day (Nagy et al. 1999), the striped hyena likely requires access to a similar quantity of dietary resources, which are easily met by the high rate of livestock mortality and, consequently, carcasses in this region. Therefore, it is likely that availability of carcasses exceeded requirements of hyenas at both sites, with differences in livestock densities not being a critical determinant of differences in hyena densities at the 2 sites. Thus, we focused on testing the importance of the other predicted limiting resource, the extent of hyena refugia available at each study site. Kumbhalgarh offered a substantially larger area of potential refugia from humans or feral dogs. This finding was consistent with our prediction.

Overall, persistence of hyenas in this arid region appears to be a function of availability of disturbance-free denning refugia in hilly terrain and abundant availability of livestock carcasses. Our study also supports the general speculation (Karanth et al. 2009, 2010) that social attitudes such as no consumption of meat from livestock carcasses and relatively higher tolerance for wildlife presence characteristic of this region contributes substantially to persistence of hyenas in these human-dominated landscapes compared to elsewhere across the range of this species.

Our findings above should be viewed cautiously as preliminary because of some limitations of our study. One constraint was the poor performance of inexpensive cameras, leading to many single-flank pictures and slow camera response time leading to loss of some pictures; these limitations resulted in sparse data sets with low recapture rates. Nonetheless, we provided evidence that hyena abundances and densities differed between the 2 sites, with both greater at Kumbhalgarh.

Another important methodological issue (Soisalo and Cavalcanti 2006) is our determination of the effective sampled area using the half mean–maximum distance moved buffer width (Karanth and Nichols 2002b). Further, our use of buffer width estimated from Kumbhalgarh hyenas with potentially smaller home ranges for the Esrana population could have caused an underestimate of the effective sampled area in Esrana, resulting in an overestimate of hyena densities. However, this possibility does not negate the main finding that hyena densities were relatively higher at Kumbhalgarh. In addition, because of our sparse data sets, we could not use recent advances in capture–recapture models that incorporate spatial data on capture locations to estimate densities directly, avoiding the buffer width approach altogether (Borchers and Efford 2008; Royle et al. 2009a, 2009b). We hope to address both these practical and analytical constraints in our future studies.

We concede that the strength of inference from our study is relatively low when viewed in the framework of classical experimental design because of its observational nature (as most field studies of large carnivores tend to be). However, we believe our results have provided increased evidence for factors important to the ecology of striped hyenas because our study was based on a priori, alternative hypotheses and explicit predictions that were tested with carefully gathered survey data on hyena densities from rigorous capture–recapture sampling. We argue that our work is better viewed in the context of evidence under the likelihood-based approach (Bolker 2008; Burnham and Anderson 2002; Hilborn and Mangel 1997; Royle and Dorazio 2008). Overall, despite limitations mentioned earlier, our study contributed to the understanding of the population status of striped hyenas in human-dominated landscapes of India and of ecological and management factors that determine it. The photographic capture–recapture method we used generated more precise estimates of hyena densities than earlier surveys. Methodological refinements we suggest above can help in rigorous assessments of ecology and conservation needs of striped hyenas across their entire range.

Our results suggest that striped hyenas, despite the ability to adapt to human-modified landscapes, require natural habitats free of anthropogenic disturbances to serve as refugia for source populations. The strikingly higher densities of hyenas in the protected reserve of Kumbhalgarh, with some degree of

regulation of human uses, support this conclusion. Thus, creation of more such protected refugia for hyenas across arid regions is a key conservation need identified from this study.

Acknowledgments

This research was made possible with funding from Save the Tiger Fund of the National Fish and Wildlife Foundation, Washington, D.C. We also acknowledge support from Wildlife Conservation Society–India Program, Centre for Wildlife Studies, Bangalore, and the National Centre for Biological Sciences (TIFR), Bangalore, for their logistic and administrative support. We are grateful to the Chief Wildlife Warden, Rajasthan, R. N. Mehrotra, and all the staff of Rajasthan Forest Department for their support at our study sites. We also thank other important contributors for their assistance: V. Srinivas for help with geographic information system analysis, D. Jathanna for help with capture–recapture analyses, R. Mogaroy and N. Patil for help with photo identification of hyenas, and N. Samba Kumar and A. Kumar for their comments, suggestions, and guidance. We also thank P. Singh, Siana and H. Singh, and Lokhit Pashu-Palak Sansthan for providing the livestock migration figures.

Literature Cited

Agarwal, A. 1991. The grass is greener on the other side. Report submitted to the International Institute for Environment and Development, London, United Kingdom.

Bagchi, S., S. P. Goyal, and K. Sankar. 2003. Prey abundance and prey selection by tigers (*Panthera tigris*) in a semi-arid, dry deciduous forest in western India. Journal of Zoology (London) 260:285–290.

Baqri, Q. H., and P. L. Kankane. 2001. Desert: Thar. Pp. 94–109 in Ecosystems of India (J. R. B. Alfred, A. K. Das, and A. K. Sanyal, eds.). ENVIS–Zoological Survey of India, Kolkata, India.

Bhandari, M. M. 1990. Flora of the Indian Desert. Revised ed. MPS Repros, Jodhpur, India.

Bolker, B. M. 2008. Ecological models and data in R. Princeton University Press, Princeton, New Jersey.

Borchers, D. L., and M. G. Efford. 2008. Spatially explicit maximum likelihood methods for capture–recapture studies. Biometrics 64:377–385.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information theoretic approach. 2nd ed. Springer-Verlag, New York.

Carbone, C., and J. L. Gittleman. 2002. A common rule for the scaling of carnivore density. Science 295:2273–2276.

Chhangani, A. K. 2009. Present status of vultures in the Great Indian Thar Desert. Pp. 65–82 in Faunal ecology and conservation of the Great Indian Desert (C. Sivaperuman, Q. H. Baqri, G. Ramaswamy, and M. Naseema, eds.). Springer-Verlag, Berlin, Germany.

Flickinger, E. L., and J. D. Nichols. 1990. Small mammal populations at hazardous waste disposal sites near Houston, Texas, USA. Environmental Pollution 65:169–180.

Hilborn, R., and M. Mangel. 1997. The ecological detective: confronting models with data. Princeton University Press, Princeton, New Jersey.

Hofer, H. 1998. Species accounts: striped hyena *Hyaena* (Hyaena) *hyaena* (Linnaeus, 1758). Pp. 21–26 in Hyaenas: status survey and conservation action plan (G. Mills and H. Hofer, comps.). IUCN/SSC Hyaena Specialist Group, IUCN, Gland, Switzerland, and Cambridge, United Kingdom.

Jarvis, A., H. I. Reuter, A. Nelson, and E. Guevara. 2006. Hole-filled seamless SRTM data v3. International Centre for Tropical Agriculture (CIAT), Cali, Colombia.

Jenness, J. 2006. Topographic position index (tpi_jen.avx) (ArcView 3.x, v.1.2. extension). Jenness Enterprises, Flagstaff, Arizona.

Kankane, P. L. 2009. Revival prospects of larger mammals in the Thar Desert of Rajasthan. Pp. 209–219 in Faunal ecology and conservation of the Great Indian Desert (C. Sivaperuman, Q. H. Baqri, G. Ramaswamy, and M. Naseema, eds.). Springer-Verlag, Berlin, Germany.

Karanth, K. K., J. D. Nichols, J. E. Hines, K. U. Karanth, and N. L. Christensen. 2009. Patterns and determinants of mammal species occurrence in India. Journal of Applied Ecology 46:1189–1200.

Karanth, K. K., J. D. Nichols, K. U. Karanth, J. E. Hines, and N. L. Christensen. 2010. The shrinking ark: patterns of large mammal extinctions in India. Proceedings of the Royal Society of London, B. Biological Sciences 277:1971–1979.

Karanth, K. U., and J. D. Nichols. 1998. Estimation of tiger densities in India using photographic captures and recaptures. Ecology 79:2852–2862.

Karanth, K. U., and J. D. Nichols. 2002a. Ecological status and conservation of tigers in India. Final Technical Report to the Division of International Conservation, United States Fish and Wildlife Service, Washington, D.C., and Wildlife Conservation Society, New York. Centre for Wildlife Studies, Bangalore, India.

Karanth, K. U., and J. D. Nichols (eds.). 2002b. Monitoring tigers and their prey: a manual for researchers, managers and conservationists in tropical Asia. Centre for Wildlife Studies, Bangalore, India.

Karanth, K. U., J. D. Nichols, N. S. Kumar, W. A. Link, and J. E. Hines. 2004. Tigers and their prey: predicting carnivore densities from prey abundance. Proceedings of the National Academy of Sciences 101:4854–4858.

Karanth, K. U., et al. 2003. Science deficiency in conservation practice: the monitoring of tiger populations in India. Animal Conservation 6:141–146.

Kruuk, H. 1976. Feeding and social behaviour of the striped hyaena (*Hyaena vulgaris* Desmarest). African Journal of Ecology 14:91–111.

Leakey, L. N., et al. 1999. Diet of striped hyaena in northern Kenya. African Journal of Ecology 37:314–326.

Mendelssohn, H., and Y. Yom-Tov. 1999. Fauna Palaestina: Mammalia of Israel. Israel Academy of Sciences and Humanities, Keterpress Enterprises, Jerusalem, Israel.

Nagy, K. A., I. A. Girard, and T. K. Brown. 1999. Energetics of free ranging mammals, reptiles and birds. Annual Review of Nutrition 19:247–277.

Otis, D. L., K. P. Burnham, G. C. White, and D. R. Anderson. 1978. Statistical inference from capture data on closed animal populations. Wildlife Monographs 62:1–135.

Prater, S. H. 1948. The book of Indian animals. 1st ed. Oxford University Press, New Delhi, India.

Rahmani, A. R., and R. G. Soni. 1997. Avifaunal changes in the Indian Thar Desert. Journal of Arid Environments 36:687–703.

Reiger, I. 1981. *Hyaena hyaena*. Mammalian Species 150:1–5.

Rexstad, E., and K. Burnham. 1991. User's guide for interactive program CAPTURE. Abundance estimation of closed animal populations. Colorado State University, Fort Collins.

Robbins, P. 1998. Nomadization in Rajasthan, India: migration, institutions, and economy. Human Ecology 26:87–112.

Robbins, P. F., A. K. Chhangani, J. Rice, E. Trigosa, and S. M. Mohnot. 2007. Enforcement authority and vegetation change at Kumbhalgarh Wildlife Sanctuary, Rajasthan, India. Environmental Management 40:365–378.

ROYLE, J. A., AND R. M. DORAZIO. 2008. Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations, and communities. Academic Press, San Diego, California.

ROYLE, J. A., K. U. KARANTH, A. M. GOPALASWAMY, AND N. S. KUMAR. 2009a. Bayesian inference in camera trapping studies for a class of spatial capture–recapture models. Ecology 90: 3233–3244.

ROYLE, J. A., J. D. NICHOLS, K. U. KARANTH, AND A. M. GOPALASWAMY. 2009b. A hierarchical model for estimating density in camera-trap studies. Journal of Applied Ecology 46:118–127.

SHARMA, K. K., AND S. P. MEHRA. 2009. The Thar of Rajasthan (India): ecology and conservation of a desert ecosystem. Pp. 1–10 in Faunal ecology and conservation of the Great Indian Desert (C. Sivaperuman, Q. H. Baqri, G. Ramaswamy, and M. Naseema, eds.). Springer-Verlag, Berlin, Germany.

SOISALO, M. K., AND S. M. C. CAVALCANTI. 2006. Estimating the density of a jaguar population in the Brazilian Pantanal using camera-traps and capture–recapture sampling in combination with GPS radio-telemetry. Biological Conservation 129:487–496.

WAGNER, A. P. 2006. Behavioral ecology of the striped hyaena (*Hyaena hyaena*). Ph.D. dissertation, Montana State University, Bozeman.

WILLIAMS, B. K., J. D. NICHOLS, AND M. J. CONROY. 2002. Analysis and management of animal populations. Academic Press, San Diego, California.

WILSON, K. R., AND D. R. ANDERSON. 1985. Evaluation of two density estimators of small mammal population size. Journal of Mammalogy 66:13–21.

Submitted 29 April 2009. Accepted 20 April 2010.

Associate Editor was Rodrigo A. Medellín.

APPENDIX I

Livestock figures for the effective study area of Kumbhalgarh, India, 2007–2008 (source: Office of the Tehsildar of Kailwara Tehsil and Desuri Tehsil).

Settlement	Cattle	Buffalo	Sheep and goats
Sumer	324	251	878
Lanpi	90	75	3,475
Desuri	1,736	1056	4,293
Joba	469	118	782
Gura Bhopsingh	418	180	680
Rajpura	203	100	900
Jaton ki Dhani	140	90	800
Ranakpur Temple	0	0	0
Roopnagar	48	15	298
Borda ki Bhagal	65	99	114
Kumbhalgarh	99	103	128
Kotra Pokharia	257	101	274
Boitra[a]	19	5	35
Nadia[a]	5	0	10
Miyawa[a]	2	11	13
Aret ki Bhagal	154	226	100
Mandigarh	183	262	1,111
Garasiya Colony[a]	210	58	461
Udavar	320	142	197
Kharni Tankri[a]	27	16	78
Total	4,769	2,908	14,627

[a] Settlements located within Kumbhalgarh Wildlife Sanctuary for which livestock data were collected by Rajasthan Forest Department personnel.

APPENDIX II

Livestock figures for the effective study area of the Esrana site, India, 2007–2008 (source: Office of the tehsildar of Jalore Tehsil and Ahore Tehsil).

Settlement	Cattle	Buffalo	Sheep	Goats
Narnawas	274	602	1,077	1,844
Naya Narnawas	89	340	440	579
Dhavala	204	559	731	1,884
Digaon	129	973	264	684
Nagni	94	451	471	692
Devada	115	279	66	419
Nabi	73	138	2,418	1,018
Bhetala	52	154	1,158	580
Mailawas	154	521	2,959	1,237
Takhatpura	717	803	604	465
Meda Uparla	512	715	18,466	5,786
Meda Nichala	1,519	1,544	1,763	261
Rajanwari	51	63	912	834
Pandgaran	70	295	1,582	1,145
Chanwarcha	318	159	764	1,311
Chipparwara	99	348	696	778
Budtara	160	126	548	706
Total	4,630	8,070	34,919	20,223

Appendix F

Tigers of trails: occupancy modeling for cluster sampling

Ecological Applications, 20(5), 2010, pp. 1456–1466
© 2010 by the Ecological Society of America

Tigers on trails: occupancy modeling for cluster sampling

J. E. Hines,[1,4] J. D. Nichols,[1] J. A. Royle,[1] D. I. MacKenzie,[2] A. M. Gopalaswamy,[3] N. Samba Kumar,[3] and K. U. Karanth[3]

[1]*United States Geological Survey, Patuxent Wildlife Research Center, Laurel, Maryland 20708 USA*
[2]*Proteus Wildlife Research Consultants, P.O. Box 5193, Dunedin, New Zealand*
[3]*Wildlife Conservation Society—India Program, Centre for Wildlife Studies, Bangalore, Karnataka-560042 India*

Abstract. Occupancy modeling focuses on inference about the distribution of organisms over space, using temporal or spatial replication to allow inference about the detection process. Inference based on spatial replication strictly requires that replicates be selected randomly and with replacement, but the importance of these design requirements is not well understood. This paper focuses on an increasingly popular sampling design based on spatial replicates that are not selected randomly and that are expected to exhibit Markovian dependence. We develop two new occupancy models for data collected under this sort of design, one based on an underlying Markov model for spatial dependence and the other based on a trap response model with Markovian detections. We then simulated data under the model for Markovian spatial dependence and fit the data to standard occupancy models and to the two new models. Bias of occupancy estimates was substantial for the standard models, smaller for the new trap response model, and negligible for the new spatial process model. We also fit these models to data from a large-scale tiger occupancy survey recently conducted in Karnataka State, southwestern India. In addition to providing evidence of a positive relationship between tiger occupancy and habitat, model selection statistics and estimates strongly supported the use of the model with Markovian spatial dependence. This new model provides another tool for the decomposition of the detection process, which is sometimes needed for proper estimation and which may also permit interesting biological inferences. In addition to designs employing spatial replication, we note the likely existence of temporal Markovian dependence in many designs using temporal replication. The models developed here will be useful either directly, or with minor extensions, for these designs as well. We believe that these new models represent important additions to the suite of modeling tools now available for occupancy estimation in conservation monitoring. More generally, this work represents a contribution to the topic of cluster sampling for situations in which there is a need for specific modeling (e.g., reflecting dependence) for the distribution of the variable(s) of interest among subunits.

Key words: cluster sampling; detection probability; India; Markov model; occupancy modeling; Panthera tigris; *spatial dependence; spatial replication; tigers; trap response model.*

Introduction

Large-scale occupancy surveys and monitoring programs are now used throughout the world to inform conservation decisions. Most occupancy modeling requires some sort of replication in order to obtain the information needed to estimate detection probability and, thereby, probabilistically separate true absence from presence and nondetection (MacKenzie et al. 2006, Royle and Dorazio 2008). The usual sampling situation involves multiple visits to each sample unit during some period of time over which the units are assumed to be closed to changes in true occupancy. In some cases it is possible to substitute spatial replication for temporal replication in order to obtain this information about detection probability. Specifically, multiple survey sites or locations are selected from each sample unit randomly and with replacement and are then surveyed a single time, usually on the same day (MacKenzie et al. 2006). Such a design permits estimation of occupancy at the level of the sample unit (not at the level of the specific sites or locations within each unit). When the species (or sign of the species) occupies a sampling unit, but is not present at all sites within the sampling unit, detection probability consists of two components: (1) Pr(present at survey site) and (2) Pr(detection | present at survey site). Resulting estimates of detection probability in such designs correspond to each specific survey site and are reasonable estimates of the product of these components.

Sometimes surveys employ spatial replication but do not follow the recommendation to sample randomly and with replacement. Such survey designs typically result from logistical constraints and represent a compromise

Manuscript received 24 February 2009; revised 28 August 2009; accepted 1 September 2009; final version received 8 October 2009. Corresponding Editor: J. J. Millspaugh.
[4] E-mail: jhines@usgs.gov

between statistical requirements and practical requirements of field surveys. It is unclear how deviations from random spatial sampling with replacement affect estimates of occupancy. In this paper we focus on a specific field design that is being used in a large-scale occupancy survey of tigers (*Panthera tigris*) in southwestern India (K. U. Karanth, N. S. Kumar, A. Gopalaswamy, and V. Srinivas, *unpublished manuscript*). This type of survey design has been adopted elsewhere in southeast Asia by several conservation organizations for use with carnivores and elephants, is currently being used for carnivore surveys in Africa, and is being considered for surveys of carnivores in the New World as well. The design uses spatial replication without the requisite random selection and replacement and is expected to produce detection data that are correlated on adjacent sample sites (replicates). We first develop a new model that deals explicitly with data resulting from this type of spatially replicated design. We also develop a new "trap response" occupancy model (see MacKenzie et al. 2006) for possible use with such data. We then use computer simulation to investigate the consequences of this kind of spatial design for estimates based on (1) standard occupancy models that ignore spatial dependence, (2) the new trap response model, and (3) the new spatial Markov process model developed for this design. Finally, we apply all three classes of model to initial data from a field survey of tigers in India (K. U. Karanth, N. S. Kumar, A. Gopalaswamy, and V. Srinivas, *unpublished manuscript*), focusing on variation among resulting estimates and on the results of model selection.

Survey Designs with Correlated Spatial Replication

Field surveys

The field survey that motivated this work was developed for tigers in the 22 000-km^2 Malnad-Mysore Tiger Landscape in Karnataka State, southwestern India. This and similar designs designate as sampling units large geographic grid cells at a scale appropriate to the study organism, depending on the biology of the species. For example, for tigers in southwestern India, the cell size was set based on expected maximum home range size of ~200 km^2 (Karanth and Sunquist 2000). Tigers are known to use forest roads and trails as travel routes and to mark them intensively with tracks, scent, and scats (Karanth and Sunquist 2000). Search for tiger signs along such trails increases detection probabilities far above those expected based on random sampling. Within each cell the design ensured that each survey team passed through a point randomly chosen before the survey. The trail was then walked by three trained surveyors looking for signs of the species of interest (e.g., tiger scat and tracks). The distance covered was subdivided into segments of equal length (e.g., 1 km) that were then treated as geographic replicates in the occupancy analysis. We initially viewed the chief virtue of this design to be logistical feasibility (but see *Discussion*).

Recommendations for the use of geographic replicates in occupancy surveys include sampling randomly selected replicates with replacement (MacKenzie et al. 2006). Logistical considerations sometimes cause biologists to select designs, such as that used in the tiger survey, that do not sample the replicates (segments) with replacement. The reasons for preferring sampling with replacement are not likely to be important for the large, wide-ranging, highly mobile species to which this type of survey design is applied. High mobility insures that all possible detection histories have some nonnegligible a priori probability of occurrence. This is not the case for spatial replication of sedentary organisms that either inhabit a replicate or not. For such species, if replicate 1 is inhabited and replicate 2 is not, then only detection histories 10 and 00 are possible if sampling is without replacement. This latter situation leads to biased estimates of occupancy and detection probability. Despite the random selection of one of the replicates (e.g., trail segments) in the carnivore survey design, all other replicates are not selected at random (Fig. 1). Trails are selected for surveys because they represent likely travel routes for study species. Because individual animals walk along trails for distances that can exceed segment length, detection of sign on one segment likely will translate to an increased probability of detecting sign on the next segment. In this paper we focus on methods for dealing with data from designs with this type of potential for spatial correlation in detection probability.

This type of sampling design can be thought of generally as a cluster sampling design (e.g., Thompson 2002). Formal inference for such a design requires a model of within-cluster variability (i.e., a description of the process generating variation among subsamples). In the present context, we require a model for variability in species presence among subsamples. The basic null model that has justified previous applications of spatial subsampling (e.g., see Williams et al. 2002:555–573) is that subsample occupancy states are independent and identically distributed Bernoulli outcomes, in which case subsample occupancy becomes confounded with detection probability (as noted in paragraph 1; see also Cam et al. 2002). Random sampling with replacement is simply an attempt to induce this null model. In cluster sampling designs such as that described here, the subsamples are spatially organized, suggesting certain types of models that accommodate spatial dependence. When such models are reasonable, they provide information that can be used to estimate additional parameters describing variation among spatial subsamples within a cluster.

New Markov process for segment occupancy, model $[\psi(.), \theta(.), \theta'(.), p(.)]$

Under the above design, we suspected some degree of positive correlation between detection events on adja-

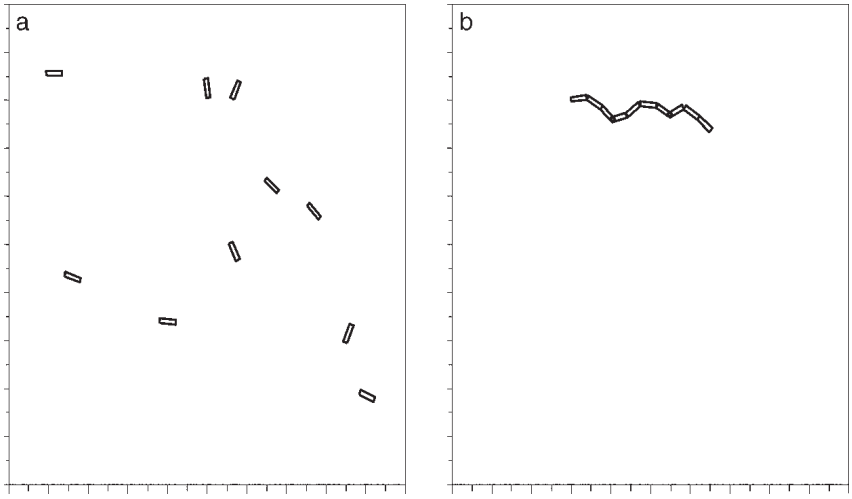


FIG. 1. Two sampling designs employing spatial replication to draw inference about occupancy and detection probabilities: (a) design 1 depicts an example of random sampling; (b) design 2 depicts sampling of segments along a trail, likely producing correlated spatial replicates.

cent sample segments. Specifically, we considered the biologically likely scenario that an individual tiger might travel along a trail for several adjacent 1-km segments. We chose to model this scenario by decomposing the detection process into two components: (1) animal presence at a segment and (2) animal detection, given presence on a segment. We then modeled the component associated with animal presence as a first-order spatial Markov process (e.g., Williams et al. 2002:197–202). Initially, we believed that this model might be so general that parameters would not be identifiable. We thus developed an additional model patterned after the trap response models of capture–recapture. Although we found that inference is indeed possible for the more general model, we present the new trap response model in Appendix A, as we believe that it may be useful in some sampling situations as well.

We assume that we begin the survey at one end of the survey route (trail) for a sample unit and proceed in order along the K total segments of the route. The data resulting from such a survey are detection histories (e.g., based on tiger sign) for each of the s sample units (large cells) selected to survey. Each detection history is simply a vector containing a "0" or a "1" for each of the K elements (segments), depending on whether the species was (1) or was not (0) detected on that segment. For example, consider the detection history for cell j, $h_j = 01011$. There were no detections of tiger sign on segments 1 or 3, but sign was detected on segments 2, 4, and 5.

Define the following parameters of a spatial process model for detection history data obtained under the above design (we drop the subscript j denoting the sample unit): $p = \Pr(\text{detection at a segment} \mid \text{sample unit occupied and species present on segment})$; $\psi = \Pr(\text{sample unit occupied})$; $\theta = \Pr(\text{species present on segment} \mid \text{sample unit occupied and species not present on previous segment})$; $\theta' = \Pr(\text{species present on segment} \mid \text{sample unit occupied and species present on previous segment})$. These parameters can be used to develop a Markov spatial process model that involves an unobservable random variable, whether the species was present or not in the previous segment. The dot notation (dot following a model parameter symbol, e.g., $\psi(.)$), in the model description indicates no variation from segment to segment in a model parameter.

Consider the detection history, $h_j = 01011$. The probability associated with this history under the Markov spatial process model is

$$\Pr(h_j = 01011) = \psi[(1-\theta)\theta + \theta(1-p)\theta'] \times p[(1-\theta')\theta + \theta'(1-p)\theta']p\theta' p.$$

The initial ψ in the above expression corresponds to the event that the species was present in the sample unit. The first set of brackets contains two additive components dealing with the uncertainty about segment 1. One possibility is that the species was present on segment 2 (the probability corresponding to this event is θ), but absent from the previous segment $(1 - \theta)$. The other possibility is that the species was present on segment 2 (θ') and present but undetected $[\theta(1-p)]$ on the previous segment. The p following these bracketed terms indicates detection on segment 2. The second set of brackets contains probabilities associated with the two possible outcomes on segment 3, absence of the species and presence with nondetection. There is no ambiguity associated with segments 4 and 5 (the species was present and detected at each of these segments), so the modeling is simpler.

The probability associated with a detection history of all zeros must include the probability of true absence from surveyed segments and the probability of presence and nondetection. The latter probability in turn depends

on the realization of the spatial Markov process that dictates true presence and absence among segments. Consider a simple design with only two segments (not recommended, but this produces a probability statement that is relatively easy to follow). The probability associated with history $h_j = 00$ can be written as

$$\Pr(h_j = 00) = (1 - \psi) + \psi[\theta(1 - p)(1 - \theta' p) + (1 - \theta)(1 - \theta p)].$$

The first term, $(1 - \psi)$, corresponds to the probability that the sample unit is not occupied. If the sample unit is occupied, then the first of the two main additive terms within the brackets corresponds to the probability that the first segment is occupied, but the species is not detected. The next segment is then either occupied with no detection $[\theta'(1 - p)]$ or not occupied $(1 - \theta')$. Note that $1 - \theta' p = \theta'(1 - p) + (1 - \theta')$. The second main additive term within the large brackets corresponds to the probability that the first segment is not occupied. The second segment then may be occupied with no detection or not occupied.

Every detection history can be modeled in this manner, and the likelihood under this model can be obtained simply as the product of the probabilities corresponding to all detection histories:

$$L(\psi, \theta, \theta', p, \mid h_1, h_2, ..., h_s) = \prod_{j=1}^{s} \Pr(h_j).$$

A general computing expression for $\Pr(h_j)$ is provided in Appendix B.

Estimates under this model can be obtained via maximum likelihood, and the senior author has incorporated this model into program PRESENCE (Hines 2006). This software can be used to obtain estimates under this and related models. In fact, the model structure incorporated into PRESENCE is more general than that described above in that it deals with missing observations. The above model can also be implemented using a Markov chain Monte Carlo (MCMC) approach. A brief description of this approach, together with WinBUGS code for this model, is presented in Supplement 1. Results of a small simulation study comparing results of the likelihood-based and MCMC approaches are provided in Appendix C.

Methods

Simulation study

A simulation study was conducted to evaluate the performance of standard occupancy models and the two new models described above for data collected under the described spatial cluster sampling. The simulation of detection history data proceeded in a straightforward manner and followed the development of model [ψ(.), θ(.), θ'(.), p(.)]. For each sample unit, occupancy was determined as a Bernoulli trial with probability ψ. For those units that were occupied, presence of the species on the first segment of the survey route was determined as a Bernoulli trial with probability θ. If the species was determined to be present on segment 1, then its presence on segment 2 was determined as a Bernoulli random variable, Bern (θ'). If the species was not present on segment 1, then presence on segment 2 was determined as Bern (θ). Species presence and absence at the remaining segments were determined similarly, with probability of presence at segment t depending on presence at segment $t - 1$. Once presence/absence was determined for all K segments of the survey route, the detection process was simulated. At each segment at which the species was present, detection was determined as Bern (p). This procedure led to detection histories for all s sample units.

Specific values were assigned to the above parameters, and detection data were generated. These detection histories were then used in conjunction with four specific occupancy models in order to assess estimator performance in the face of the described spatial sampling with correlation. Specifically, we simulated data from a study of 200 sample units, each sampled by a "trail" consisting of 10 segments. Occupancy at the level of the sample unit was set at $\psi = 0.75$. For occupied sample units, probability of animal presence for segment 1 and for all subsequent segments, t, for which there was no presence on the previous segment, $t - 1$, was set at $\theta = 0.1$. For occupied sample units, probability of animal presence on any segment, t, for which there was presence on the previous segment, $t - 1$, was set at $\theta' = 0.5$. The detection probability for each segment, conditional on presence of animals on the segment, was set at $p_t = p = 0.80$. The detection data were thus generated according to a spatial Markov process and showed a strong positive spatial correlation between successive segments.

In order to provide a set of simulations to serve as a sort of control, we also generated data that corresponded to standard occupancy model assumptions (MacKenzie et al. 2002, 2006, Royle and Dorazio 2008). These simulations were carried out with the same sample sizes and parameter values as above, with the exception that we set $\theta = \theta' = 0.3$. This latter constraint corresponds to the situation in which occupancy of a particular segment is the same regardless of whether the previous segment is occupied (we have removed the Markovian dependence).

These values were used to generate 1000 sets of detection history data corresponding to each of the above spatial processes (Markovian and non-Markovian). These data were then used to estimate occupancy and detection probability under two standard occupancy models that were not designed to account for the possibility of spatial correlation between segments, [ψ(.), p(.)] and [ψ(.), $p(t)$]. The first model assumes constant occupancy among sites and constant detection probability among sites and segments. Note that this model corresponds to the second set of simulations described above. The second model assumes

constant occupancy among sites and constant detection probability among sites, but permits variation in detection probability among segments. We then evaluated the performance of estimators from the new model developed specifically for spatially correlated data, [$\psi(.)$, $\theta(.)$, $\theta'(.)$, $p(.)$]. For comparison, we also fit the new trap response model [$\psi(.)$, $p(.)$, $p'(.)$] developed as an approximation to the process generating the data (Appendix A).

In evaluating the performance of all of these models, we focused on the bias of the occupancy estimator, $\hat{\psi}$, and computed bias as

$$\text{Bias}(\hat{\psi}) = \frac{\sum_{k=1}^{n}(\hat{\psi}_k - \psi)}{n}$$

where $\hat{\psi}_k$ is the estimate obtained for simulated data set k, n is the number of simulations ($n = 1000$), and ψ is the true occupancy value used to generate the data. Relative bias is obtained by dividing bias by the true parameter value, e.g.,

$$\text{RelBias}(\hat{\psi}) = \text{Bias}(\hat{\psi})/\psi.$$

We also evaluated the performance of the estimates of standard errors for various estimators, $\hat{\phi}$. We computed bias of these estimators as

$$\text{Bias}[\widehat{\text{SE}}(\hat{\phi})] \approx \frac{\sum_{k=1}^{n}\widehat{\text{SE}}(\hat{\phi}_k)}{n} - \sqrt{\frac{\sum_{k=1}^{n}(\hat{\phi}_k - \bar{\hat{\phi}})^2}{n-1}}$$

basically computing the difference between the average model-based standard error and the iteration-based estimate based on the 1000 estimates of ϕ.

Large-scale tiger occupancy survey

The spatial distribution survey of tigers in Karnataka State, India (K. U. Karanth, N. S. Kumar, A. Gopalaswamy, and V. Srinivas, *unpublished manuscript*), was conducted between February 2006 and June 2007 across a 22 000-km^2 area, assuming that tiger spatial distribution remained unchanged during this relatively short period. Based on previous studies of tiger density, home range size, and relationship to abundance of ungulate prey (Karanth and Sunquist 2000, Karanth et al. 2004), we assumed an expected maximum home range size of 150 km^2 for tigers. The grid cell size selected was larger than this area, primarily with the goal of eventually linking the occupancy parameter to a measure of tiger abundance using the Royle-Nichols (2003) model.

The survey was carried out over 205 grid cells of 188 km^2 each that coincided with the survey map-grid feature to facilitate fieldwork. The surveyed distance walked within each cell was 40 km if the cell entirely comprised tiger habitat based on land-cover features. This distance was proportionately reduced depending on extent of habitat, and cells with $<10\%$ forest cover were not surveyed, as they were unlikely to shelter tigers. Thus, the number of 1-km replicate segments surveyed per cell ranged from 4 to 42. The tiger signs (tracks, scats) encountered were verified and recorded (K. U. Karanth, N. S. Kumar, A. Gopalaswamy, and V. Srinivas, *unpublished manuscript*).

As noted above, within each sample unit (grid cell) the design ensured that each survey team passed through a point randomly chosen before the survey. However, the starting point of the survey was not necessarily located at the beginning of the trail. The first segment surveyed was typically in the interior of the trail. This design requires a slight modification of the probability structure described here for model $\psi(.)$, $\theta(.)$, $\theta'(.)$, $p(.)$. Specifically, it is not appropriate to model the probability of segment occupancy for the initial segment as θ, because the initial surveyed segment may or may not be preceded by an occupied segment. Instead, we need an expression for the probability that a randomly selected segment from the interior of a trail is occupied. A reasonable expression for this probability is the equilibrium probability of occupancy for a spatial Markov process defined by θ and θ', as given by

$$\frac{\theta}{\theta + (1 - \theta')}. \quad (1)$$

(e.g., see MacKenzie et al. 2006:208). Thus, the probability associated with segment-level occupancy of the initial surveyed segment can be modeled either as θ, when this segment is at the beginning of a trail, or as expression 1 for surveys such as the tiger survey in which the initial surveyed segment is in the trail interior.

We fit six models to the tiger data set. The first three models were parameterized with constant occupancy and three different models of the detection process, [$\psi(.)$, $\theta(.)$, $\theta'(.)$, $p(.)$], [$\psi(.)$, $p(.)$, $p'(.)$], and [$\psi(.)$, $p(.)$]. The additional models reflected the same three models of the detection process, but with occupancy modeled as a linear-logistic function of the number of segments surveyed (because this number reflected the proportion of the cell comprised of tiger habitat, denoted h in model notation), [$\psi(h)$, $\theta(.)$, $\theta'(.)$, $p(.)$], [$\psi(h)$, $p(.)$, $p'(.)$], and [$\psi(h)$, $p(.)$]. All models were implemented in program PRESENCE (Hines 2006), and maximum-likelihood estimates were computed. Akaike's information criterion (AIC) was computed as a model selection statistic, and AIC weights were computed for the six models (Burnham and Anderson 2002). We considered use of AIC_c AIC adjusted for small sample size. However, the issue of just what constitutes sample size in the case of occupancy models has not been resolved, so we opted for the unmodified AIC. Note that the issue of defining sample size is problematic not only for occupancy modeling, but for many other situations as well (Burnham and Anderson 2002:332–333).

For a given model of occupancy (e.g., constant), we expected the model with spatial segment-level depen-

TABLE 1. Simulation results for 1000 sets of detection history data for 200 sites under each of two spatial correlation scenarios.

Parameter estimates by model	True segment occupancy parameter values	
	$\theta = 0.1, \theta' = 0.5$	$\theta = \theta' = 0.3$
$\psi(.), p(.)$		
$\hat{\psi}$	0.5194	0.7532
$\widehat{SE}(\hat{\psi})$	0.0429	0.0351
$\widehat{SD}(\hat{\psi})$	0.0448	0.0363
$\hat{p}$	0.1792	0.2395
$\widehat{SE}(\hat{p})$	0.0148	0.0124
$\widehat{SD}(\hat{p})$	0.0189	0.0128
$\psi(.), p(t)$		
$\hat{\psi}$	0.5178	0.7515
$\widehat{SE}(\hat{\psi})$	0.0429	0.0352
$\widehat{SD}(\hat{\psi})$	0.0447	0.0432
$\hat{p}_5$	0.1855	0.2393
$\widehat{SE}(\hat{p}_5)$	0.0387	0.0349
$\widehat{SD}(\hat{p}_5)$	0.0406	0.0358
$\psi(.), p(.), p'(.)$		
$\hat{\psi}$	0.7039	0.7534
$\widehat{SE}(\hat{\psi})$	0.0835	0.0356
$\widehat{SD}(\hat{\psi})$	0.0894	0.0366
$\hat{p}$	0.0980	0.2397
$\widehat{SE}(\hat{p})$	0.0170	0.0146
$\widehat{SD}(\hat{p})$	0.0159	0.0152
$\hat{p}'$	0.3977	0.2390
$\widehat{SE}(\hat{p}')$	0.0491	0.0250
$\widehat{SD}(\hat{p}')$	0.0383	0.0226
$\psi(.), \theta(.), \theta'(.), p(.)$		
$\hat{\psi}$	0.7648	0.7534
$\widehat{SE}(\hat{\psi})$	0.1068	0.0355
$\widehat{SD}(\hat{\psi})$	0.1054	0.0360
$\hat{\theta}$	0.0996	0.4843
$\widehat{SE}(\hat{\theta})$	0.0188	0.0793
$\widehat{SD}(\hat{\theta})$	0.0188	0.2647
$\hat{\theta}'$	0.5082	0.4887
$\widehat{SE}(\hat{\theta}')$	0.0686	0.0858
$\widehat{SD}(\hat{\theta}')$	0.0733	0.2670
$\hat{p}$	0.7955	0.6548
$\widehat{SE}(\hat{p})$	0.1054	0.0499
$\widehat{SD}(\hat{p})$	0.1169	0.3210
Number converged	851	753

Notes: Detection history data were generated for a true site occupancy of $\psi = 0.75$ and segment-level detection probability of $p = 0.8$. Detection history data were generated both with ($\theta = 0.1$, $\theta' = 0.5$) and without ($\theta = \theta' = 0.3$) spatial correlation of segment-level occupancy between adjacent segments. Four different models were fit to each simulated data set, and parameter estimates under these models are summarized. Presented are the mean parameter estimates from all 1000 simulated data sets (e.g., $\hat{\psi}$), the mean of the model-based estimates of standard error (e.g., $\widehat{SE}(\hat{\psi})$), and the replication-based estimate of standard deviation of parameter estimates ($\widehat{SD}(\hat{\psi})$). The model that included spatial correlation [$\psi(.)$, $\theta(.)$, $\theta'(.)$, $p(.)$] did not always converge in the simulations, and the number of cases for which convergence was obtained is included.

dency, [$\psi(.)$, $\theta(.)$, $\theta'(.)$, $p(.)$], to be selected as the most appropriate model for the data. This model was developed for exactly this kind of spatial process. Simulation study results led us to suspect that the trap response model, [$\psi(.)$, $p(.)$, $p'(.)$], would provide a fair description of the data and that the standard occupancy model, [$\psi(.)$, $p(.)$], would neither describe the data well nor provide a good estimate of occupancy. For a given model of the detection process (e.g., constant), we expected the model with occupancy written as a function of the habitat covariate, [$\psi(h)$, $p(.)$], to perform better than the constant occupancy model, [$\psi(.)$, $p(.)$]. Overall, we thus expected model [$\psi(h)$, $\theta(.)$, $\theta'(.)$, $p(.)$] to perform best and model [$\psi(.)$, $p(.)$] to perform worst.

RESULTS

Simulations

When data were simulated under the case of no spatial correlation among segments (Table 1, column $\theta = \theta' = 0.3$), parameter estimates for the two standard occupancy models, [$\psi(.)$, $p(.)$] and [$\psi(.)$, $p(t)$], were very close to expectations. Mean occupancy estimates were only slightly larger than the true value of 0.75; bias and relative bias were small. True segment-level detection probability for these standard models can be obtained as the product of segment-level occupancy and detection, conditional on occupancy, $\theta p = 0.3 \times 0.8 = 0.24$. The mean values of $\hat{p}$ under these two models were again nearly identical to 0.24 (Table 1). Similarly, the model-based estimates of standard error performed well, as they matched the empirical estimates of standard deviation of parameter estimates quite closely (Table 1).

However, when data were simulated with spatial correlation (Table 1, column $\theta = 0.1$, $\theta' = 0.5$) the standard models performed very poorly. For example, relative bias of the occupancy estimators under these standard models was approximately −0.30, indicating that occupancy estimates were ~30% too small. The spatial Markov process induced heterogeneity among segments with respect to segment-level occupancy, with segments preceded by an occupied segment having very different probabilities of being occupied than segments not preceded by an occupied segment.

We recognized that the Markov detection process (trap response) model, [$\psi(.)$, $p(.)$, $p'(.)$], did not capture the underlying model of spatial dependence perfectly, but we hoped that it would provide an approximation that yielded reasonable estimates. The occupancy estimator under this model performed much better than the standard occupancy models, but still exhibited negative bias (relative bias nearly −0.07). The estimated

TABLE 2. Model selection statistics for six models fit to tiger (*Panthera tigris*) survey data from southwestern India.

Model	ΔAIC	w	$-2\log(L)$	Parameters
$\psi(h)$, $\theta(.)$, $\theta'(.)$, $p(.)$	0.00	0.87	1564.67	5
$\psi(.)$, $\theta(.)$, $\theta'(.)$, $p(.)$	3.77	0.13	1570.44	4
$\psi(h)$, $p(.)$, $p'(.)$	25.18	0.00	1591.85	4
$\psi(.)$, $p(.)$, $p'(.)$	31.30	0.00	1599.97	3
$\psi(h)$, $p(.)$	123.93	0.00	1692.60	3
$\psi(.)$, $p(.)$	134.14	0.00	1704.81	2

Note: The statistics are the change in Akaike's information criterion, ΔAIC; AIC weight, w; -2 times the logarithm of the likelihood, $-2\log(L)$; and the number of parameters in the model.

detection probability for segments preceded by segments with a detection ($\hat{p}'$) should estimate the product, $\theta' p = 0.5 \times 0.8 = 0.4$, and the mean of the estimates was very close to this value (Table 1). However, detection probability for segments not preceded by segments with a detection (p) pertain to: (1) some segments for which preceding segments were unoccupied and (2) other segments for which previous segments were occupied yet not detected. It is this detection parameter that is not estimated properly, leading to the negative bias in the occupancy estimator.

The model [$\psi(.)$, $\theta(.)$, $\theta'(.)$, $p(.)$] that was developed explicitly for the underlying spatial model used to generate the data performed well with small positive bias in the occupancy estimator (relative bias <0.02). Our main question about implementing this model was whether the underlying parameters were identifiable. The model was not as numerically stable as the other three models with convergence failure occurring in approximately 150 of the 1000 simulations. Because this was a simulation study, there was no attempt to use alternative starting values or try other approaches to obtaining convergence in problem cases. When this model was fit to data generated with no spatial dependence ($\theta = \theta' = 0.3$), it produced reasonable estimates of occupancy and its variance (Table 1). However, the absence of Markovian spatial dependence in the process used to generate the data produced difficulties in estimating the parameters associated with the detection process (the local occupancy and detection parameters) and also led to increased numerical instability (convergence failure in nearly one-quarter of the simulations).

Tiger survey

As expected, model selection results for the tiger survey data strongly support the use of the Markov process models for spatial dependence, [$\psi(h)$, $\theta(.)$, $\theta'(.)$, $p(.)$] and [$\psi(.)$, $\theta(.)$, $\theta'(.)$, $p(.)$] (Table 2). The former model including the effects of habitat on occupancy was favored, as expected, and received a model weight of ~0.87. The coefficient associated with the proportion of the cell in tiger habitat was estimated to be $\hat{\beta}_1 = 0.053$, $\widehat{\mathrm{SE}}(\hat{\beta}_1) = 0.021$. Thus, under the selected model, $\psi(h)$, $\theta(.)$, $\theta'(.)$, $p(.)$, the probability of occupancy at the level of the 200-km^2 cell increased with the proportion of habitat in the cell as predicted. The occupancy estimate for a cell with the average proportion of habitat was $\hat{\psi}(\bar{h}) = 0.50$, $\widehat{\mathrm{SE}}[\hat{\psi}(\bar{h})] = 0.069$. The other model that received support, [$\psi(.)$, $\theta(.)$, $\theta'(.)$, $p(.)$], yielded an occupancy estimate of $\hat{\psi} = 0.57$, $\widehat{\mathrm{SE}}(\hat{\psi}) = 0.077$ (Table 3). The trap response models, [$\psi(h)$, $p(.)$, $p'(.)$] and [$\psi(.)$, $p(.)$, $p'(.)$], developed as crude approximations to the true detection process, received virtually no support, but still provided a relatively better description of the data than did the standard occupancy models, [$\psi(h)$, $p(.)$] and [$\psi(.)$, $p(.)$]. As predicted based on simulation results, the trap response and standard occupancy models yielded progressively smaller estimates of occupancy (Table 3). The naïve estimate computed as the proportion of cells at which tiger sign was detected was 0.36.

The parameter estimates corresponding to the detection process provided strong evidence of the kind of process for which the model was developed. Under the top model, the segment-level occupancy for segments not preceded by an occupied segment was only $\hat{\theta} = 0.07$, whereas occupancy for a segment preceded by an

TABLE 3. Estimates of occupancy and parameters related to the detection process for tiger survey data from southwestern India.

Model	$\hat{\psi}$ ($\widehat{\mathrm{SE}}[\hat{\psi}]$)†	$\hat{p}$ ($\widehat{\mathrm{SE}}[\hat{p}]$)	$\hat{p}'$ ($\widehat{\mathrm{SE}}[\hat{p}']$)	$\hat{\theta}$ ($\widehat{\mathrm{SE}}[\hat{\theta}]$)	$\hat{\theta}'$ ($\widehat{\mathrm{SE}}[\hat{\theta}']$)
$\psi(h)$, $\theta(.)$, $\theta'(.)$, $p(.)$	0.50 (0.069)	0.42 (0.063)	···	0.07 (0.017)	0.79 (0.061)
$\psi(.)$, $\theta(.)$, $\theta'(.)$, $p(.)$	0.57 (0.077)	0.42 (0.059)	···	0.07 (0.017)	0.80 (0.057)
$\psi(h)$, $p(.)$, $p'(.)$	0.42 (0.044)	0.09 (0.008)	0.36 (0.031)	···	···
$\psi(.)$, $p(.)$, $p'(.)$	0.46 (0.045)	0.09 (0.008)	0.36 (0.031)	···	···
$\psi(h)$, $p(.)$	0.38 (0.039)	0.14 (0.008)	···	···	···
$\psi(.)$, $p(.)$	0.41 (0.039)	0.13 (0.008)	···	···	···

Notes: Definitions are as follows: ψ is the probability that a 188-km^2 sample unit is occupied; definitions of p vary among models, but for the first model, it denotes the probability of detecting sign on a segment, given the presence of tigers on the segment (local occupancy); θ denotes occupancy of a segment, given that the preceding segment was unoccupied; θ' denotes occupancy of a segment, given that the preceding segment was occupied.

† For models in which occupancy is a function of habitat, $\psi(h)$, the table shows the estimated occupancy (and SE) at the mean value of the habitat covariate, $\hat{\psi}(\bar{h})$. For all three habitat models, these estimates were within 0.01 of the mean occupancy taken over all cells, $\hat{\bar{\psi}}(h)$.

occupied segment was estimated to be $\hat{\theta}' = 0.79$. The segment-level detection probability, conditional on segment-level occupancy, was estimated to be relatively high as expected, $\hat{p} = 0.42$. This value is larger than the detection parameter estimates under the other two classes of models (Table 3), because these latter parameters incorporate both segment-level occupancy and detection given occupancy.

Discussion

This work was designed to investigate ways of estimating occupancy using spatial replication in a case in which replicate-level occupancy follows a one-dimensional spatial Markov process. We described the basic kind of field survey design that motivated this work and noted that this design is frequently used for surveying large, wide-ranging mammals that typically use trails for marking and movement. We believe this approach has great potential utility because it explicitly models the process that generates signs along trails (animal behavior) and also meets key logistical and practical needs of field surveys in tropical forests (easy movement of survey teams across rugged landscapes using trails and ease of detecting animal signs on them). In this design spatial replicates are visited in a specified order (e.g., segments along a trail), and we suspected that replicate-level occupancy was correlated for adjacent segments (because of behavior of tigers; Karanth and Sunquist 2000). We developed a new occupancy model to deal explicitly with this sort of spatial process. We also developed a new trap response occupancy model as a rough approximation for data obtained under such sampling designs. We then conducted a simulation study generating detection history data according to the hypothesized spatial model for such a survey in order to evaluate estimators obtained under standard occupancy models (MacKenzie et al. 2006) and the two new models that we developed.

In the case of higher probability of local occupancy of a trail segment given local occupancy of the preceding trail segment, occupancy estimators for standard models showed substantial negative bias. The new trap response occupancy model [$\psi(.)$, $p(.)$, $p'(.)$] simply places a Markovian dependence on detection probabilities, such that detection probability for a spatial replicate (trail segment) depends on whether or not sign of the species was detected on the preceding replicate or trail segment. We were confident that such a model could be fit to detection history data and speculated that it might provide a reasonable approximation to the hypothesized underlying spatial process. Simulation results indicated that occupancy estimates from this model were indeed much improved over those of the standard occupancy models, but they were still negatively biased.

The other new occupancy model [$\psi(.)$, $\theta(.)$, $\theta'(.)$, $p(.)$] was developed specifically for the Markov spatial process that was thought to characterize the tiger survey data and that was used to generate the simulated data. However, we were uncertain about whether this model could be readily fit to detection history data and whether the model parameters were really identifiable. The model failed to converge for ~15% of the simulated data sets, but we made no efforts to use different starting values or otherwise deal with the convergence problems. For the remainder of the data sets, convergence was obtained and model estimators performed very well, exhibiting negligible bias.

We fit six models to the data from the large-scale occupancy field survey for tigers in southwestern India that motivated the simulation study. We modeled occupancy as either a constant or a function of the proportion of the grid cell that contained suitable tiger habitat. For each type of occupancy model, we modeled the detection process either using a constant detection probability, the new trap response model, or the new spatial Markov process model. The AIC model selection statistics provided support for the habitat model of occupancy and strong support for the Markov spatial process model, as predicted. The parameter estimates associated with segment-level occupancy provided evidence of a large increase in the probability of local occupancy of a segment when the preceding segment was occupied. The trap response models specifying a Markov process for detections indicated a much higher segment-level detection probability for segments preceded by a detection. The results of these models were consistent with our predictions based on tiger behavior and field survey methods.

Results of the field survey analysis provided evidence that tiger biologists are capable of identifying habitat elements of the Malenad–Mysore Tiger Landscape of Karnataka that are essential for tiger existence and strengthen their arguments for protection of the areas of suitable habitat that do remain. Cells containing large amounts of habitat showed the largest probabilities of tiger occupancy, providing stronger inferences than expert statements about specific areas most important to tigers. The estimated fraction of surveyed cells that were occupied under the most appropriate model was ~0.50, whereas the naïve occupancy estimate obtained as the proportion of cells at which tigers were detected was 0.36. The Karnataka survey demonstrates the ability to assess current range of secretive animals using large-scale field surveys, and the analytic results indicate the need to adequately deal with nondetection in analysis of resulting data. K. U. Karanth and associates are currently investigating specific factors within each cell, in addition to available habitat, that may influence probability of a cell being occupied by tigers.

The occupancy estimates under the six models were also consistent with expectations for a Markov spatial process based on simulation results. The standard occupancy models yielded the smallest occupancy estimates, whereas the Markov spatial process models produced the largest occupancy estimates ($\hat{\psi}(\bar{h}) = 0.50$, $\hat{\psi} = 0.57$). We conclude that the spatial process models

were appropriate for the data and provided an occupancy estimate that was a substantial improvement over the naïve estimate.

We have not extended the kind of modeling presented here to the case of dealing with abundance-induced heterogeneity and abundance estimation (e.g., Royle and Nichols 2003). At present, those interested in such models (or indeed in any other occupancy models requiring independent spatial replicates) can use the model presented here to test for spatial dependence at different segment lengths. For example, the tiger survey data were collected at 1-km intervals, but they can be aggregated to create longer segments of 2 km, 3 km, 4 km, etc. Spatial dependence is expected to decline with segment length, and the model presented here can be used to test for spatial independence in order to select a length at which independence is a reasonable approximation (K. U. Karanth, N. S. Kumar, A. Gopalaswamy, and V. Srinivas, *unpublished manuscript*). Models requiring independence can then be used with these aggregated data sets to draw inferences.

The ability to break trails into segments of various lengths poses the interesting design question of whether there is an optimal segment length with respect to precision or mean squared error. It would be unwise to use segment lengths so short that most of them contain no detections even when animals are present at the level of the sample unit. Similarly, segments should not be so long that fewer than three segments exist in most sample units. Beyond those simple recommendations, it does not seem possible to provide guidelines, or even construct a simulation study, that would be generally useful. Instead, this is the sort of design issue that should be investigated (e.g., via simulation) on a case-by-case basis, and we suspect that the conclusion will depend very heavily on the logistics and biology of the survey situation.

Both new models (Markovian spatial dependence, Markovian detections [Appendix A]) were developed specifically for the use of spatial replicates for occupancy estimation in the presence of Markov spatial processes in detection or replicate-level occupancy. In addition to the increasing use of the described type of survey design for large mammals worldwide, we also foresee application to any road- or trail-based surveys (track surveys, scent station surveys) of species that occasionally use roads and trails as travel routes. Even for species that do not use roads or trails as travel routes, we believe that the models developed here may prove useful. For example, the North American Breeding Bird Survey (e.g., Peterjohn and Sauer 1993) is a road-based survey using avian point counts conducted at 50 stops located along the route at 0.8-km intervals. Adjacent stops are expected to exhibit greater similarity of habitat, on average, than stops located farther apart. In some instances, individual birds with range centers located between two adjacent stops may be detected at each of two stops. Both of these possibilities would be expected to generate the sort of Markovian spatial dependence for which our models were developed.

In addition to applications for certain spatial sampling designs, we believe that these models will also have broad application to occupancy studies that use temporal replication. Consider an occupancy study of a territorial species such that only a single individual or pair or group (e.g., a wolf pack) is likely to use any particular sampling unit during the survey season. Further assume that individuals of the species travel widely, such that a sample unit is likely to be used by the species on some days and not on others. If this use is correlated in time (e.g., a tiger or a wolf pack cycling through a large range, perhaps spending multiple days at a kill site and then moving on), then either of the proposed new models might be a reasonable candidate for describing that situation.

Other situations for which temporal Markov processes may be useful involve strong seasonality of animal presence and use of surveyed sites. For example breeding anuran occupancy can be modeled as a function of date (season) and air temperature, with substantial pulses of breeding activity (Weir et al. 2005). Markovian modeling, with initial local occupancy (θ) a function of such environmental covariates and subsequent local occupancy (θ') probability larger following initial emergence, might be useful in such a situation. Occupancy analyses of butterfly surveys may show multiple peaks of emergence within a season (Kery et al. 2009), and Markovian models for such data should be useful as well. In such situations the suggested Markovian modeling should not only permit reasonable inference about sample unit occupancy, but also permit inference about the temporal pattern of local occupancy (see Kery et al. 2009). In the case of temporal Markovian dependence, estimates of θ and θ' pertain to an occupancy process analogous to temporary emigration of individual animals in capture–recapture studies (e.g., Kendall et al. 1997).

When viewed in the broader context of models for estimation of animal abundance and occupancy, model $[\psi(.), \theta(.), \theta'(.), p(.)]$ represents an explicit decomposition of the detection process into two components: availability and detection given availability (see discussion in Nichols et al. 2008*b*). With respect to abundance estimation, availability refers to an individual animal having a nonzero probability of detection, for example because it is on the water surface, not submerged (e.g., Marsh and Sinclair 1989), at the time of a visual survey or because it vocalizes at the time of an auditory survey (e.g., Farnsworth et al. 2002) or because it is present in the sampled area (not a temporary emigrant) during a capture period in a trapping study (Kendall et al. 1997). In cases of occupancy estimation for mobile animals that use areas larger than a single sample unit, the detection process can also be decomposed into two components: (1) local presence of species at the temporal or spatial replicate and (2) detection given local

presence. Similar decomposition was also used to estimate occupancy at two different spatial scales using data from multiple detection devices at the local sampling site (Nichols et al. 2008*a*). This sort of decomposition of the detection process may be needed to properly model detection, as in the present study, and it may also provide the ability to address interesting ecological questions (Nichols et al. 2008*a*).

The kind of modeling used here may also be relevant to the still broader application area of cluster sampling (e.g., Thompson 2002). Specifically, the process and/or observation model for the variable of interest (e.g., occupancy) might differ between the two levels at which sampling occurs, the cluster and the subunit within a cluster. It is possible that there are other situations in which explicit modeling of the process governing the distribution of the variable across subunits might yield improved inference. In the spirit of placing this modeling within a general context, we also note that this work represents a special application of inference for hidden Markov chains (e.g., Cappé et al. 2005).

We note the potential to extend the type of modeling presented here to capture–recapture modeling of Markovian temporary emigration in demographically "closed" animal populations. Kendall et al. (1997) described approaches to the modeling of temporary emigration for open populations using Pollock's (1982) robust design, and they included Markovian models. Kendall (1999) showed that abundance estimates under capture–recapture models for closed populations are robust to random (non-Markovian) temporary emigration. However, Markovian temporary emigration, such that presence of an animal on a trapping array at one sample period depends on whether it was present or not the previous sample period, was shown to induce bias in abundance estimates (Kendall 1999). The type of Markovian modeling used here can be used in closed capture–recapture modeling to permit estimation of abundance in the face of such Markovian temporary emigration.

We conclude with a consideration of the two spatial sampling designs that motivated this work (Fig. 1). We introduced the topic by noting that use of spatial replicates with standard occupancy models strictly requires that spatial replicates be selected randomly and with replacement from each sample unit (e.g., grid cell). Kendall and White (2009) provide some results on the magnitudes of bias that can arise during certain sampling situations when sampling is not conducted with replacement. In the absence of spatial dependence of segment-level occupancy (this is the spatial model that we attempt to induce by random sampling), standard occupancy models appear to perform reasonably. However, in the presence of spatial dependence, standard models yield biased estimates of grid cell occupancy. The model that we present here can be used to test for such dependence, and if it is present, to provide reasonable parameter estimates. Our initial development of the Markov spatial model was motivated by biological and logistical considerations associated with certain kinds of field surveys. However, the Markovian dependence that we initially viewed as a nuisance actually permits decomposition of the detection process in a manner that is not possible with independent segments (e.g., see the poor estimator performance for $\hat{\theta}$, $\hat{\theta}'$, and $\hat{p}$ in the last column of Table 1 for model $[\psi(.), \theta(.), \theta'(.), p(.)]$). Thus, if we are interested either in decomposing the detection process or in focusing for other reasons on lower level occupancy, then sampling designs that lead to Markovian spatial dependence provide an advantage.

Acknowledgments

We thank Shannon Barber-Meyer and Sunarto Sunarto for asking questions about consequences of trail-based sampling designs for occupancy estimation. Two anonymous referees provided constructive comments on the initial submission that were useful in preparing this final revision.

Literature Cited

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference. Second edition. Springer, New York, New York, USA.

Cam, E., J. D. Nichols, J. E. Hines, and J. R. Sauer. 2002. On the estimation of species richness from species accumulation data. Ecography 25:102–108.

Cappé, O., E. Moulines, and T. Rydén. 2005. Inference in hidden Markov models. Springer, New York, New York, USA.

Farnsworth, G. L., K. H. Pollock, J. D. Nichols, T. R. Simons, J. E. Hines, and J. R. Sauer. 2002. A removal model for estimating detection probabilities from point count surveys. Auk 119:414–425.

Hines, J. E. 2006. PRESENCE2: software to estimate patch occupancy and related parameters. U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel, Maryland, USA. ⟨http://www.mbr-pwrc.usgs.gov/software/presence.html⟩

Karanth, K. U., J. D. Nichols, N. S. Kumar, W. A. Link, and J. E. Hines. 2004. Tigers and their prey: predicting carnivore densities from prey abundance. Proceedings of the National Academy of Sciences USA 101:4854–4858.

Karanth, K. U., and M. E. Sunquist. 2000. Behavioural correlates of predation by tiger, leopard and dhole in Nagarahole, India. Journal of Zoology, London 250:255–265.

Kendall, W. L. 1999. Robustness of closed capture–recapture methods to violations of the closure assumption. Ecology 80: 2517–2525.

Kendall, W. L., J. D. Nichols, and J. E. Hines. 1997. Estimating temporary emigration and breeding proportions using capture–recapture data with Pollock's robust design. Ecology 78:563–578.

Kendall, W. L., and G. C. White. 2009. A cautionary note on substituting spatial subunits for repeated temporal sampling in studies of site occupancy. Journal of Applied Ecology 46: 1182–1188.

Kery, M., J. A. Royle, M. Plattner, and R. M. Dorazio. 2009. Species richness and occupancy estimation in communities subject to temporary emigration. Ecology 90:1279–1290.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy when detection probabilities are less than one. Ecology 83:2248–2255.

MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. A. Bailey, and J. E. Hines. 2006. Occupancy modeling and estimation. Academic Press, San Diego, California, USA.

Marsh, H., and D. F. Sinclair. 1989. Correcting for visibility bias in strip transect aerial surveys of aquatic fauna. Journal of Wildlife Management 53:1017–1024.

Nichols, J. D., L. L. Bailey, A. F. O'Connell, Jr., N. W. Talancy, E. H. Campbell, E. H. C. Grant, A. T. Gilbert, E. M. Annand, T. P. Husband, and J. E. Hines. 2008*a*. Multi-scale occupancy estimation and modeling using multiple detection methods. Journal of Applied Ecology 45:1321–1329.

Nichols, J. D., L. Thomas, and P. B. Conn. 2008*b*. Inferences about landbird abundance from count data: recent advances and future directions. Pages 201–235 *in* D. L. Thomson, E. G. Cooch, and M. J. Conroy, editors. Modeling demographic processes in marked populations. Springer, New York, New York, USA.

Peterjohn, B. G., and J. R. Sauer. 1993. North American breeding bird survey annual summary 1990–91. Bird Populations 1:1–15.

Pollock, K. H. 1982. A capture–recapture design robust to unequal probability of capture. Journal of Wildlife Management 46:752–756.

Royle, J. A., and R. M. Dorazio. 2008. Hierarchical modeling and inference in ecology. Academic Press, New York, New York, USA.

Royle, J. A., and J. D. Nichols. 2003. Estimating abundance from repeated presence–absence data or point counts. Ecology 84:777–790.

Thompson, S. K. 2002. Sampling. Wiley, New York, New York, USA.

Weir, L. A., J. A. Royle, P. Nanjappa, and R. E. Jung. 2005. Modeling anuran detection and site occupancy on North American Amphibian Monitoring Program (NAAMP) routes in Maryland. Journal of Herpetology 39:627–639.

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations. Academic Press, San Diego, California, USA.

APPENDIX A

Markov process for detections, model [$\psi(.)$, $p(.)$, $p'(.)$] (*Ecological Archives* A020-053-A1).

APPENDIX B

Computing expression for probabilities associated with detection histories, $\Pr(h_j)$, under model [$\psi(.)$, $\theta(.)$, $\theta'(.)$, $p(.)$] (*Ecological Archives* A020-053-A2).

APPENDIX C

Simulation results comparing maximum-likelihood estimates with those obtained using the Markov chain Monte Carlo method (*Ecological Archives* A020-053-A3).

SUPPLEMENT

WinBUGS code for model [$\psi(.)$, $\theta(.)$, $\theta'(.)$, $p(.)$] (*Ecological Archives* A020-053-S1).

Appendix G

Monitoring carnivore populations at the landscape scale: occupancy modelling of tigers from sign surveys

Journal of Applied Ecology

British Ecological Society

Journal of Applied Ecology 2011, **48**, 1048–1056 doi: 10.1111/j.1365-2664.2011.02002.x

Monitoring carnivore populations at the landscape scale: occupancy modelling of tigers from sign surveys

Kota Ullas Karanth[1,2]*, Arjun M. Gopalaswamy[2,3,4], Narayanarao Samba Kumar[2,3], Srinivas Vaidyanathan[3,5], James D. Nichols[6] and Darryl I. MacKenzie[7]

[1]*Wildlife Conservation Society – Global Conservation Program, 2300 Southern Boulevard, Bronx, NY 10460, USA;* [2]*Centre for Wildlife Studies, 26-2, Aga Abbas Ali Road (Apt: 403), Bengaluru 560 042, India;* [3]*Wildlife Conservation Society – India Program, 1669, 31st Cross, 16th Main, Banashankari 2nd Stage, Bengaluru 560 070, India;* [4]*Wildlife Conservation Research Unit (WildCRU), The Recanati-Kaplan Centre, Department of Zoology, Tubney House, University of Oxford, Abingdon Road, Tubney, Abingdon OX13 5QL, UK;* [5]*Foundation for Ecological Research, Advocacy and Learning, No 27, 2nd Cross Appavounagar, Vazhakulam, Pondicherry 605 012, India;* [6]*United States Geological Survey, Patuxent Wildlife Research Center, Laurel, MD 20708, USA; and* [7]*Proteus Wildlife Research Consultants, PO Box 5193, Dunedin, New Zealand*

Summary

1. Assessing spatial distributions of threatened large carnivores at landscape scales poses formidable challenges because of their rarity and elusiveness. As a consequence of logistical constraints, investigators typically rely on sign surveys. Most survey methods, however, do not explicitly address the central problem of imperfect detections of animal signs in the field, leading to underestimates of true habitat occupancy and distribution.

2. We assessed habitat occupancy for a tiger *Panthera tigris* metapopulation across a *c.* 38 000-km^2 landscape in India, employing a spatially replicated survey to explicitly address imperfect detections. Ecological predictions about tiger presence were confronted with sign detection data generated from occupancy sampling of 205 sites, each of 188 km^2.

3. A recent occupancy model that considers Markovian dependency among sign detections on spatial replicates performed better than the standard occupancy model (ΔAIC = 184·9). A formulation of this model that fitted the data best showed that density of ungulate prey and levels of human disturbance were key determinants of local tiger presence. Model averaging resulted in a replicate-level detection probability $\hat{\bar{p}}_t(S\hat{E}[\hat{\bar{p}}_t])$ = 0·17 (0·17) for signs and a tiger habitat occupancy estimate of $\hat{\psi}(S\hat{E}[\hat{\psi}])$ = 0·665 (0·0857) or 14 076 (1814) km^2 of potential habitat of 21 167 km^2. In contrast, a traditional presence-versus-absence approach underestimated occupancy by 47%. Maps of probabilities of local site occupancy clearly identified tiger source populations at higher densities and matched observed tiger density variations, suggesting their potential utility for population assessments at landscape scales.

4. *Synthesis and applications*. Landscape-scale sign surveys can efficiently assess large carnivore spatial distributions and elucidate the factors governing their local presence, provided ecological and observation processes are both explicitly modelled. Occupancy sampling using spatial replicates can be used to reliably and efficiently identify tiger population sources and help monitor metapopulations. Our results reinforce earlier findings that prey depletion and human disturbance are key drivers of local tiger extinctions and tigers can persist even in human-dominated landscapes through effective protection of source populations. Our approach facilitates efficient targeting of tiger conservation interventions and, more generally, provides a basis for the reliable integration of large carnivore monitoring data between local and landscape scales.

*Correspondence author. E-mail ukaranth@gmail.com

© 2011 The Authors. Journal of Applied Ecology © 2011 British Ecological Society

Key-words: conservation, density variation, detection probability, landscape ecology, metapopulations, occupancy sampling, spatial models

Introduction

Large carnivores are among the most threatened mammals in the world (Ceballos *et al.* 2005). The decline of the tiger *Panthera tigris* Linnaeus, 1758, typifies the challenges of recovering large carnivore populations. Driven by synergistic impacts of habitat fragmentation, prey depletion and direct hunting (Karanth *et al.* 2004; Walston *et al.* 2010), tigers have suffered a range contraction of *c.* 93% in the past two centuries (Dinerstein *et al.* 2006). Global species recovery now depends on effective management of the remaining *c.* 40 'tiger source populations' (see Walston *et al.* 2010 for details) that occur as interconnected clusters within larger landscapes (Dinerstein *et al.* 2006; Ranganathan *et al.* 2008). Therefore, rigorous assessments of tiger populations are required at two spatial scales: protected reserves of a few hundred square kilometres each and wider landscapes of several thousand square kilometres.

Within reserves, tiger demographic parameters can be assessed reliably using intensive capture–recapture sampling of photographs or faecal DNA from individuals (Karanth *et al.* 2004, 2006; Mondol *et al.* 2009). Such intensive methods explicitly deal with the central problem of imperfect detections: that a proportion of animals in the surveyed area go undetected (Williams, Nichols & Conroy 2002). However, because of logistical constraints, landscape-scale surveys typically cannot employ intensive methods and depend on detecting tiger signs such as tracks and scats (Miquelle *et al.* 1999; Smith *et al.* 1999a,b; Johnsingh *et al.* 2004; Carroll & Miquelle 2006; Barlow *et al.* 2008; Jhala, Gopal & Qureshi 2008; Jhala, Qureshi & Gopal 2011). However, these 'presence-versus-absence' surveys do not discriminate between nondetection of sign and true absence of tigers, leading to underestimates of spatial distribution (MacKenzie *et al.* 2002). In past tiger surveys, environmental covariates have been incorporated into analyses in attempts to address this problem (Smith *et al.* 1999a,b; Miquelle *et al.* 2005; Barlow *et al.* 2008; Jhala, Gopal & Qureshi 2008; Jhala, Qureshi & Gopal 2011). Such results are, however, difficult to interpret because the covariates used may influence detection probability or occupancy or even both these quantities. More generally, investigators have used habitat suitability models (Engler, Guisan & Rechsteiner 2004) of various kinds, including regressions (Tyre, Possingham & Lindenmayer 2001; Gu & Swihart 2004), resource-selection functions (Boyce & McDonald 1999), ecological niche factor analysis (Hirzel *et al.* 2002) and simulations (Manley *et al.* 2004) for the analyses of animal presence/absence data. Even these approaches, however, do not explicitly model or estimate detection probabilities.

On the other hand, habitat occupancy models (MacKenzie *et al.* 2002, 2006; Royle & Dorazio 2008) do explicitly deal with imperfect detections and are increasingly being favoured (e.g. Magoun *et al.* 2007 for wolverines *Gulo gulo* Linnaeus, 1758). If carefully designed, they can meet conservation needs (MacKenzie & Royle 2005) by explicitly confronting ecological or management predictions with sign survey data without ignoring imperfect detections. In practice, 'replications' necessary for estimating occupancy probabilities can be either 'temporal' or 'spatial' (Kendall & White 2009).

In this study, we confronted predictions incorporating *a priori* ecological and management hypotheses about local (site-level) tiger presence with data from a carefully designed, spatially replicated survey of tiger signs across a *c.* 38 000-km^2 landscape. This landscape currently harbours one of the largest wild tiger populations globally (Walston *et al.* 2010) holding an estimated 20% of India's wild tigers (Jhala, Gopal & Qureshi 2008). Using simple sign survey data, we modelled the ecological process of interest to us (tiger distribution and local presence) as well as the observation process (survey method), within a single inferential framework (MacKenzie *et al.* 2006; Royle & Dorazio 2008).

Based on prior knowledge of tiger ecology in the landscape (Karanth & Sunquist 2000; Karanth *et al.* 2004, 2006), we hypothesized that tiger 'source populations' that reproduce above replacement levels (Walston *et al.* 2010) would be confined to a few well-protected reserves and that tigers would occur at very low densities or be absent elsewhere. We further hypothesized that abundance of large ungulate prey species and protection from human disturbance (Karanth *et al.* 2004) would be the key determinants of local tiger presence. Our specific objectives were two-fold: First, to design a practical, landscape-scale, spatially replicated sign survey method to generate reliable estimates of tiger distribution that explicitly deals with imperfect detection of signs. Secondly, to measure local tiger habitat occupancy rates in the landscape and assess which key ecological variables (e.g. prey abundance) and management factors (e.g. human disturbances) influence these occupancy patterns. We present a practical application of a new occupancy model developed by Hines *et al.* (2010) specifically for spatially replicated sign surveys. These results have wider utility for monitoring of many other species of large carnivores (and mammals in general) at landscape scales.

Materials and methods

THE STUDY AREA, ECOLOGY AND HUMAN IMPACTS

The Western Ghats region of India (see Das *et al.* 2006 for a description; Fig. 1) is globally important for tigers (Dinerstein *et al.* 2006; Walston *et al.* 2010). Our survey was conducted in the central part of this region, known as the Malenad–Mysore Tiger Landscape (MMTL) in Karnataka State. The land cover is a matrix of natural moist-evergreen, moist-deciduous and dry-deciduous forests, interspersed with horticultural, agricultural and forestry crops covering 38 350 km^2 (Krishnaswamy, Kiran & Davande 2003). In this matrix,

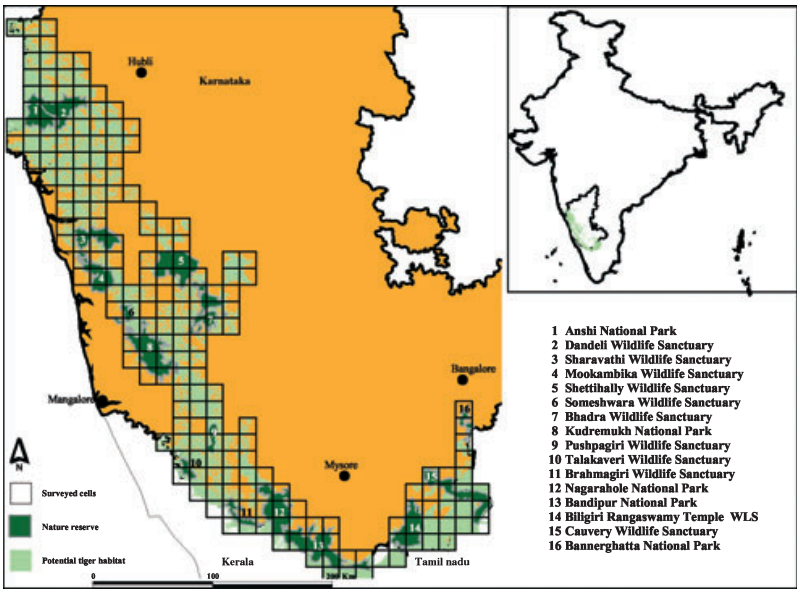


Fig. 1. Potential tiger habitat in Karnataka, India. Inset map: Location of the study area referred to as Malenad–Mysore Tiger Landscape (MMTL) in India.

potential tiger habitat consists of a forested area of 21 167 km^2 (Fig. 1) that includes 14 wildlife reserves (5 500 km^2). This landscape, however, also supports ≈10·5 million people (Directorate of Census Operation 2004) and is characterized by pressures from illegal hunting, forest product collection, livestock grazing and rapid economic growth. Despite these pressures, tiger populations in better protected reserves are at higher densities and reproducing above replacement levels, facilitating dispersals through the wider landscape (Karanth *et al.* 2004; K. U. Karanth unpublished data). The forest connectivity appears to be sufficient (Das *et al.* 2006) for the tiger metapopulation to persist.

The principal ungulate prey of tigers (as well as of leopards *Panthera pardus* Linnaeus, 1758 and dholes *Cuon alpinus* Pallas, 1811) in this landscape are as follows: gaur *Bos frontalis* Lambert, 1804; sambar *Rusa unicolor* Kerr, 1792; chital *Axis axis* Erxleben, 1777; wild pig *Sus scrofa* Linnaeus, 1758; and muntjac *Muntiacus muntjak* Zimmermann, 1780. Based on tiger: prey density ratios established by Karanth *et al.* (2004), we expected the densities of wild ungulates to be an important covariate of local tiger presence. The effectiveness of official wildlife protection varies across MMTL, being high in Bandipur, Nagarahole, Bhadra reserves, moderate in BR Temple, Kudremukh and Dandeli-Anshi reserves and somewhat poor at other locations (Fig. 1). Free grazing by livestock is widespread, often accompanied by illegal hunting and depletion of wild prey (Madhusudan & Mishra 2003; Rayar 2010). Because officials try to prevent intrusions by humans and livestock in nature reserves, and herders typically tend to stay within *c.* 10 km from their settlements, we assumed that encounter rate with signs of livestock is an easily detected surrogate for the intensity of negative human impacts (and for lack of effective protection) at the site level.

FIELD SURVEY PROTOCOLS

Survey design

The proportion of habitat in MMTL occupied by tigers, occupancy, ψ, was a key parameter of interest. Our 15-month-long survey (February 2006–May 2006 and December 2006–May 2007), with each sample

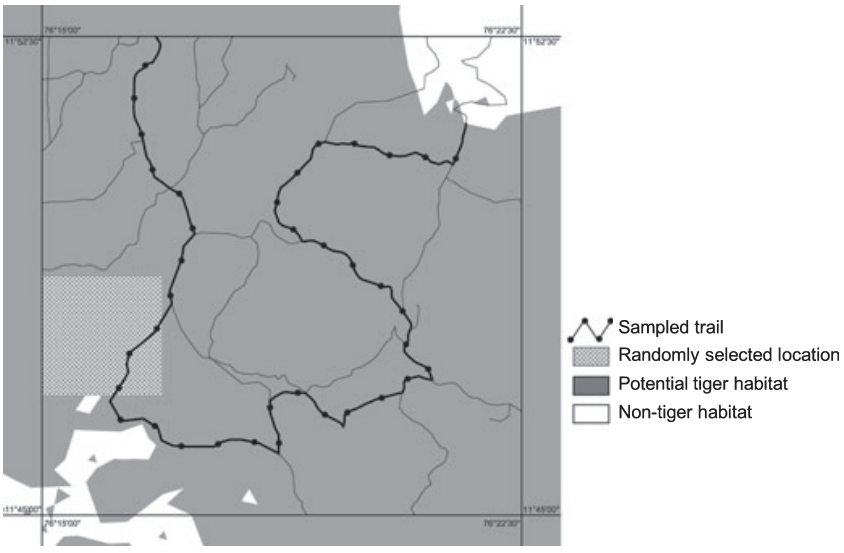


Fig. 2. Diagram of a sampled site (grid cell) depicting sequential coverage of a 1-km-long spatial replicates.

unit being surveyed within < 12–36 h, reasonably met the assumption of 'closure' (MacKenzie *et al.* 2006). The total survey effort was 2021 man-days, involving 4 174 km of walks along trails, and lead to the detection of 403 tiger signs (223 scat deposits and 180 sets of tracks).

The sample units consisted of a grid of square cells (sites) superimposed on the land-cover matrix in MMTL (Fig. 1). Because our goal was to measure true occupancy (proportion of area occupied) rather than intensity of habitat use by tigers (MacKenzie & Royle 2005; MacKenzie *et al.* 2006), the cell size chosen was, on average, larger than the maximum expected tiger home range size. Because nonresident male tigers have the largest home ranges (*c.* 80–90 km^2; Karanth & Sunquist 2000), we used the cells of 188 km^2 size, whose boundaries coincided with graticules on our field maps (Fig. 2). Such 'large' grid cells also enabled us to logistically deploy survey teams more efficiently.

Occupancy studies typically survey only a subsample of cells (MacKenzie & Royle 2005; MacKenzie *et al.* 2006) and thereafter try to extend inference to the unsurveyed cells using covariate data from surveyed cells. However, we surveyed all the cells in MMTL because a key objective was to map tiger distribution to identify 'source populations' at higher densities and sufficient manpower was available.

We opted for spatial replication over temporal (Kendall & White 2009) primarily for logistical reasons (Hines *et al.* 2010). The survey effort/cell was a compromise that considered sufficiency of sign detections, replications and spatial coverage achieved. Thus, we fixed the total sampling effort (distance walked) in a cell with 100% habitat at 40 km, reducing the effort in decreasing proportion to habitat available. Cells with < 10% tiger habitat (20 km^2) and forest fragments of < 10 km^2 size were excluded because tigers were unlikely to occupy these. Advance planning ensured adequate spatial coverage of cells (Kendall & White 2009) and field teams passed through one randomly chosen location in each cell (Fig. 2).

Field protocol based on tiger behaviour and environmental factors

The survey was conducted in dry seasons (October–May) to minimize variations in animal sign detection probabilities induced by rainfall. Because surveys targeted tracks/dung deposited along trails, variations in forest cover type did not influence sign detections.

Typically, tigers move *c.* 1–20 km day^{-1} along forest trails to hunt or to locate, avoid or deter conspecifics (Karanth & Sunquist 2000). Their passage is marked by tracks and occasional scat deposits (Sunquist 1981; Smith, McDougal & Miquelle 1989). Fresh signs of tigers, leopards, dholes and major ungulate prey species in MMTL could be identified by our trained observers. Only unambiguously identified signs were recorded. Overall, the number of spatial replicates in a cell (sampling effort) varied from 4 to 40, depending on the extent of potential habitat in the cell.

Survey teams comprising of 2–3 skilled trackers walked the trails from morning to dusk (Fig. 2). They photographed, georeferenced and recorded signs of tigers, leopards, dholes and ungulate prey species, as well as signs of livestock presence. Each type of sign detection was assigned only once to each 100-m trail segment, thus yielding the standard '1' (detection) or '0' (nondetection) histories required for occupancy analyses (MacKenzie *et al.* 2006; Hines *et al.* 2010). These sign detection data were aggregated at 1 km length to form 'spatial replicates'. At this replicate length, we were reasonably sure to detect the signs of prey as well as of human disturbances if present. Furthermore, as these metrics were being used only as covariates influencing tiger presence in our models, obtaining measures of their absolute density using intensive methods addressing imperfect detection was unnecessary as well as impractical at this spatial scale.

MODELLING AND ANALYSIS OF DATA

We used a recent refinement of the standard occupancy model (MacKenzie *et al.* 2002) developed by Hines *et al.* (2010), which explicitly deals with Markovian dependence of animal sign detection events on spatial replicates. General model notations are provided elsewhere and additional notations are provided under specific model descriptions. Analyses were performed using the program PRESENCE (Hines 2006).

Model structure and spatial dependency

The standard occupancy model is based on two key parameters:

ψ is the probability that a cell is occupied by tigers.

p is the probability of detecting tiger presence in a replicate, given that the cell is occupied by tigers.

The likelihood formulation for this model has been presented elsewhere (see MacKenzie *et al.* 2002 for details). However, occupancy models of MacKenzie *et al.* (2002) assume that replicate surveys can be modelled as statistically independent Bernoulli trials. Our replicates were aggregations of 100-m trail segments at 1 km. Because tigers might walk longer than 1 km along trails (Smith, McDougal & Miquelle 1989; Karanth & Sunquist 2000), we assumed that sign detection events on successive spatial replicates could potentially lack the statistical independence required by the occupancy models of MacKenzie *et al.* (2002). Therefore, we also considered the model developed by Hines *et al.* (2010), which explicitly accounts for such potential dependency using a first-order Markov process (Gillespie 1992). Of the two variants of the Hines *et al.*'s (2010) model implemented in program PRESENCE, we chose the variant that assumes that surveys can begin on any randomly chosen replicate (Fig. 2). In < 10% of the cells surveyed, when teams encountered unsuitable habitat along the trail (e.g. human settlements, large reservoirs) the resulting disjointed trail segments were combined sequentially (Fig. 2). The general computing expressions adapted from Hines *et al.* (2010) for probabilities associated with such sign detection histories are given in Appendix S1, Supporting Information.

The global model and ecological determinants of tiger presence

We first compared the models of MacKenzie *et al.* (2002) and Hines *et al.* (2010) without additional covariates, to choose the appropriate model type for conducting further analyses. All model comparisons were based on Akaike Information Criterion (AIC) values (Burnham & Anderson 2002), because sample size of cells was adequate (> 200, with 4–40 spatial replicates per cell).

We used program PRESENCE (Hines 2006) to estimate the overall occupancy rate, ψ, with the inference applying to the entire landscape matrix of 38 350 km^2 framed by the 205 surveyed cells (Fig. 1) including potential tiger habitat of 21 167 km^2. To estimate overall tiger habitat occupancy within MMTL, the cell-specific occupancy parameter, ψ, was weighted by the area of potential tiger habitat in each cell (Bevington & Robinson 2003; see detailed computations in Appendix S2, Supporting Information).

Based on earlier work (Karanth *et al.* 2004), which demonstrates a tiger: prey density ratio of *c.* 1 : 500, we hypothesized that local tiger presence would be positively influenced by the density of wild ungulate prey. We further wanted to test whether prey in certain body size classes influenced tiger presence. These 'prey-density covariates' for each cell were proportions of 1-km-long replicates that contained signs of each prey species. We thus considered 12 plausible alternative

occupancy models, incorporating the following covariates: density of all prey species combined and density of only larger ungulate species. To assess whether tiger presence is influenced by human disturbance, the proportion of replicates with livestock sign was used as an additional covariate in our models. Further, we also incorporated proportion of tiger habitat/cell (or sampling effort) as a covariate (Appendix S2, Supporting Information).

Model selection and parameter estimation

The ecological process parameter of interest was the cell-level tiger habitat occupancy rate ψ. We were also interested in the important covariates that influenced this site-level tiger presence. The sampling process parameters of interest were the replicate-level occupancy parameters θ and θ' that demonstrate the degree of dependence between replicated samples. However, we expected that some of the covariates influencing tiger presence may also contribute to variation in tiger abundance and thereby influence sign detection probability p_t (Royle & Nichols 2003). We also expected that other unknown factors could also influence the detectability of tiger signs. Thus, the nuisance parameter p_t had to be explicitly accounted for in our modelling process. Consequently, our model selection process initially focused on determining a suitable covariate model structure for p_t and subsequently on using this model structure to estimate cell-level occupancy ψ.

We defined a global occupancy model based on the recommendations of Burnham & Anderson (2002) and MacKenzie *et al.* (2006) and formulated 12 alternative plausible models to identify a covariate structure for the replicate-level detectability p_t, using the Hines *et al.*'s (2010) model structure. No covariates were included for replicate-level occupancy parameters θ and θ' in the global model described below:

ψ (*all prey + livestock + proportion of habitat*), θ (.),θ' (.), p_t(*all prey + livestock + proportion of habitat*).

Using program PRESENCE (Hines 2006), we selected the highest ranked 12 alternative models based on AIC values, to fix the model structure for p_t (Table 1). Thereafter, we kept this model structure component unchanged and ran further occupancy analyses for comparing models involving different covariate structures for ψ. We computed the final estimates of cell-specific occupancy rate $\hat{\psi}_i$, the total fraction of area occupied $\hat{\psi}$ and the replicate-level occupancy parameters $\hat{\theta}$ and $\hat{\theta}'$ and other parameters by model averaging using the top 10 models, all of which had similar support (ΔAIC < 2 units; Burnham & Anderson 2002). We note that the current version of PRESENCE (Hines 2006) does not facilitate any goodness-of-fit test for the spatial dependence model (Hines *et al.* 2010).

Results

MODEL SELECTION

In the first step of model selection, we compared the standard occupancy model of MacKenzie *et al.* (2002) which assumed statistical independence among tiger sign detections made on 1-km-long spatial replicates, with the Hines *et al.* (2010) model that provides for a first-order Markovian dependence among detections (See Materials and methods). These comparisons very clearly showed a lack of independence in sign detections on 1-km-long replicates. The Hines *et al.* (2010) model performed significantly better (ΔAIC = 184·9) in comparisons with the standard model (MacKenzie *et al.* 2002). Therefore, all further occupancy analyses were performed using Hines *et al.* (2010) modelling approach implemented in program PRESENCE.

In the second step, we compared the adequacy of 12 plausible alternative occupancy models, which describe the expected influence of different combinations of covariates (see Materials and methods) on tiger occupancy as well as on detection probability parameters at the replicate as well as cell levels (Table 1). In these comparisons, a general model structure for the covariates likely to influence ψ remains fixed, whereas the covariates associated with replicate-level detection probability p_t are varied. Based on these comparisons, a model that incorporates the influence of combined abundance of larger prey (gaur, sambar, chital and wild pig), human disturbance (via relative abundance of livestock signs) and proportion of habitat surveyed (sampling effort) on replicate-level detection probability p_t emerged as the top-ranked model (AIC weight = 0·309). We note that the covariate structure for the detection probability in the top-ranked model is the most parameterized, thus limiting bias on occupancy estimates. Therefore, in all subsequent occupancy analyses, this basic

Table 1. Model selection results; role of covariates in determining probability of detecting tiger sign p_t on 1-km-long replicates, based on covariates for probability of tiger occupancy from the global model (see Materials and methods), using the Hines *et al.*'s (2010) model. Number of sites = 205. Covariates considered all prey (AP), large prey (LP), livestock (LVS) and proportion of available habitat (PH)

Model	AIC	ΔAIC	AIC weight	Model likelihood	Number of parameters	Deviance
ψ (AP + LVS + PH), θ (.), θ' (.), p_t (LP + LVS + PH)	1446·84	0	0·309	1	10	1426·84
ψ (AP + LVS + PH), θ (.), θ' (.), p_t (AP + LVS + PH)	1447·43	0·59	0·230	0·745	10	1427·43
ψ (AP + LVS + PH), θ (.), θ' (.), p_t (LP + LVS)	1447·66	0·82	0·205	0·664	9	1429·66
ψ (AP + LVS + PH), θ (.), θ' (.), p_t (AP + LVS)	1447·89	1·05	0·183	0·592	9	1429·89
ψ (AP + LVS + PH), θ (.), θ' (.), p_t (LP + PH)	1452·37	5·53	0·019	0·063	9	1434·37
ψ (AP + LVS + PH), θ (.), θ'(.), p_t (LP)	1452·44	5·6	0·019	0·061	8	1436·44
ψ (AP + LVS + PH), θ (.), θ' (.), p_t (AP)	1452·47	5·63	0·019	0·060	8	1436·47
ψ (AP + LVS + PH), θ (.), θ' (.), p_t (AP + PH)	1452·78	5·94	0·016	0·051	9	1434·78
ψ (AP + LVS + PH), θ (.), θ' (.), p_t (LVS + PH)	1466·35	19·51	0	0	9	1448·35
ψ (AP + LVS + PH), θ (.), θ' (.), p_t (LVS)	1468·51	21·67	0	0	8	1452·51
ψ (AP + LVS + PH), θ (.), θ' (.), p_t (.)	1497·79	50·95	0	0	7	1483·79
ψ (AP + LVS + PH), θ (.), θ' (.), p_t (PH)	1498·81	51·97	0	0	8	1482·81

Table 2. Model selection results; role of covariates in determining probability of tiger occupancy ψ in Malenad–Mysore Tiger Landscape, based on modelling probability of detecting tiger sign p_t on 1-km-long replicates using the Hines *et al.* (2010). Covariate structure for p_t is obtained from the top model in Table 1 (see Materials and Methods). Number of sites = 205. Covariates considered all prey (AP), large prey (LP), livestock (LVS) and proportion of available habitat (PH)

Model	AIC	ΔAIC	AIC weight	Model likelihood	Number of parameters	Deviance
ψ (AP + PH), θ (.), θ′ (.), p_t (LP + LVS + PH)	1445·77	0	0·158	1	9	1427·77
ψ (LP + PH), θ (.), θ′ (.), p_t (LP + LVS + PH)	1445·99	0·22	0·142	0·896	9	1427·99
ψ (LVS + PH), θ (.), θ′ (.), p_t (LP + LVS + PH)	1445·99	0·22	0·142	0·896	9	1427·99
ψ (PH), θ (.), θ′ (.), p_t (LP + LVS + PH)	1446·71	0·94	0·099	0·625	8	1430·71
ψ (AP + LVS + PH), θ (.), θ′ (.), p_t (LP + LVS + PH)	1446·84	1·07	0·093	0·586	10	1426·84
ψ (LP + LVS + PH), θ (.), θ′(.) p_t (LP + LVS + PH)	1446·99	1·22	0·086	0·543	10	1426·99
ψ (.), θ (.), θ′ (.), p_t (LP + LVS + PH)	1447·34	1·57	0·072	0·456	7	1433·34
ψ (AP), θ (.), θ′ (.), p_t (LP + LVS + PH)	1447·35	1·58	0·072	0·454	8	1431·35
ψ (LVS), θ (.), θ′ (.), p_t (LP + LVS + PH)	1447·44	1·67	0·069	0·434	8	1431·44
ψ (LP), θ (.), θ′ (.), p_t (LP + LVS + PH)	1447·46	1·69	0·067	0·430	8	1431·46

structure of the detection probability component of the occupancy model was retained, with only the covariate structure describing ψ being varied.

In the third step of the model selection process, we had a candidate set of 12 models by varying covariate model structures on cell-level occupancy parameter ψ while using our best supported structure for replicate-level detection probability p_t (*large prey + livestock + proportion of habitat*). Comparisons between 10 higher-ranked models are reported in Table 2. We note all these models fit the data well (ΔAIC < 1·69), and therefore, we used model averaging to derive final parameter estimates (Burnham & Anderson 2002).

ESTIMATES OF OCCUPANCY AND OTHER PARAMETERS

We used parametric bootstrapping (Efron 1982) to derive the estimates of variance for the total fraction of the area occupied ψ for all models (see Appendix S2, Supporting Information). The final parameter estimates derived for tiger habitat occupancy and sign detection probabilities from model averaging are reported as follows:

$\hat{\theta}_0(S\hat{E}[\hat{\theta}_0])$, probability of tiger presence on the first replicate = 0·418 (0·0787).

$\hat{\theta}(S\hat{E}[\hat{\theta}])$, probability of tiger presence on a replicate, given absence on previous replicate = 0·194 (0·0399).

$\hat{\theta'}(S\hat{E}[\hat{\theta'}])$, probability of tiger presence on a replicate, given presence on previous replicate = 0·730 (0·0857).

$\hat{p}_t(S\hat{E}[\hat{p}_t])$, the probability of detecting a tiger sign on a replicate, given presence of tigers on the replicate = 0·17 (0·168).

$\hat{\psi}(S\hat{E}[\hat{\psi}])$, the total fraction of area occupied by tigers in MMTL = 0·665 (0·0857).

Thus, of the 21 167 km^2 potential tiger habitat available in MMTL, we estimate that tigers actually occupied 66%, or an area of 14 076 km^2 (SE = 1814 km^2). In contrast, a naïve estimate derived from the traditional 'presence-versus-absence' approach is only 7 537 km^2 and underestimated true occupancy by *c.* 47%.

As hypothesized, tiger distribution appears to be patchy across MMTL because of ecological as well as management-related factors. Therefore, we examined the β coefficient values for different covariates we expected to influence tiger habitat occupancy. These coefficient values/signs, which express the strength/direction of their expected influence (MacKenzie *et al.* 2006), are reported in Table 3. The β coefficient estimates from 10 plausible models (Table 2) were similar. Relative density of all prey, and larger prey, and the proportion of available tiger habitat were key determinants of tiger presence, indicated by positive β values associated with these covariates (Table 3). Moreover, relative intensity of livestock presence, our surrogate measure of human disturbance, proved to be a negative

Table 3. Model-specific *β* coefficient estimates for covariates determining tiger occupancy ψ in Malenad–Mysore Tiger Landscape (MMTL). Models were used to determine tiger occupancy in MMTL. Number of sites = 205. Covariates considered all prey (AP), large prey (LP), livestock (LVS) and proportion of available habitat (PH)

Model	$\hat{\beta}_0(S\hat{E}[\hat{\beta}_0])$	$\hat{\beta}_{AP}(S\hat{E}[\hat{\beta}_{AP}])$	$\hat{\beta}_{LP}(S\hat{E}[\hat{\beta}_{LP}])$	$\hat{\beta}_{LVS}(S\hat{E}[\hat{\beta}_{LVS}])$	$\hat{\beta}_{PH}(S\hat{E}[\hat{\beta}_{PH}])$
ψ (AP + PH), θ (.), θ′ (.), p_t (LP + LVS + PH)	−2·347 (1·173)	2·181 (1·162)	–	–	3·089 (1·469)
ψ (LP + PH), θ (.), θ′ (.), p_t (LP + LVS + PH)	−2·257 (1·156)	–	2·109 (1·163)	–	3·057 (1·473)
ψ (LVS + PH), θ (.), θ′ (.), p_t (LP + LVS + PH)	0·108 (1·011)	–	–	−1·561 (0·931)	2·928 (1·589)
ψ (PH), θ (.), θ′ (.), p_t (LP + LVS + PH)	−0·712 (0·963)	–	–	–	3·077 (1·747)
ψ (AP + LVS + PH), θ (.), θ′ (.), p_t (LP + LVS + PH)	−1·325 (1·569)	1·527 (1·343)	–	−1·002 (1·020)	2·974 (1·439)
ψ (LP + LVS + PH), θ (.), θ′ (.), p_t (LP + LVS + PH)	−1·212 (1·556)	–	1·431 (1·349)	−1·034 (1·022)	2·956 (1·444)
ψ (.), θ (.), θ′ (.), p_t (LP + LVS + PH)	1·218 (0·442)	–	–	–	–
ψ (AP), θ (.), θ′ (.), p_t (LP + LVS + PH)	−0·244 (0·917)	2·009 (1·309)	–	–	–
ψ (LVS), θ (.), θ′ (.), p_t (LP + LVS + PH)	1·815 (0·677)	–	–	−1·537 (1·024)	–
ψ (LP), θ (.), θ′ (.), p_t (LP + LVS + PH)	−0·201 (0·908)	–	1·986 (1·326)	–	–

© 2011 The Authors. Journal of Applied Ecology © 2011 British Ecological Society,

influence on local tiger presence (with negative β values; Table 3).

Given variations in the densities of tigers and prey at some sites (Karanth *et al.* 2004) and effectiveness of protection across MMTL, we expected substantial variations in site-level tiger abundances and consequently in site-level probabilities of tiger occupancy.

We note that the two covariates contributed significantly to the replicate-level variation in sign detection probability p_t (ΔAIC = 50·95 between the top-ranked model in Table 1 and the model that does not have a covariate model structure on p_t). However, evidence for the effect of these covariates on cell-level tiger occupancy ψ was not as strong (ΔAIC = 1·57 between the top model in Table 2 and the model that does not have a covariate model structure on ψ). Furthermore, the covariate relationships were all in predicted directions, suggesting that these covariates have a strong influence on cell-specific tiger abundance. If these variations in cell-specific tiger abundance are the key determinant of variations in sign detection probabilities, then the estimated cell-specific occupancy rates could potentially be linked to variations in relative tiger abundance using some of the new occupancy models proposed by Royle & Nichols (2003) or Nichols *et al.* (2007). We also graphically mapped these variations in cell-specific occupancy probabilities across the entire landscape (Fig. 3).

Discussion

OCCUPANCY MODELS FOR CARNIVORE SURVEYS

To address increasing threats to large carnivores (Ceballos *et al.* 2005), substantial investments are being made on landscape-scale sign surveys. Although these surveys use sophisticated GIS analyses (tigers: Wikramanayake *et al.* 2005; jaguars: Sanderson *et al.* 2002; lions: Bauer & Van Der Merwe 2004), they are fundamentally rooted in the traditional 'presence-versus-absence' approach to inference. Our study showed that such naïve estimates of occupancy can be substantially biased downward because detection parameters associated with the actual sampling process not being reliably estimated (in our case θ, θ′, and p_t being < 1). Our results reaffirm the value of using occupancy modelling approach to explicitly deal with the difficult problem of imperfect detections. Similarly, development of a specific probabilistic model of the process of searching for tiger signs along trails (Hines *et al.* 2010) proved important. Occupancy models using spatial replicates and incorporating *a priori* hypotheses generated results that could be reasonably interpreted in terms of ecological or management factors (see Results).

Overall, we demonstrate the benefits of designing carnivore sign surveys based on species biology (e.g. tiger home range size and movements on trails) as well as on field survey protocols (e.g. distances walked and spatial coverage of cells). By modelling just three key covariates (available habitat, prey density and human disturbance levels), we were able to provide reasonable explanations for the patterns of tiger distribution. We note that tiger surveys sometimes sacrifice ecological clarity and model parsimony by fitting many covariates relying on omnibus statistical tools such as regressions (e.g. Jhala, Gopal & Qureshi 2008).

MEASURING TIGER DISTRIBUTION AND ABUNDANCE

Local tiger presence is strongly governed by prey density and protection effectiveness, which cannot be remotely sensed. Landscape analyses not augmented by such field data (e.g. Dinerstein *et al.* 2006; Ranganathan *et al.* 2008) therefore have inherent limitations for modelling tiger distributions. Site occupancy estimates from our models permit rapid field assessment of tiger population status across large landscapes (Fig. 3). Temporal extensions of these models (MacKenzie *et al.* 2003) can monitor tiger range contractions or expansions. Overall, we suggest that occupancy modelling offers a robust alternative to previous tiger sign survey methods, such as effort-adjusted encounter rates (Johnsingh *et al.* 2004), adjusted track count-based estimates (Hayward *et al.* 2002) and extrapolations from sign encounter rates (Smith *et al.* 1999a,b; Wikramanayake *et al.* 2005; Carroll & Miquelle 2006; Barlow *et al.* 2008).

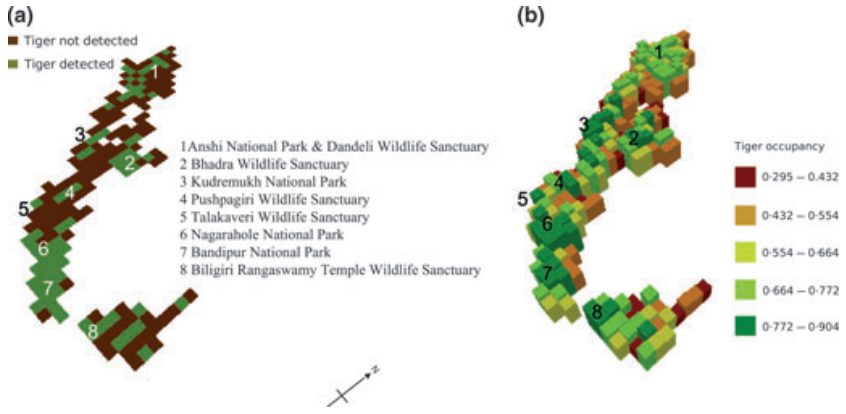


Fig. 3. Patterns of spatial variation in probability of site occupancy of tigers in the Malenad–Mysore Landscape of Karnataka (MMTL). (a) Based on naïve estimate derived from the traditional presence-versus-absence approach and (b) based on occupancy analysis under the model developed by Hines *et al.* (2010).

We note, however, that recent reports by Jhala, Gopal & Qureshi (2008) suggest that spatially replicated sign surveys can estimate tiger distribution reliably without explicitly modelling imperfect detections. On the contrary, our results show replicate-level tiger sign detection probability can be as low as 17%, which, if ignored, would lead to underestimates of occupancy. We find these differences puzzling and worth further exploration.

Our survey did not fully meet the assumptions of abundance models of occupancy (Royle & Nichols 2003; Nichols *et al.* 2007; Conroy *et al.* 2008). We are currently exploring further applications of such abundance models of occupancy, which can be integrated with rigorous site-specific tiger density estimates from camera trapping to reliably estimate tiger numbers across large landscapes.

IMPLICATIONS FOR THE MANAGEMENT OF WILD TIGER POPULATIONS

Prey depletion by human hunters was first hypothesized as a critical factor driving local declines of tiger populations by Karanth & Stith (1999). Although alternative models hypothesize a greater influence of direct hunting (Chapron *et al.* 2008), the prey depletion model has since then been supported by rigorous field studies of tiger:prey density ratios (Karanth *et al.* 2004), as well as a long-term population dynamics study (Karanth *et al.* 2006). This study finds support for the models proposing that ungulate prey densities are strong determinants of tiger presence at the landscape scale also (Karanth *et al.* 2010).

Our results (Figs 1 and 3) show that tigers have persisted better in MMTL compared with far more extensively forested, sparsely populated and economically underdeveloped landscapes in India, a pattern earlier reported by Jhala, Gopal & Qureshi (2008). Their persistence appears to be dependent upon a few effectively protected 'source populations' (Figs 1 and 3), which are reproducing above replacement levels (Karanth *et al.* 2006). A recent global conservation analysis (Walston *et al.* 2010) shows that about 70% of wild tigers now survive in source populations occupying just 6% of the habitat. These sources are currently identified based on anecdotal evidence. Our survey approach can rapidly and objectively identify such 'sources' for targeting conservation efforts.

Walston *et al.* (2010) are concerned that the widely promoted 'landscape approach' to tiger conservation (e.g. Wikramanayake *et al.* 2005; Dinerstein *et al.* 2006) does not adequately prioritize protection of source populations and therefore may not help in arresting the current tiger decline. The proponents of the 'landscape approach', however, insist that conservation investments should continue to target wider regions. The population assessment approach we present provides a basis for reliably monitoring tiger meta-populations by integrating site-level, intensive, capture–recapture surveys with cost-effective, landscape-scale sign surveys. Furthermore, this approach of confronting alternative ecological/management predictions with survey data may also prove useful for future adaptive management (Walters 1986; Williams, Nichols & Conroy 2002) of wild tiger populations. We also believe these methods are relevant to efficient monitoring of spatial distributions and abundances of many other wide-ranging large carnivores whose signs are relatively easily detected.

Acknowledgements

We are grateful to the Wildlife Conservation Society, New York, and USGS Patuxent Wildlife Research Center, Laurel, MD, for supporting our research work. Additionally, we are indebted to the following donors for funding support: Save the Tiger Fund of National Fish and Wildlife Foundation, and Rhino & Tiger Conservation Fund of US Fish and Wildlife Service, Washington, DC. In India, we acknowledge the facilitation of the study by Karnataka State Forest Department, Centre for Wildlife Studies, Bangalore, and National Tiger Conservation Authority, New Delhi. We thank J. E. Hines, J. A. Royle, B. K. Williams, J. Krishnaswamy, M. Delampady for useful discussions and D. Jathanna, P. Singh and N. Patil for analytical support. We thank the numerous field assistants and volunteers who enthusiastically participated in our field surveys and acknowledge the superb leadership of D. V. Girish, B. Hegde, N. Jain, K. Rayar, K. Nag, C. Krishna and A. Kadam in leading the field surveys. We also thank three reviewers and the editors for comments that greatly improved the manuscript.

References

Barlow, A.C.D., Ahmed, M.I.U., Rahman, M.M., Howlader, A., Smith, A.C. & Smith, J.L.D. (2008) Linking monitoring and intervention for improved management of tigers in the Sundarbans of Bangladesh. *Biological Conservation*, **141**, 2032–2040.

Bauer, H. & Van Der Merwe, S. (2004) Inventory of free-ranging lions *P. leo* in Africa. *Oryx*, **38**, 26–31.

Bevington, P.R. & Robinson, D.K. (2003) *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill, New York.

Boyce, M.S. & McDonald, L.L. (1999) Relating populations to habitats using resource selection functions. *Trends in Ecology and Evolution*, **14**, 268–272.

Burnham, K.P. & Anderson, D.R. (2002) *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*. Springer-Verlag, New York.

Carroll, C. & Miquelle, D.G. (2006) Spatial viability analysis of Amur tiger *Panthera tigris altaica* in the Russian Far East: the role of protected areas and landscape matrix in population persistence. *Journal of Applied Ecology*, **43**, 1056–1068.

Ceballos, G., Ehrlich, P.R., Soberón, J., Salazar, I. & Fay, J.P. (2005) Global mammal conservation: what must we manage? *Science*, **309**, 603–607.

Chapron, G., Miquelle, D.G., Lambert, A., Goodrich, J.M., Legrandre, S. & Clobert, J. (2008) The impact on tigers of poaching versus prey depletion. *Journal of Applied Ecology*, **45**, 1667–1674.

Conroy, M.J., Runge, J.P., Barker, R.J., Schofield, M.R. & Fonnesbeck, C.J. (2008) Efficient estimation of abundance for patchily distributed populations via two-phase, adaptive sampling. *Ecology*, **89**, 3362–3370.

Das, A., Krishnaswamy, J., Bawa, K.S., Kiran, M.C., Srinivas, V., Kumar, N.S. & Karanth, K.U. (2006) Prioritization of conservation areas in the Western Ghats, India. *Biological Conservation*, **133**, 16–31.

Dinerstein, E., Loucks, C., Wikramanayake, E., Ginsberg, J., Sanderson, E., Seidensticker, J., Forrest, J., Bryja, G., Heydlauff, A., Klenzendorf, S., Leimgruber, P., Mills, J., O'Brien, T.G., Shrestha, M., Simons, R. & Songer, M. (2006) The fate of wild tigers. *BioScience*, **57**, 508–514.

Directorate of Census Operation. (2004) *Primary Census Abstract for Andhra Pradesh, Karnataka & Lakshadweep*. Technical Report Vol. X. 00-063-2001-Cen-CD. Office of the Registrar General and Census Commissioner, Government of India, New Delhi.

Efron, B. (1982) *The jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Engler, R., Guisan, A. & Rechsteiner, L. (2004) An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology*, **41**, 263–274.

Gillespie, D.T. (1992) *Markov Processes: An Introduction to Physical Scientists*. Academic Press, London, UK.

Gu, W. & Swihart, R.K. (2004) Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological Conservation*, **116**, 195–203.

Hayward, G.D., Miquelle, D.G., Smirnov, E.N. & Nations, C. (2002) Monitoring Amur tiger populations: characteristics of track surveys in snow. *Wildlife Society Bulletin*, **30**, 1150–1159.

Hines, J.E. (2006) PRESENCE2 – Software to estimate patch occupancy and related parameters. Version 2.4. USGS-PWRC. http://www.mbr-pwrc.usgs.gov/software/presence.html.
Hines, J.E., Nichols, J.D., Royle, J.A., MacKenzie, D.I., Gopalaswamy, A.M., Kumar, N.S. & Karanth, K.U. (2010) Tigers on trails: occupancy modeling for cluster sampling. *Ecological Applications*, **20**, 1456–1466.
Hirzel, A.H., Hausser, J., Chessel, D. & Perrin, N. (2002) Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology*, **83**, 2027–2036.
Jhala, Y.V., Gopal, R. & Qureshi, Q. (2008) *Status of the Tigers, Co-predators, and Prey in India*. National Tiger Conservation Authority, Government of India, New Delhi and Wildlife Institute of India, Dehradun.
Jhala, Y.V., Qureshi, Q. & Gopal, R. (2011) Can the abundance of tigers be assessed from their signs? *Journal of Applied Ecology*, **48**, 14–24.
Johnsingh, A.J.T., Ramesh, K., Qureshi, Q., David, A., Goyal, S.P., Rawat, G.S., Rajapandian, K. & Prasad, S. (2004) *Conservation Status of Tiger and Associated Species in the Terai Arc Landscape, India*. Final Technical Report, Wildlife Institute of India, Dehradun, pp. viii + 110.
Karanth, K.U. & Stith, B.M. (1999) Prey depletion as a critical determinant of tiger population viability. *Riding the Tiger: Tiger Conservation in Human Dominated Landscapes* (eds J. Seidensticker, S. Christie & P. Jackson), pp. 100–113. Cambridge University Press, Cambridge, UK.
Karanth, K.U. & Sunquist, M.E. (2000) Behavioral correlates of predation by tiger, leopard and dhole in Nagarahole, India. *Journal of Zoology*, **250**, 255–265.
Karanth, K.U., Nichols, J.D., Kumar, N.S., Link, W.A. & Hines, J.E. (2004) Tigers and their prey: predicting carnivore densities from prey abundance. *Proceedings of the National Academy of Sciences, USA*, **101**, 4854–4858.
Karanth, K.U., Nichols, J.D., Kumar, N.S. & Hines, J.E. (2006) Assessing tiger population dynamics using photographic capture-recapture sampling. *Ecology*, **87**, 2925–2937.
Karanth, K.K., Nichols, J.D., Karanth, K.U., Hines, J.E. & Christensen, N.L. (2010) The shrinking Ark: patterns of mammal extinctions in India. *Proceedings of the Royal Society-B*, **227**, 1971–1979.
Kendall, W.L. & White, G.C. (2009) A cautionary note on substituting spatial subunits for repeated temporal sampling in studies of site occupancy. *Journal of Applied Ecology*, **46**, 1182–1188.
Krishnaswamy, J., Kiran, M.C. & Davande, S. (2003) *Eco-climatic Vegetation Classification for the Western Ghats Using Multi-season NDVI*. Ashoka Trust for Research in Ecology and the Environment, Bangalore.
MacKenzie, D.I. & Royle, J.A. (2005) Designing efficient occupancy studies: general advice and tips on allocation of survey effort. *Journal of Applied Ecology*, **42**, 1105–1114.
MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. & Langtimm, C.A. (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**, 2248–2255.
MacKenzie, D.I., Nichols, J.D., Knutson, M.G. & Franklin, A.B. (2003) Estimating site occupancy, colonization and local extinction when species are detected imperfectly. *Ecology*, **84**, 2200–2207.
MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Hines, J.E. & Bailey, L.L. (2006) *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. Elsevier, San Diego.
Madhusudan, M.D. & Mishra, C. (2003) Why big, fierce animals are threatened: conserving large mammals in densely populated landscapes. *Battles Over Nature: Science and the Politics of Conservation* (eds V. Saberwal & M. Rangarajan), pp. 31–55. Permanent Black, New Delhi.
Magoun, A.J., Ray, J.C., Johnson, D.S., Valkenburg, P., Dawson, F.N. & Bowman, J. (2007) Modeling wolverine occurrence using aerial surveys of tracks in snow. *Journal of Wildlife Management*, **71**, 2221–2229.
Manley, P.N., Zielinski, W.J., Schlesinger, M.D. & Mori, S.R. (2004) Evaluation of a multiple-species approach to monitoring species at the ecoregional scale. *Ecological Applications*, **14**, 296–310.
Miquelle, D.G., Smirnov, E.N., Merrill, T.W., Myslenkov, A.E., Quigley, H.B., Hornocker, M.G. & Schleyer, B. (1999) Hierarchical spatial analysis of Amur tiger relationships to habitat and prey. *Riding the Tiger: Tiger Conservation in Human Dominated Landscapes* (eds J. Seidensticker, S. Christie & P. Jackson), pp. 71–99. Cambridge University Press, Cambridge, UK.
Miquelle, D.G., Stephens, P.A., Smirnov, E.N., Goodrich, J.M., Zaumyslova, O.Y. & Myslenkov, A.E. (2005) Tigers and wolves in the Russian Far East: competitive exclusion, functional redundancy and conservation implications. *Large Carnivores and the Conservation of Biodiversity* (eds J.C. Ray, J. Berger, K.H. Redford & R. Steneck), pp. 179–207. Island Press, Washington, DC.
Mondol, S., Karanth, K.U., Kumar, N.S., Gopalaswamy, A.M., Andheria, A. & Ramakrishnan, U. (2009) Evaluation of non-invasive genetic sampling methods for estimating tiger population size. *Biological Conservation*, **142**, 2350–2360.
Nichols, J.D., Hines, J.E., MacKenzie, D.I., Seamans, M.E. & Gutierrez, R.J. (2007) Occupancy estimation and modelling with multiple states and state uncertainty. *Ecology*, **88**, 1395–1400.
Ranganathan, J., Chan, K.M.A., Karanth, K.U. & Smith, J.L.D. (2008) Where can tigers persist in the future? A landscape-scale, density-based population model for the Indian subcontinent. *Biological Conservation*, **141**, 67–77.
Rayar, K. (2010) *Assessing Potential Tiger Habitats in Cauvery Wildlife Sanctuary, Karnataka Using Occupancy Modeling Approaches*. MS thesis, Tata Institute of Fundamental Research, National Centre for Biological Sciences, Bangalore, India.
Royle, J.A. & Dorazio, R.M. (2008) *Hierarchical Models and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities*. Academic Press, New York.
Royle, J.A. & Nichols, J.D. (2003) Estimating abundance from repeated presence–absence data or point counts. *Ecology*, **84**, 777–790.
Sanderson, E.W., Redford, K.H., Chetkiewicz, C.-L.B., Medellin, R.A., Rabinowitz, A., Robinson, J.G. & Taber, A.B. (2002) Planning to save a species: the jaguar as a model. *Conservation Biology*, **16**, 58–72.
Smith, J.L.D., McDougal, C. & Miquelle, D. (1989) Scent marking in free ranging tigers, *Panthera tigris*. *Animal Behavior*, **37**, 1–10.
Smith, J.L.D., McDougal, C., Ahearn, S.C., Joshi, A. & Conforti, K. (1999a) Metapopulation structure of tigers in Nepal. *Riding the Tiger: Tiger Conservation in Human-dominated Landscapes* (eds J. Seidensticker, S. Christie & P. Jackson), pp. 176–189. Cambridge University Press, Cambridge.
Smith, J.L.D., Tunhikorn, S., Tanhan, S., Simcharoen, S. & Kanchanasaka, B. (1999b) Metapopulation structure of tigers in Thailand. *Riding the Tiger: Tiger Conservation in Human-dominated Landscapes* (eds J. Seidensticker, S. Christie & P. Jackson), pp. 166–175. Cambridge University Press, Cambridge.
Sunquist, M.E. (1981) Social organization of tigers (*Panthera tigris*) in Royal Chitwan National Park, Nepal. *Smithsonian Contributions to Zoology*, **336**, 1–98.
Tyre, A.J., Possingham, H.P. & Lindenmayer, D.B. (2001) Inferring process from pattern: can territory occupancy provide information about life history parameters? *Ecological Applications*, **11**, 1722–1737.
Walston, J., Robinson, J.G., Bennett, E.L., Breitenmoser, U., da Fonseca, G.A.B., Goodrich, J. *et al.* (2010) Bringing the tiger back from the brink—the six percent solution. *PLoS Biology*, **8**, e1000485 doi: 10.1371/journal.pbio.1000485.
Walters, C.J. (1986) *Adaptive Management of Renewable Resources*. Macmillan, New York.
Wikramanayake, E., Mcknight, M., . Dinerstein, E., Joshi, A., Gurung, B. & Smith, D. (2005) Designing a conservation landscapes for tigers in human-dominated environments. *Conservation Biology*, **18**, 839–844.
Williams, B.K., Nichols, J.D. & Conroy, M.J. (2002) *Analysis and Management of Animal Populations*. Academic Press, San Diego.

Received 10 October 2010; accepted 7 April 2011
Handling Editor: Johan du Toit

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. A general computing expression adapted from Hines *et al.* (2010) for modelling detection histories, h_j, obtained for detections of tiger sign along a continuous trail, composed of spatial replicates.

Appendix S2. Estimation of mean and standard deviation of the fraction of total Malenad–Mysore Tiger Landscape (MMTL) occupied by tigers.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Appendix H

Counting India's tigers reliably

LETTERS

edited by Jennifer Sills

Counting India's Wild Tigers Reliably



THE INDIAN GOVERNMENT REPORTED A 16% increase in tiger numbers over the past 4 years (News of the Week, Around the World item "Tiger numbers up? Maybe," 1 April, p. 18). This implies an average increase of 49% in local tiger densities, despite the reported range contraction of 22%. Yet these assertions cannot be verified because details of tiger photo-captures at sampled locations, as well as of spatial extrapolations from these data, are incomplete (*1*–*3*).

Reported tiger numbers were based on calibrations of tiger sign encounter rates against estimated local tiger densities (*2*), but the recently released values of correlation coefficients were much higher than have been reported previously (*4*, *5*). Moreover, the extrapolation of tiger numbers to wider regions is reportedly based on standard methods of sampling and estimation (*1*), but it is not clear from reports (*1*–*3*) whether the survey protocols used actually match these standard practices (*6*, *7*).

A recent global analysis (*8*) showed that 70% of wild tigers survive in 42 "source populations" that occupy a mere 6% of remaining habitat. Although such source populations may suffer annual losses of more than 20%, studies of tiger population dynamics show that high recruitment rates compensate for these losses when there is adequate prey and protection (*9*, *10*). Therefore, future efforts to reverse tiger declines must focus on reliably monitoring tiger numbers, as well as survival and recruitment rates, in these threatened source populations. There is a scientific consensus that monitoring should be conducted annually—within 30 to 45 days to avoid population fluctuations—and cover an area of more than 500 km^2, at intensities of more than 500 trap-nights per 100 km^2 (*11*).

The monitoring protocol for India's national animal requires a major overhaul if it is to generate transparent, reliable measures of tiger conservation successes (or failures) in the future.

K. ULLAS KARANTH,[1,2]* ARJUN M. GOPALASWAMY,[1,3] N. SAMBA KUMAR,[1,4] MOHAN DELAMPADY,[5] JAMES D. NICHOLS,[6] JOHN SEIDENSTICKER,[7] BARRY R. NOON,[8] STUART L. PIMM[9]

[1]Centre for Wildlife Studies, Bangalore, 560042, India. [2]Wildlife Conservation Society–Global Conservation Program, Bronx, NY 10460–1099, USA. [3]Wildlife Conservation Research Unit (WildCRU), The Recanati-Kaplan Centre, Department of Zoology, University of Oxford, Tubney, Abingdon, OX13 5QL, UK. [4]Wildlife Conservation Society–India Program, Bangalore, 560 070, India. [5]Statistics and Mathematics Unit, Indian Statistical Institute, Bangalore Centre, Bangalore, 560059, India. [6]Patuxent Wildlife Research Center, U.S. Geological Survey, Laurel, MD 20708–4017, USA. [7]Smithsonian Conservation Biology Institute, National Zoological Park, Washington, DC 20013–7012, USA. [8]Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO 80521–1474, USA. [9]Nicholas School of the Environment, Duke University, Durham, NC 27708, USA.

*To whom correspondence should be addressed. E-mail: ukaranth@gmail.com

CREDIT: FOTOSEARCH

References

1. Y. V. Jhala, R. Gopal, Q. Qureshi, "Status of the tigers, co-predators, and prey in India" (National Tiger Conservation Authority, Government of India, New Delhi, and Wildlife Institute of India, Dehradun, 2008).
2. Y. V. Jhala, Q. Qureshi, R. Gopal, *J. Appl. Ecol.* **48**, 14 (2011).
3. Ministry of Environment and Forests, "India tiger estimate 2010" (Government of India, National Tiger Conservation Authority and Wildlife Institute of India, 2011).
4. J. E. Hines *et al.*, *Ecol. Appl.* **20**, 1456 (2010).
5. K. U. Karanth *et al.*, *J. Appl. Ecol.*, 10.1111/j.1365-2664.2011.02002.x (2011).
6. K. H. Pollock *et al.*, *Environmetrics* **13**, 105 (2002).
7. D. I. MacKenzie *et al.*, *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence* (Elsevier, San Diego, CA, 2006).
8. J. Walston *et al.*, *PLoS Biol.* **8**, e1000485 (2010).
9. K. U. Karanth, J. D. Nichols, N. S. Kumar, W. A. Link, J. E. Hines, *Proc. Nat. Acad. Sci. U.S.A.* **101**, 4854 (2004).
10. K. U. Karanth, J. D. Nichols, N. S. Kumar, J. E. Hines, *Ecology* **87**, 2925 (2006).
11. A. F. O'Connell *et al.*, *Camera Traps in Animal Ecology: Methods and Analyses* (Springer, Tokyo, 2011).

Practical Implications of Test Anxiety Tools

IN THEIR REPORT "WRITING ABOUT TESTING worries boosts exam performance in the classroom" (Reports, 14 January, p. 211), G. Ramirez and S. L. Beilock showed that letting students write about their worries for 10 minutes before an exam substantially diminishes the link between test anxiety and test performance. Their second study replicates and extends our previous work, in which we showed that letting students write down attributes of successful problem-solvers for 10 minutes diminishes the relationship between cognitive test anxiety and test performance (*1*).

We differ regarding the theoretical interpretation of the relationship and the resulting practical implications. Although Ramirez and Beilock's intervention is not identical to ours, the fundamental mechanisms are similar. Ramirez and Beilock argue that the effect is mediated by a state of worry during the test situation, but they did not test this proposition. In one of our investigations, we tested this idea and found no support. Instead, we found cognitive test anxiety to affect situational task engagement: Students who have high cognitive test anxiety do not engage in the task because they underestimate their probability for success and consequently do not fully engage in solving the problems at hand. This idea

Downloaded from www.sciencemag.org on May 12, 2011

Appendix I

Dispersing tiger makes a point

Conservation news

Dispersing tiger makes a point...

Data on the long-range dispersal of tigers are limited and most of the available information comes from only a few telemetry studies. Long-term camera-trap data over a large region can potentially provide insights into poorly known and less understood aspects of tiger ecology and behaviour. The Wildlife Conservation Society (WCS) through its collaborating partner, the Centre for Wildlife Studies in India, has been using camera trapping to monitor key source populations of tigers in the Malenad-Mysore Tiger Landscape in Karnataka, India (in Nagarahole since 1990, Bhadra since 1997, Bandipur since 1999, Dandeli-Anshi since 2008, and Biligiri Rangaswamy Temple Sanctuary since 2010). This has produced what is possibly the largest photo-database of individual tigers (> 2,000 profiles of 500+ distinct individuals). This database serves many useful purposes, including that of tracking long-range dispersal, tracing the origin of seized tiger-skins, and monitoring mortality. Here we describe two long-range dispersal events that demonstrate the value of tiger-permeable landscapes for the long-term persistence of tiger populations.

A male (BDT-130) first camera-trapped as a 4-month old cub in April 2006 in Bhadra Tiger Reserve was later photographed by Sudhir Shivaram, a wildlife photographer, in October 2006, along with its sibling and mother, in Bhadra, by which time it was almost full-grown. In May 2008 WCS camera-trapped BDT-130 in Dandeli-Anshi Tiger Reserve, indicating the individual had dispersed radially by at least c. 200 km, traversing some of the most densely human-populated forest patches in the Malenad-Mysore Tiger Landscape. The subsequent re-captures of this tiger in April 2009 and April 2010 in Dandeli-Anshi indicate the successful establishment of a territory by this dispersing male.

On 1 May 2011 a male tiger strayed into Gama village in Shimoga Wildlife Division and was subsequently captured by forest officials following complaints of predation on cattle and attempted attacks on humans. This individual was identified by WCS as BPT-241, which was first camera-trapped near Gundre in Bandipur Tiger Reserve in February 2010. The distance between Gama and Gundre is nearly 280 km, which is probably the longest dispersal distance ever recorded by camera trapping. In consultation with WCS scientists BPT-241 was released by the Karnataka Forest Department in the nearby Bhadra Tiger Reserve on 7 May 2011, as this individual was a healthy, injury-free, young dispersing male and the Reserve had increasing numbers of prey but fewer tigers than its carrying capacity. Nearly 100 days after its release this tiger was found dead near a gaur kill with grave injuries sustained from conspecific aggression.

Telemetry studies in Nepal, Russia and India have shown that young tigers disperse from their natal areas at 18–24 months old, searching for vacant territories. Dispersing male tigers are known to cover long distances before establishing new territories whereas dispersing females are likely to establish home ranges close to their mother. However, information on how far individuals travel and under what habitat conditions they disperse is scarce. Furthermore, there are few recorded instances of transients dispersing through human-dominated landscapes.

The recent dispersal events recorded in the Malenad-Mysore Tiger Landscape indicate the remarkable ability of tigers to disperse stealthily long distances from their natal range. These events also suggest it is possible to support, maintain and sustain dispersal corridors using sustainable landscape approaches (*Conservation Biology*, 7, 20–28) that facilitate tiger movements. Understanding the nature and extent of the forest corridors required for tiger dispersal is the key to sustaining tiger populations over large regions, and the utility of the photo-database for monitoring movements emphasizes the role of long-term scientific monitoring in enhancing our understanding of the biological needs of tigers and our ability to conserve them effectively.

Narendra Patil and N. Samba Kumar Wildlife Conservation Society—India Program, 1669, 31st Cross, 16th Main, Bansahankari 2nd Stage, Bengaluru, India, and Centre for Wildlife Studies, Bengaluru, India. E-mail narendra.patil.cws@gmail.com

Arjun M. Gopalaswamy Wildlife Conservation Research Unit, Department of Zoology, University of Oxford, Abingdon, UK

K. Ullas Karanth Wildlife Conservation Society—Global Conservation Program, Bronx, New York, USA

Open trade in Kaiser's spotted newt in South-East Asia

Kaiser's spotted newt *Neurergus kaiseri* is endemic to Iran's southern Zagros Mountains, where it is confined to four streams in a single catchment area. Because of its small range (< 10 km^2) and population size (< 1,000 mature individuals) and a range of threats (habitat loss, droughts, introduction of non-native fish and capture for the pet trade) the species is categorized as Critically Endangered on the IUCN Red List. The threat posed by trade for the international pet market recently resulted in the inclusion of this newt in Appendix I of CITES, precluding all international commercial trade as of June 2010 (http://www.cites.org/eng/cop/15/prop/E-15-Prop-14.pdf). The species is protected in Iran and no export permits have been

© 2011 Fauna & Flora International,

Appendix J

Assessing patterns of Human-Wildlife Conflicts and Compensation around a Central Indian Protected Area

OPEN ACCESS Freely available online

PLOS ONE

Assessing Patterns of Human-Wildlife Conflicts and Compensation around a Central Indian Protected Area

Krithi K. Karanth[1,2,3]*, Arjun M. Gopalaswamy[2,4,5], Ruth DeFries[1], Natasha Ballal[2]

1 Ecology, Evolution, and Environmental Biology, Columbia University, New York, New York, United States of America, **2** Centre for Wildlife Studies, Bangalore, India, **3** Nicholas School of Environment, Duke University, Durham, North Carolina, United States of America, **4** Wildlife Conservation Research Unit, Department of Zoology, University of Oxford, Tubney, United Kingdom, **5** Wildlife Conservation Society-India program, Bangalore, India

Abstract

Mitigating crop and livestock loss to wildlife and improving compensation distribution are important for conservation efforts in landscapes where people and wildlife co-occur outside protected areas. The lack of rigorously collected spatial data poses a challenge to management efforts to minimize loss and mitigate conflicts. We surveyed 735 households from 347 villages in a 5154 km^2 area surrounding Kanha Tiger Reserve in India. We modeled self-reported household crop and livestock loss as a function of agricultural, demographic and environmental factors, and mitigation measures. We also modeled self-reported compensation received by households as a function of demographic factors, conflict type, reporting to authorities, and wildlife species involved. Seventy-three percent of households reported crop loss and 33% livestock loss in the previous year, but less than 8% reported human injury or death. Crop loss was associated with greater number of cropping months per year and proximity to the park. Livestock loss was associated with grazing animals inside the park and proximity to the park. Among mitigation measures only use of protective physical structures were associated with reduced livestock loss. Compensation distribution was more likely for tiger related incidents, and households reporting loss and located in the buffer. Average estimated probability of crop loss was 0.93 and livestock loss was 0.60 for surveyed households. Estimated crop and livestock loss and compensation distribution were higher for households located inside the buffer. Our approach modeled conflict data to aid managers in identifying potential conflict hotspots, influential factors, and spatially maps risk probability of crop and livestock loss. This approach could help focus allocation of conservation efforts and funds directed at conflict prevention and mitigation where high densities of people and wildlife co-occur.

Citation: Karanth KK, Gopalaswamy AM, DeFries R, Ballal N (2012) Assessing Patterns of Human-Wildlife Conflicts and Compensation around a Central Indian Protected Area. PLoS ONE 7(12): e50433. doi:10.1371/journal.pone.0050433

Editor: Brian Gratwicke, Smithsonian's National Zoological Park, United States of America

Received July 18, 2012; **Accepted** October 22, 2012; **Published** December 5, 2012

Copyright: © 2012 Karanth et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: National Science Foundation Grant Number 1029219 supported this research. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exists.

* E-mail: krithi.karanth@gmail.com

Introduction

Reducing conflict between wildlife and people is considered a top conservation priority, particularly in landscapes where high densities of people and wildlife co-occur [1–2]. Increased visibility for conflict incidents may be attributed to actual increase in incidents taking place or just greater reporting by affected local people [3]. Dearth of knowledge about conflict loss and compensation distribution contributes to poor allocation of conservation resources [4–5]. Failure to address emerging issues with conflict losses and distribution of compensation may lead to escalation of tensions between people and wildlife, and promote retaliatory actions leading to extirpations of species [6–7]. Preventing conflict and improving distribution of compensation are important to fostering co-existence in landscapes that surround protected areas and function as critical buffers for wildlife [8–9].

Indian protected areas (PAs) support a huge array of wildlife that are prone to conflict with people. People tolerate some species such as Nilgai *Boselaphus tragocamelus*, Chinkara *Gazzella bennetti* and Blackbuck *Antilope cervicapra* but are less tolerant of other species such as wild pigs *Sus scorfa* and elephants *Elephas maximus* [10–11]. Crop loss is more common than livestock loss, human injury and death [12]. Local residents most often directly bear the costs of living alongside wildlife and may have limited ability to cope with losses [12]. Understanding the factors associated with conflict and where they are likely to occur is important for conservation management of conflicts [2,13].

In this study, we assess and map perceived conflict and compensation distribution to households around Kanha National Park in Central India. Loss refers to probabilities of surveyed households reporting crop raiding or livestock predation. By spatially modeling and mapping perceived conflict and self-reported compensation distribution, we potentially identify household characteristics and practices along with environmental factors that influence conflict and compensation. Data limitations on the timing and locations of actual conflicts and lack of access to detailed records of compensation restrict us to surveys of local people as the primary source of information (see Methods).

In this paper, we examined:

a) What agricultural and demographic factors (e.g., land area, total number of crops grown in a year, cropping months in a year, household members, proportion of men), environmental factors (proximity to PA and water, elevation, forest cover), and mitigation measures (e.g. fencing, lighting,

physical structures, night guards) are associated with households' self-reported conflict?

b) How does loss vary between crop raiding versus livestock predation?

c) Which demographic factors, conflict type, wildlife species, household location and reporting effort are associated with reported compensation received by households?

We expected households in closer proximity to the PA and water, surrounded by forest cover and in lower elevations with greater availability of forage from crops (numbers or types of crops grown in a year by a household and growing season) to be more prone to conflict [4,14–15]. In contrast, we might expect demographic factors (more household members, more men), household characteristics (such as land size or agricultural income) and use of mitigation measures to lower household loss [16]. We also examined reported compensation distribution around Kanha. We expected households reporting conflict incidents, as well as those with more educated members, larger land size, and more livestock to receive better compensation [14]. We also expected compensation distribution to be influenced by species involved, perhaps better for tigers and leopards compared to ungulates and other carnivores [17–19].

Materials and Methods

Study Sites

Kanha National Park (22° 7′–22° 27′N, 80° 26′–81° 3′E), established in 1955, is one of India's most well known Tiger Reserves. Kanha (from here onwards) covers an area of 940 km^2 and is surrounded by an administrative buffer of 1005 km^2 (Fig. 1). Vegetation is comprised of Sal and mixed deciduous forests interspersed with grasslands, and supports carnivores such as tiger *Panthera tigris*, leopard *Panthera pardus* and wild dog *Cuon alpinus*, and herbivores including sambar *Cervus unicolor*, chital *Cervus axis*, barasingha *Cervus duvaucelii* and gaur *Bos gaurus* [9]. Local livelihood activities such as grazing and collection of resources are common occurrences in Kanha although all of these activities are legally prohibited inside Kanha [12]. Human population densities range between 182–195/km^2, and livestock densities range between 65–79/km^2 in the three districts adjoining the PA [9]. Unlike most other Indian PAs, Kanha has an administratively designated buffer and is surrounded by forested patches that are interspersed with agricultural and barren land [9]. In the administrative buffer activities such as grazing, collection of forest resources and all activities are permitted. The presence of a buffer provides an opportunity to compare conflict experienced and compensation effectiveness among households within and outside the buffer.

Social Surveys and Interviews

Trained assistants conducted 735 structured and open-ended surveys with households selected from 347 villages surrounding Kanha (Table 1). Although there are households inside the park, we only surveyed households outside the park including within the administrative buffer. To enable systematic spatial sampling, we used a grid based approach to sample 5154 km^2 area by placing 218 grid cells (13 km^2 in size) in a 20 km radius outside the PA (Fig. 1). We selected a 20 km radius as a reasonable distance within which we would expect wildlife, particularly ungulates, from the PA to travel outside [20, Karanth et al. unpublished]. Due to the interest in a large geographic area, we sampled 50% of the grid cells in a checkerboard pattern (area of 2319 km^2), with 193 cells (we excluded 9 cells that were forested without villages, and 16 cells that were logistically inaccessible). Survey of India topographic maps, imagery and Google Earth were used to digitize 632 villages around Kanha (v6.1). Our goal was to sample at least 60% of villages in a cell and survey at least one household from each village. In each village we opportunistically selected and approached individual households. In some cells that had just 1 or 2 villages, we surveyed more than one household in such villages but ensured that these households were located as spatially apart as possible. We were able to opportunistically sample an average of 4 villages in each grid cell (ranging from 1–5 villages in a cell). We approached adult male and female respondents and they were questioned about household demographic and socio-economic characteristics and mitigation measures (use of fencing, lighting, guard animals etc., Table 1). Respondents were specifically questioned about recent (2010 and 2011) experience with all conflict incidents (crop raiding, livestock predation and human injury or death), as well as compensation reported and received. All surveys were conducted in October 2011 in Hindi and recorded responses were translated to English. All survey protocols were approved by Columbia University's IRB.

We also collected official year wise summary records of total compensation paid out to individual households by the authorities for crop and livestock loss for the years 2009–2011. These records only cover households located inside the PA and administrative buffer. Compensation by the park authorities is determined by the geographic location of households. Households apply to either territorial revenue or forest divisions in which they are located and the authorities compensate those households filing within their administrative division (Karanth et al. unpublished). Typically the amount of compensation provided is determined by officials assessing the extent of damage and do not differentiate among species (Shukla pers comm. 2012) and do not differentiate among species.

Variable and Model Selection

We used model selection to identify factors associated with self-reported household incidents of crop raiding, livestock predation and compensation distribution around Kanha. The variables chosen are in Table 2. To avoid colinearity of variables used in the model, we computed Pearsons' correlation coefficients for all pairs of variables. High correlation (>0.5) resulted in elimination of some variables [used distance to PA variable over distance to closest forest patch, (Table S1A–B) [21–22]. To permit direct comparison of estimated coefficient values, we scaled the regression inputs by dividing by two standard deviations [23].

Our models represented different hypotheses about factors potentially influencing reported household loss. We defined an a priori set of candidate models, assessed model fit and identified the variables significantly associated with crop and livestock loss using the Corrected Akaike's Information Criterion (AICc) as the model selection tool since the ratio of the number of data points to the number of structural parameters in our model set was, on average less than 40 [24] (Table S2). Additionally, the AICc criterion is also known to perform well under some degree of correlation among explanatory variables [24]. AICc weights represented the relative measure of appropriateness of a given model relative to the entire model set and rank the models (including a 'global model' that includes all explanatory variables of interest). The top model which provided the most optimal trade-off between model parsimony and fit, in our goal to assess association, and closely competing top models (a subset of models accounting for >95% of the model uncertainty) was used to estimate probabilities of crop raiding, livestock predation and compensation distribution by multi-model inference via model averaging [24].

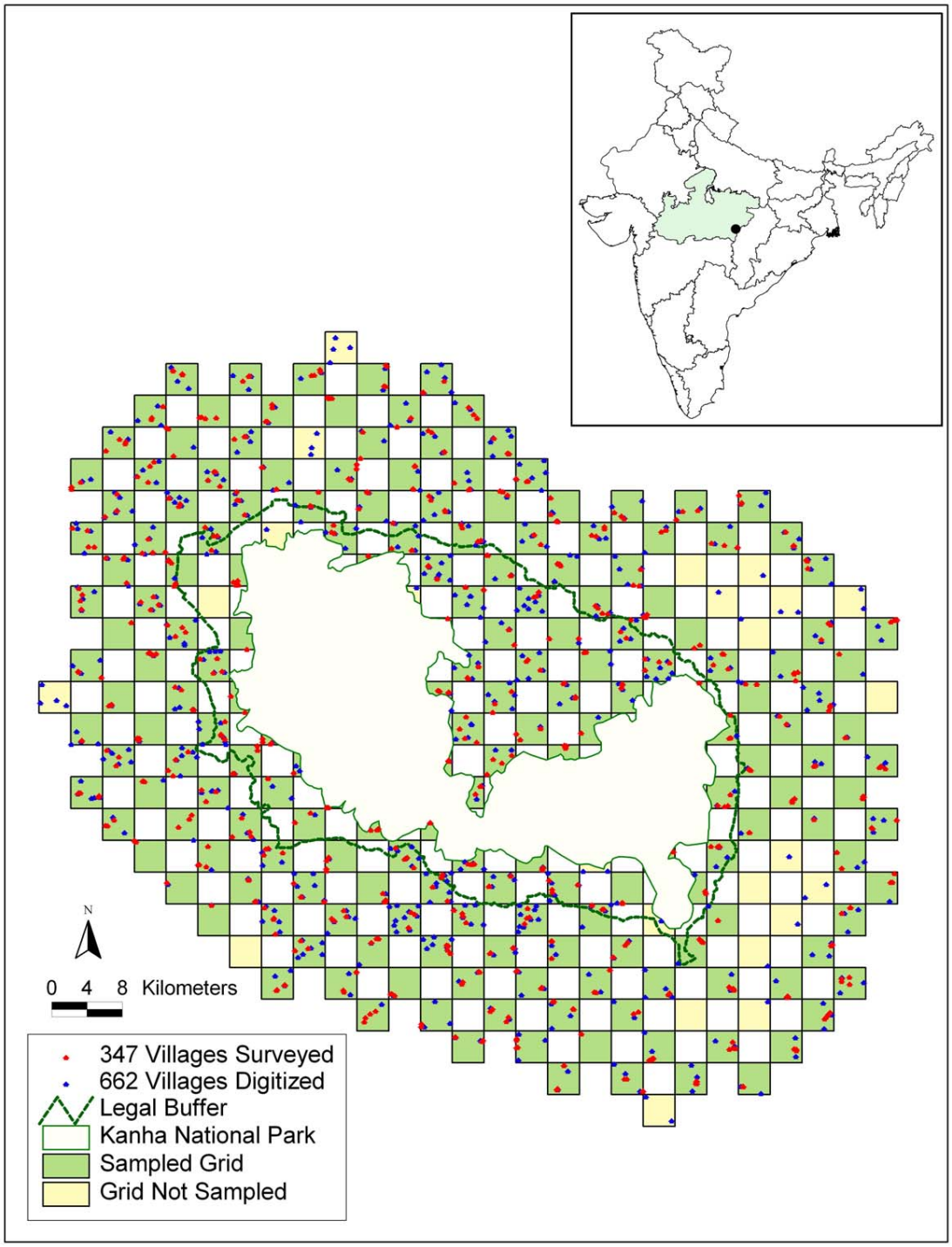


Figure 1. Sampling design and households surveyed around Kanha National Park.
doi:10.1371/journal.pone.0050433.g001

Modeling Household Risk of Crop-raiding and Livestock Predation

We defined household loss as a household reporting recent experiences with crop raiding or livestock predation (in 2010 and 2011). We chose to ask respondents only about the most recent year as we believed that respondents might forget or confuse incidents over multiple years. Since our responses were binary, we fit logistic regression models to model household conflict loss.

We modeled self-reported household crop raiding and livestock predation as a function of environmental factors. The environmental factors included distance to PA, distance to water, elevation and percentage of forest cover within 3 km of household for each household (data derived based on 9). The selection of

Table 1. Household characteristics of 735 surveyed households around Kanha.

Characteristics	Sub-characteristics	Household details
Villages Sampled		347
Number of people in HH (average)		6
Livestock* ownership (average/hh)		7
Annual agricultural income (average/hh)		Rs 12070
Distance to park in km (average)		10.1 (0.01–22.2)
Compensation		61% received Rs 1000 to 5000
Top-ranked crop raiding species	a. Wild Pig	48%
	b. Chital	17%
	c. Hanuman Langur	14%
	d. Rhesus Macaque	7%
	e. Jackal	5%
Income loss from crop raiding		Average Rs 4324, Maximum Rs 175000
Top ranked predators	a. Jackal	29%
	b. Wolf	28%
	c. Tiger	22%
	d. Leopard	17%
Income loss from predation		Rs 1078 (maximum Rs 75,000)
HH visited by authorities		18% (67% within a week, 17% within a month and 16% after a month)

doi:10.1371/journal.pone.0050433.t001

3 km buffer around a household represented a reasonable estimate for home ranges of crop raiding wildlife in tropical forests ([20] and references therein). We also examined the influence of agricultural factors (number of crop types grown in a year, number of cropping months, land area owned by households reported in the surveys). We examined demographic characteristics (number of household members, proportion of men, gender of respondent). We also compared household use of different mitigation efforts: lighting versus fencing versus guard animals. We modeled reported compensation received by households as a function of individual respondents factors (gender, age), household characteristics (land size, total livestock owned), conflict type (crop loss, livestock loss, human injury, human death), reporting effort by households to authorities, wildlife species, proximity to the PA and location inside administrative buffer.

We constructed 31 a priori models representing our hypotheses about causes of 'crop-raiding' (Table S2), including a global model with 18 explanatory variables. We modeled household livestock predation as a function of 20 variables represented by 30 models (Table S2). Similarly, we proposed 36 models of compensation distribution using 22 variables (Table S3). We computed the weighted frequency of each variable among all the models to assess the relative importance of each variable. For each surveyed household, we estimated crop loss, livestock loss and compensation access probabilities for all the candidate models. The weighted average of these individual model estimates, weighted by the Akaike weights [24], was used to generate overall estimates for individual households crop loss, livestock loss and compensation access.

Spatial Mapping of Conflict Risk and Compensation Distribution

We derived estimated probabilities for crop loss, livestock loss and compensation access for every household from each individual model from the entire model set and these values are then weighted to derive model-averaged estimates for each household. We used ordinary kriging to estimate crop loss, livestock loss and compensation distribution probabilities in non-sampled areas. Kriging is a geostatistical interpolation technique that permits inferences about a parameter of interest (in our case, the probability of conflict or compensation) in non-sampled areas by using information available in sampled locations and accounting for uncertainty as the distance between spatial locations increases. The approach produces a semivariogram that describes the spatial correlation between the points. Several models are available for fitting a semivariogram, such as spherical, circular, exponential, Gaussian and linear [25–26]. Based on the Akaike's information criterion (AIC), we fit the spherical model to the observed data, using the Kriging Interpolator 3.2 extension in ArcView Spatial Analyst [27]. In our application, we treated the weighted estimate of conflict loss probability at each sampled location as ''truth'', assuming no variability around the estimate (i.e. no nugget effect) [28], to generate maps of conflict loss probabilities for crop loss and livestock predation, as well as distribution of compensation reported by households around Kanha.

Results

1 Household Characteristics, Agricultural Practices and Reported Crop Loss

Household respondents were largely men (87%), with many (77%) having completed 8th grade or less education. Majority of households (93%) were engaged in agriculture, 88% legally owned land with average land size of 4.3 acres. Households reported growing 12 different crops. The major crops grown were rice, millet and legumes, and the average length of time for crops in the field was 5 months (range 4–12 months, Table 1).

Households reporting crop loss listed 17 species as crop raiders including 10 herbivores, 4 carnivores, 2 primates and peacocks. Animals reported causing the most crop damage were wild pig *Sus*

Table 2. Details on variables collected from surveys and used in models.

Variable	Crop loss	Livestock loss	Compensation
Distance to park	√	√	√
Distance to water	√	√	–
Forest cover	√	√	–
Elevation	√	√	–
Total land area	√	√	√
Gender of respondents	√	√	√
Caste	√	√	√
Total number of people in household	√	√	–
Number of crops	√	–	√
Cropping months	√	–	√
Agriculture land area	√	–	–
Legal agriculture title	√	–	√
Fencing	√	–	–
Night watching	√	√	–
Guard animals	√	–	–
Lighting	√	–	–
Scare devices	√	–	–
Total livestock owned	–	√	√
Grazing goat inside	–	√	–
Grazing cattle inside			
Keeping closer eye on owned animals	–	√	–
Physical structures	–	√	–
Crop loss reported to authorities	–	–	√
Livestock-loss reported to authorities	–	–	√
Human injury reported to authorities	–	–	√
Human death reported to authorities	–	–	√
Household located inside buffer	–	–	√
Any crop raiding mitigation measure	√	–	–
Any livestock predation mitigation measure	–	√	–
Pig related incident	–	–	√
Chital related incident	–	–	√
Langur related incident	–	–	√
Macaque related incident	–	–	√
Jackal related incident	–	–	√
Tiger related incident	–	√	√
Leopard related incident	–	√	√
Jackal related incident	–	√	√
Wolf related incident	–	√	–

doi:10.1371/journal.pone.0050433.t002

scrofa, chital and langur *Semnopithecus entellus* and used multiple mitigation measures (Table 1, Fig. 2). The highest numbers of raiding incidents were reported from September to December, peaking in October. Sixty-four percent of households reported experiencing more than five incidents per year and 32% households reported 2–5 incidents per year.

From modeling factors associated with self-reported household crop loss, we found five top-ranked models with substantial weights (cumulative AICc >0.95, Table 3). As predicted, the total number of crops grown in a year was associated with increased households' loss (positive β coefficient in Table 3) but no other household characteristics were relevant. Although use of mitigation measures, in general, appeared important, no individual measure, such as fencing or lighting, appeared noticeably more important (Table 3). We suspect that the strong positive associations of any general mitigation measure (as indicated by positive beta coefficients in Table 3) are a result of these mitigation measures being put in place because households have historically experienced many incidents of crop raiding. Among the environmental factors distance to PA was associated with decreased crop loss (negative beta coefficient in Table 3). Distance to water was associated with increased

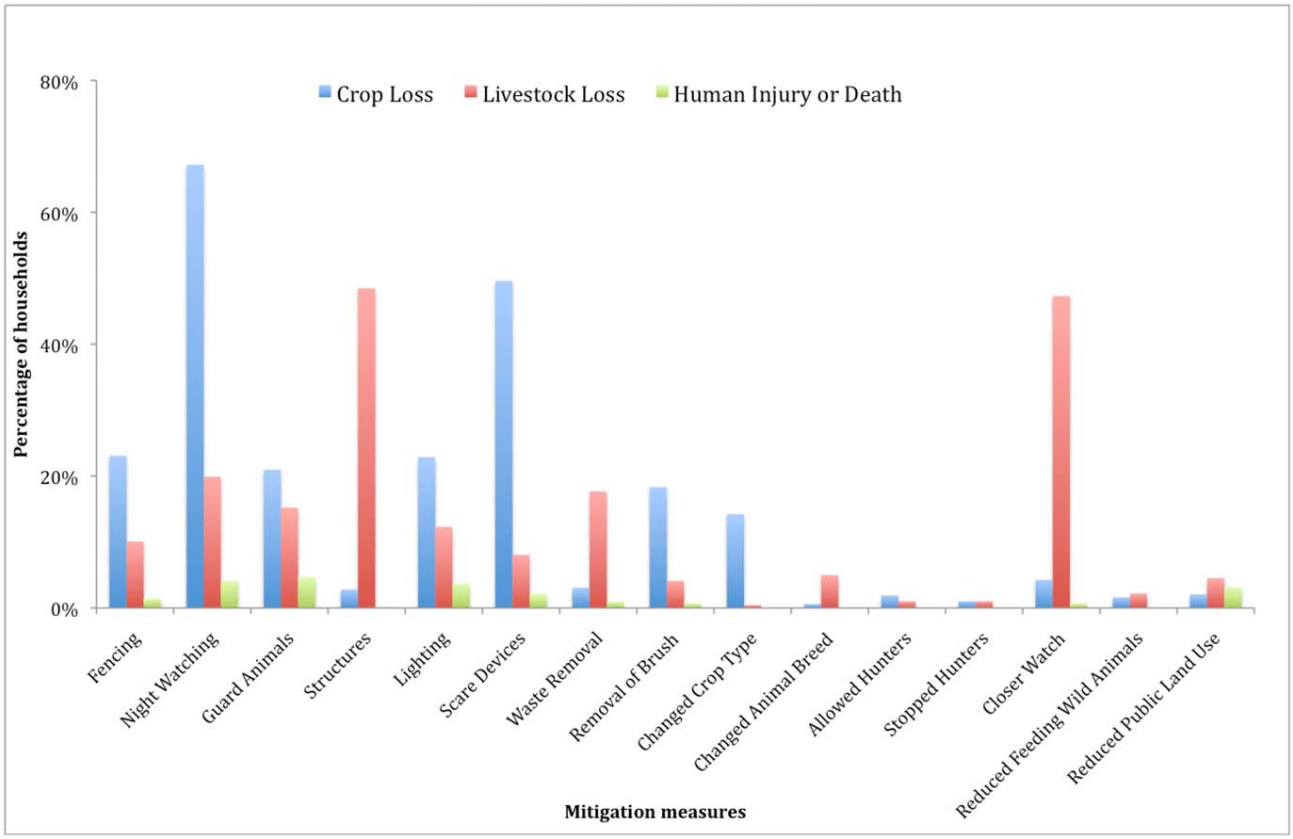


Figure 2. Mitigation measures reported by surveyed households around Kanha National Park.
doi:10.1371/journal.pone.0050433.g002

household loss (positive beta coefficient in Table 4) but high uncertainty in the estimates suggested that this parameter must be interpreted with caution. Contrary to our expectations other modeled factors did not appear to influence crop loss (Table 2 and Table S2), although we recognize that a much larger sample size might have assisted in providing greater support for other explanatory factors as well.

We mapped the probability of crop loss for households around Kanha (Fig. 3). Our modeling indicated that crop loss risk was high in most places, average estimated probabilities of crop loss was 0.93 (range 0.66–0.99). Probabilities of crop loss averaged for households inside the administrative buffer was 0.95 (S.E = 0.004, range 0.77–0.99) and for households outside the buffer was 0.92 (S.E = 0.003, range 0.66–0.99). Spatial modeling based on ordinary kriging (without the nugget effect) suggests that households located closer to the PA have higher risk regardless of location within or outside the administrative buffer. The spatial model locates hot spots of conflict (Fig. 3) and identified high risk particularly around some villages in the buffer (Ramepur, Moharai, Kirsari), and for other villages outside the buffer (Baila, Budhanwara, Maharajpur, Khorja).

2 Livestock Ownership and Perceived Predation Risk

Most (91%) of households owned livestock and ninety percent of animals were stall fed or locally grazed compared to 10% being park grazed. Noticeably households reported an increase in stall-feeding by 7.6% over 10 years and an increase in the number of days their animals were grazed in the last ten years (Table 1). This is perhaps indicative of declining forage areas in close proximity to people's households requiring them to graze livestock further away from home and well inside the PA.

Livestock losses to ten carnivores were reported and most troublesome species were jackal *Canis aureus*, wolf *Canis lupus*, tiger and leopard (Table 1). Livestock losses were distributed across the year but higher from April through July, which are the driest summer months in Kanha (9). Some (34%) households reported more than five incidents per year, 34% households reported 2–5 incidents per year and 31% reported one incident per year. Seven-percent of households reported experiencing injury and 1% reported instances of human death (Table 1).

From modeling factors associate with self-reported household risk to livestock loss outside Kanha, we identified nine top-ranked models (cumulative AICc >0.95, Table 4). These models indicate that as predicted grazing cows inside Kanha was associated with increased loss (positive beta coefficient in Table 4). Among the environmental factors, distance to PA was associated with decreased livestock loss (negative beta coefficient in Table 4). Distance to water was associated with increased household loss (positive beta coefficient in Table 4) but high standard errors suggest caution with interpreting this result. Two mitigation factors night watching and physical structures appeared in top ranked models (Table 3). Only use of physical structures (negative beta coefficient in Table 4) was associated with lower losses but high standard errors suggest that this must be interpreted with caution. Other modeled factors did not appear to be associated with household livestock loss (Table S2).

We mapped risk of livestock predation for households around Kanha (Fig. 4). Our modeling clearly indicates that livestock loss

Table 3. Top models (cumulative weight>0.95) and beta coefficients for predicting household crop loss around Kanha National Park.

Models	30	29	27	26	28
	$w_i = 0.35$	$w_i = 0.28$	$w_i = 0.17$	$w_i = 0.12$	$w_i = 0.07$
Intercept	1.62 (0.11)	1.62 (0.11)	1.62 (0.11)	1.63 (0.11)	1.60 (0.11)
Number of crops	1.02 (0.25)	1.03 (0.25)	1.02 (0.25)	0.94 (0.25)	0.89 (0.23)
Average cropping months	−0.37 (0.21)	−0.40 (0.21)	−0.40 (0.21)	−0.53 (0.22)	NA
Agriculture land area	NA	NA	NA	−0.06 (0.20)	NA
Legal agriculture title	NA	NA	NA	0.51 (0.27)	NA
Use of any mitigation measure	1.84 (0.22)	1.88 (0.22)	1.86 (0.22)	1.82 (0.22)	1.77 (0.21)
Distance to PA	−0.55 (0.21)	−0.55 (0.21)	−0.54 (0.21)	−0.50 (0.21)	−0.56 (0.21)
Distance to water	NA	NA	0.22 (0.21)	0.21 (0.21)	0.22 (0.21)
Forest cover	NA	−0.25 (0.20)	−0.27 (0.20)	−0.30 (0.20)	−0.23 (0.20)
Model AICc	636.67	637.14	638.08	638.8	639.75
Δ AICc	0	0.47	1.41	2.13	3.08

*Note: Standard errors in brackets and top-ranked models are shown, wi is the AIC model weight ΔAICc is the difference in values between lowest AIC model and each model.

doi:10.1371/journal.pone.0050433.t003

risk occurs in some places and average estimated probability was 0.60 (range 0.30–0.85). Probability of livestock loss averaged for households inside the administrative buffer was 0.74 (S.E = 0.004, range 0.57–0.85) and loss probability averaged for households outside buffer was 0.55 (S.E = 0.005, range 0.30–0.83). Spatial modeling by ordinary kriging identified hot spots of risk around the villages of Ramepur, Moharai and Chilpi inside the buffer and Baila outside the buffer (Fig. 4).

3 Mitigation Measures

Households reported several mitigation measures used by them to protect property and lives. To protect crops the most common mitigation measures were night watching (67%) and scare devices (49%) and to protect livestock measures deployed by households were closer watch on animals (48%) and physical structures (47%, Table 1, Fig. 2). Our models suggest that potentially physical structures and guard animals appeared to be associated with lower crop loss, and no mitigation appeared to be associated lower livestock loss.

4 Compensation Distribution

Although 73% of those surveyed reported that they experienced crop loss, only 26% of these households reported losses to authorities and 22% of those reporting received compensation

Table 4. Top models (cumulative weight >0.95) and beta coefficients for predicting household livestock loss around Kanha National Park.

Models	27	28	5	3	29	4	6	2	1
	$w_i = 0.25$	$w_i = 0.19$	$w_i = 0.16$	$w_i = 0.12$	$w_i = 0.09$	$w_i = 0.08$	$w_i = 0.06$	$w_i = 0.02$	$w_i = 0.02$
Intercept	0.24 (0.08)	0.24 (0.08)	0.24 (0.08)	0.24 (0.08)	0.24 (0.08)	0.24 (0.08)	0.24 (0.08)	0.24 (0.08)	0.24 (0.08)
Distance to PA	−0.93 (0.16)	−0.93 (0.16)	−0.97 (0.16)	−0.94 (0.16)	−0.93 (0.16)	−0.95 (0.16)	−0.97 (0.16)	−0.83 (0.17)	−0.95 (0.16)
Distance to water	0.46 (0.16)	0.47 (0.16)	0.49 (0.16)	0.46 (0.16)	0.46 (0.16)	0.47 (0.16)	0.49 (0.16)	0.40 (0.18)	0.51 (0.16)
Forest cover	−0.05 (0.16)	−0.03 (0.16)	−0.05 (0.16)	−0.06 (0.16)	−0.05 (0.16)	−0.05 (0.16)	−0.05 (0.16)	−0.03 (0.15)	−0.04 (0.15)
Elevation	NA	NA	NA	NA	NA	NA	NA	0.29 (0.19)	NA
Grazing cows inside PA	0.42 (0.19)	0.40 (0.19)	0.41 (0.19)	0.43 (0.19)	0.42 (0.19)	0.42 (0.19)	0.41 (0.19)	NA	NA
Night watching	NA	NA	NA	0.38 (0.21)	NA	0.37 (0.20)	NA	NA	NA
Physical structures	NA	NA	NA	−0.08 (0.16)	NA	−0.08 (0.16)	NA	NA	NA
Any mitigation measure	0.29 (0.17)	0.30 (0.17)	NA	NA	0.29 (0.17)	NA	NA	NA	NA
Total number of people in household	NA	0.19 (0.16)	NA	NA	NA	0.16 (0.16)	NA	NA	NA
Total livestock	NA	NA	NA	NA	0.001 (0.16)	NA	0.04 (0.16)	0.09 (0.16)	0.10 (0.16)
Model AICc	961.53	962.12	962.55	963.07	963.57	964.03	964.53	966.74	967.03
Δ AICc	0	0.59	1.03	1.54	2.04	2.5	3	5.21	5.5

*Note: Standard errors in brackets and top-ranked models are shown, wi is the AIC model weight Δ AICc is the difference in values between lowest AIC model and each model.

doi:10.1371/journal.pone.0050433.t004

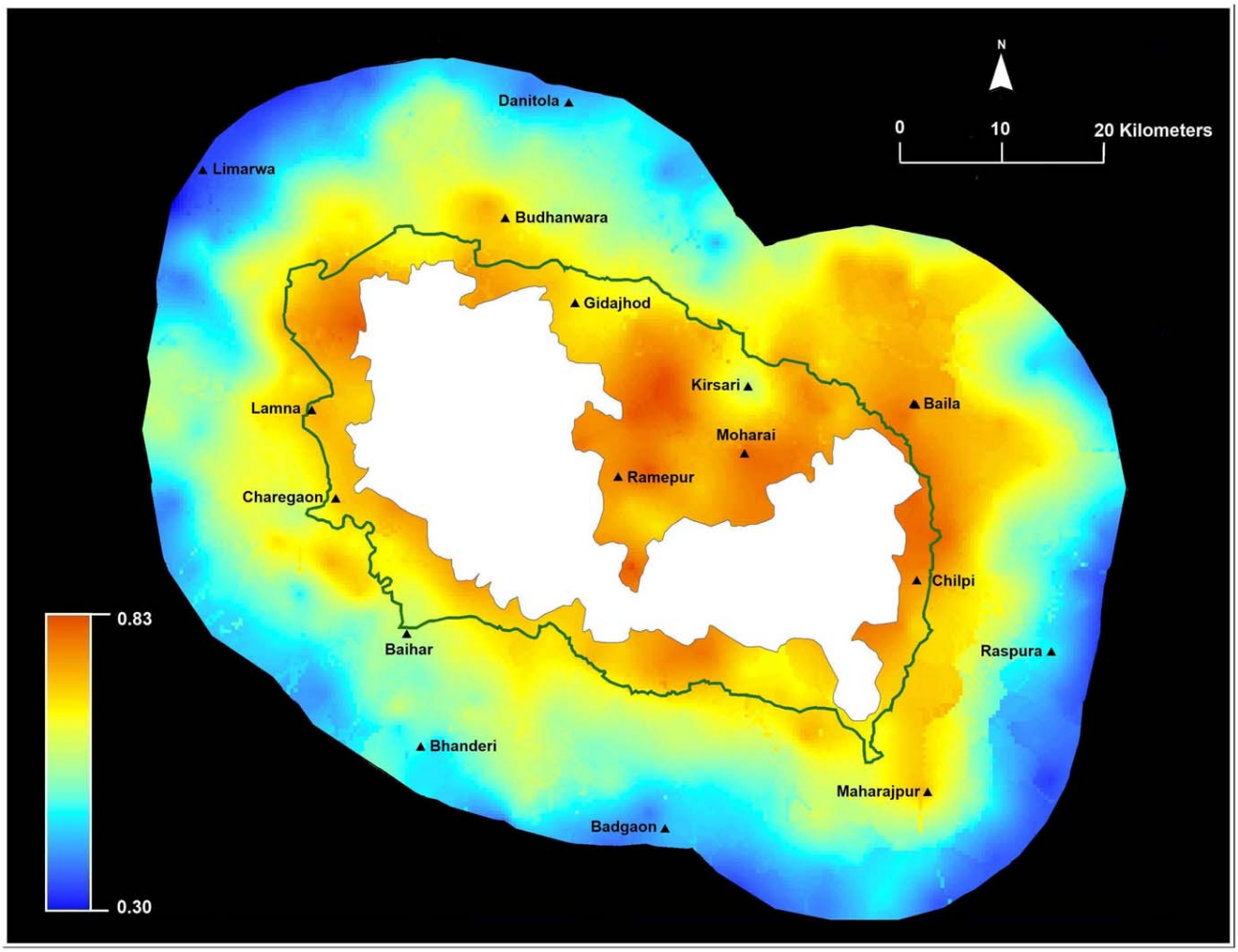


Figure 3. Predicted crop loss within 20 km around Kanha National Park (dark green is the administrative buffer). Kriging generates probabilities for crop loss in the landscape with blue areas depicting low risk and red areas depicting high risk.
doi:10.1371/journal.pone.0050433.g003

(ranging from $21 to $104, 1 US$ = 48 Indian Rupees at the time of the study). Most (63%) households reported losing up to 75% of crops to wildlife (Table 1). Authorities reportedly visited 18% of affected households, and 68% of visits were within the first week. In Kanha the average household reported annual loss of income from crop loss was US $90 (where average annual income is $517 per household).

Thirty percent of surveyed households reported to have experienced livestock loss, but only 34% of these affected households reported losses to authorities and 41% received compensation upon reporting. In Kanha official compensation paid from 2006–2011 to 524 households inside the PA was $84 and higher than the $63 compensation paid to 1644 households located in the buffer (Kanha Forest Department 2012). In our study, households reported receiving compensation averaging $22 (ranging $20–100). Although we cannot directly compare official compensation paid to individual households, the financial amounts of compensation from our surveyed households compare with official records. Surveys conducted in parks around Karnataka also find similar self-reporting levels of loss and compensation ([12], Karanth et al. unpublished).

From modeling factors associated with reported household access to compensation we identify important factors associated with compensation receipt (Table 5). Households filing claims to authorities and located in the administrative buffer were more likely to receive compensation and this is supported by positive beta coefficients for reporting crop and livestock loss (Table 5). Species involved was also associated with compensation received by households, particularly reporting tigers had positive influence (positive beta coefficient in Table 5). Conflict incidents related to other species do not appear to have yielded compensatory benefits to households (negative beta coefficients in Table 5). Reporting incidents of human injury and death occur in fewer models but did not appear to be associated with improved compensation (positive beta coefficients in Table 5). Households' location in the legal buffer appeared to be associated with higher compensation (positive beta coefficients in Table 5). Other modeled factors did not appear to influence compensation (Table S3).

We compared losses experienced and reported to authorities, and compensation received by households inside and outside the buffer. Households reporting crop loss was similar inside and outside the buffer but households reporting livestock loss was higher inside the buffer (Table 6). We find that households inside the administrative buffer were more likely to report losses (46% for crop loss and 62% for livestock loss) compared to households outside the buffer (19% for crop loss and 18% for livestock loss).

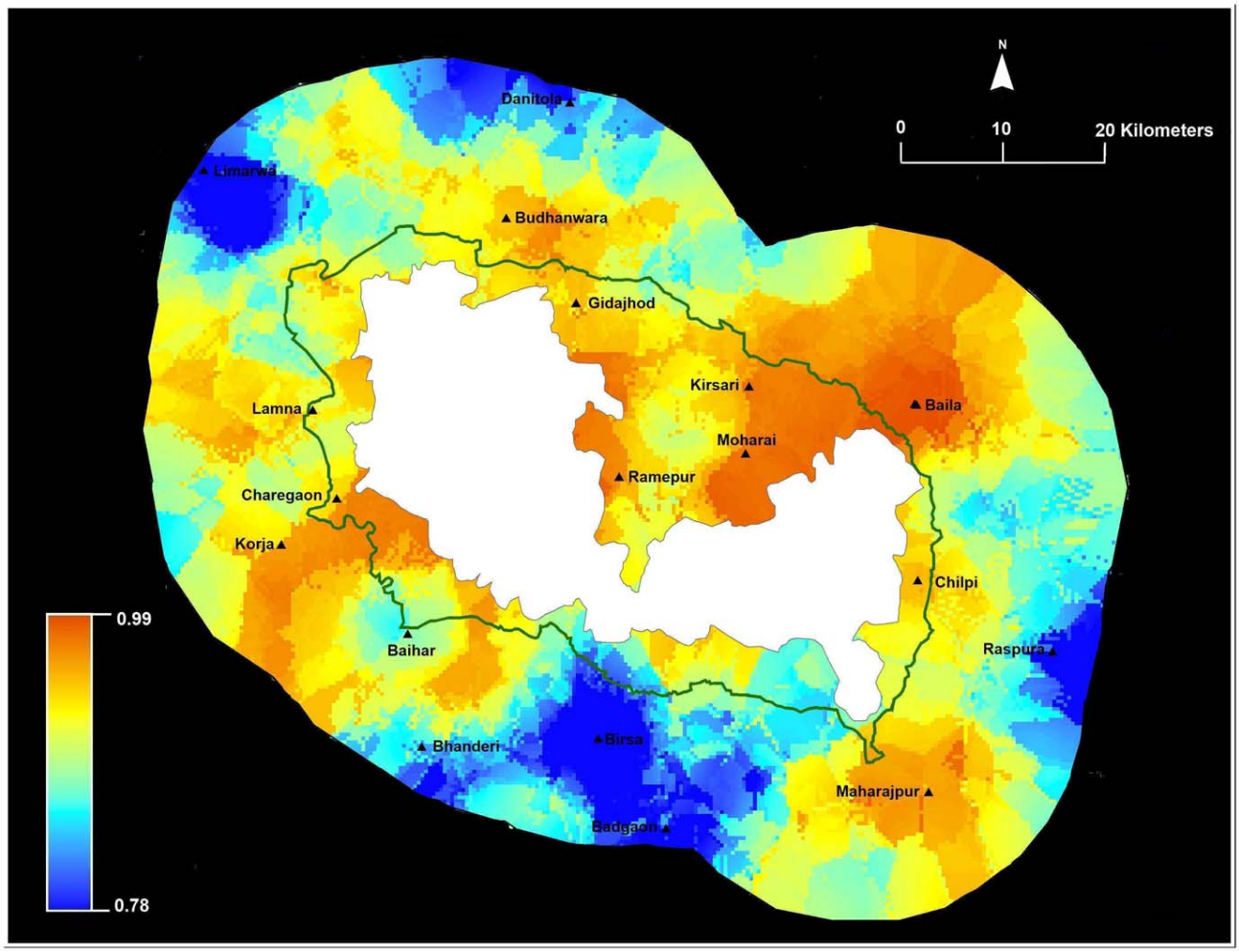


Figure 4. Predicted livestock predation loss within 15 km around Kanha National Park (dark green is the administrative buffer). Kriging generates probabilities for livestock loss in the landscape with blue areas depicting low risk and red areas depicting high risk.
doi:10.1371/journal.pone.0050433.g004

Of the households that reported loss to authorities, households inside the buffer were more likely to receive compensation (35% for crop loss and 48% for livestock loss) compared to households outside (11% for crop loss and 29% for livestock loss).

We mapped reported compensation distribution for households around Kanha. The estimated probabilities averaged 0.09 (range 0.01–0.72, Fig. 5). Households located within the administrative buffer averaged 0.21 (S. E = 0.02, range 0.03–0.72) compared to households outside 0.05 (S. E = 0.003, range 0.01–0.45). Compensation distribution for households located in the administrative buffer is higher than households located outside indicating some positive influence of management inside the buffer. Our spatial model suggests higher compensation around Moharai, Ramepur, Kirsari, Lamna, Umaerdehi and Chilpi inside the buffer (Fig. 5).

Discussion

Our results identified and quantified main factors associated with self-reported household crop raiding and livestock predation losses, as well as compensation distribution (Figures 3, 4, and 5), in 5145 km^2 area surrounding Kanha. Similar to other studies, we find spatial and temporal variation in conflict losses [4,14,18]. Overall, the probability of crop loss was high (average 0.93) for households and higher for households located within the administrative buffer. Crop loss was associated with greater cropping months and proximity to Kanha. No individual mitigation measure was associated with lowering crop loss. Overall, the probability of livestock loss was comparatively lower compared to crop loss (average 0.60) and higher for households located in the administrative buffer compared to households located outside. Livestock loss was associated with households grazing cows inside Kanha and located closer to the reserve. However, use of physical structures to protect livestock appeared to be the only mitigation measure associated with lowering of livestock loss. Interestingly, over the last ten years households reported a gradual increase in stall-feeding practices compared to free range grazing, suggesting the gradual modification of behavior. This result, however, may merely be a reflection of the inability of people to recall practices from the past. Our findings mirror other studies, where crop and livestock loss are a function of multiple factors, with individual choices and behavior influencing loss and often difficult to modify [29].

Considerable conservation attention and monies are currently being invested in mitigation factors that have rarely been evaluated for effectiveness [16,30–31]. We find only use of physical structures was potentially associated with lowering

Table 5. Top models (cumulative weight >0.95) and beta coefficients for predicting household compensation distribution around Kanha National Park.

Models	26	27	28	30	29	11	31	6
	$w_i = 0.37$	$w_i = 0.31$	$w_i = 0.10$	$w_i = 0.08$	$w_i = 0.06$	$w_i = 0.04$	$w_i = 0.03$	$w_i = 0.01$
Intercept	−2.93 (0.19)	−2.99 (0.20)	−2.95 (0.19)	−3.00 (0.20)	−2.93 (0.19)	−2.87 (0.18)	−2.97 (0.20)	−2.93 (0.19)
Crop-raiding reported to authorities	0.52 (0.38)	0.58 (0.38)	0.58 (0.38)	0.50 (0.38)	0.48 (0.39)	0.66 (0.37)	0.50 (0.39)	0.74 (0.37)
Livestock-predation reported to authorities	1.44 (0.44)	1.49 (0.45)	1.53 (0.45)	1.47 (0.46)	1.42 (0.44)	1.63 (0.43)	1.51 (0.46)	1.67 (0.45)
Wild pig	−0.83 (0.41)	−0.79 (0.41)	−0.74 (0.41)	−0.76 (0.42)	−0.83 (0.41)	−0.75 (0.40)	−0.71 (0.42)	−0.69 (0.40)
Chital	−0.05 (0.34)	−0.14 (0.35)	−0.08 (0.35)	−0.11 (0.35)	−0.03 (0.34)	−0.07 (0.34)	−0.05 (0.35)	−0.14 (0.35)
Tiger	1.42 (0.40)	1.44 (0.41)	1.37 (0.41)	1.4 (0.42)	1.40 (0.40)	1.49 (0.41)	1.33 (0.42)	1.50 (0.42)
Leopard	−0.08 (0.40)	0.05 (0.41)	0.09 (0.41)	0.05 (0.42)	−0.07 (0.4)	−0.13 (0.39)	0.10 (0.42)	−0.002 (0.41)
Langur	NA	−0.61 (0.43)	−0.36 (0.40)	−0.74 (0.45)	NA	NA	−0.51 (0.43)	−0.65 (0.43)
Wolf	NA	−0.42 (0.48)	−0.46 (0.48)	−0.51 (0.50)	NA	NA	−0.55 (0.50)	−0.47 (0.48)
Jackal	NA	−0.28 (0.45)	−0.19 (0.45)	−0.31 (0.46)	NA	NA	−0.23 (0.46)	−0.12 (0.44)
Bonnet macaque	NA	1.00 (0.47)	NA	0.98 (0.48)	NA	NA	NA	0.78 (0.46)
Human injury reported to authorities	NA	NA	NA	0.66 (0.56)	0.41 (0.53)	NA	0.68 (0.57)	NA
Human death reported to authorities	NA	NA	NA	0.24 (0.89)	−0.19 (0.85)	NA	0.37 (0.90)	NA
Household located inside buffer	0.83 (0.32)	0.95 (0.33)	0.84 (0.32)	0.96 (0.33)	0.84 (0.32)	NA	0.86 (0.32)	NA
Model AIC	346.33	346.69	348.91	349.34	349.85	350.88	351.23	352.79
Δ AIC	0	0.36	2.58	3.01	3.52	4.54	4.89	6.46

†Note: Standard errors in brackets and top-ranked models are shown, wi is the AIC model weight ΔAICc is the difference in values between lowest AIC model and each model.

doi:10.1371/journal.pone.0050433.t005

livestock loss. Our results indicate that blind investment in mitigation activities should be replaced by targeted and focused strategies that work at the individual household level.

Although 73% of households reported crop loss and 33% reported livestock loss, only 26% reported crop loss and 34% reported livestock loss to authorities. Among reporting households, 22% received compensation for crop loss and 41% received compensation for livestock loss.

Likelihood of compensation distribution received by households was higher for households that report incidents related to tigers (although most damage was reported by pig, jackal and wolf incidents, Table 1) and households located in the buffer. There is

Table 6. Comparison loss and compensation reported by households surveyed inside and outside administrative buffer around Kanha National Park.

Features	Inside Buffer	Outside Buffer	Total
Number of Households	190	543	733
HH experiencing crop loss	145 (76%)	391 (72%)	536 (73%)
HH reporting crop loss to authorities	66 (46%)	73 (19%)	139 (26%)
HH reporting crop loss to authorities receiving compensation	23 (35%)	8 (11%)	31 (22%)
HH experiencing livestock loss	87 (46%)	153 (28%)	240 (33%)
HH reporting livestock loss to authorities	54 (62%)	28 (18%)	82 (34%)
HH reporting loss to authorities receiving compensation	26 (48%)	8 (29%)	34 (41%)
Survey estimated probability of crop loss	0.95 (SE = 0.004)	0.92 (SE = 0.003)	0.93 (SE = 0.003)
	Range: 0.77–0.99	Range: 0.66–0.99	Range 0.66–0.99
Survey estimated probability of livestock loss	0.74 (SE = 0.004)	0.55 (SE = 0.005)	0.60 (SE = 0.005)
	Range: 0.57–0.85	Range: 0.30–0.83	Range 0.30–0.85
Survey estimated probability of compensation distribution	0.21 (SE = 0.02)	0.05 (SE = 0.003)	0.09 (SE = 0.005)
	Range: 0.03–0.72	Range: 0.01–0.45	Range 0.01–0.72

doi:10.1371/journal.pone.0050433.t006

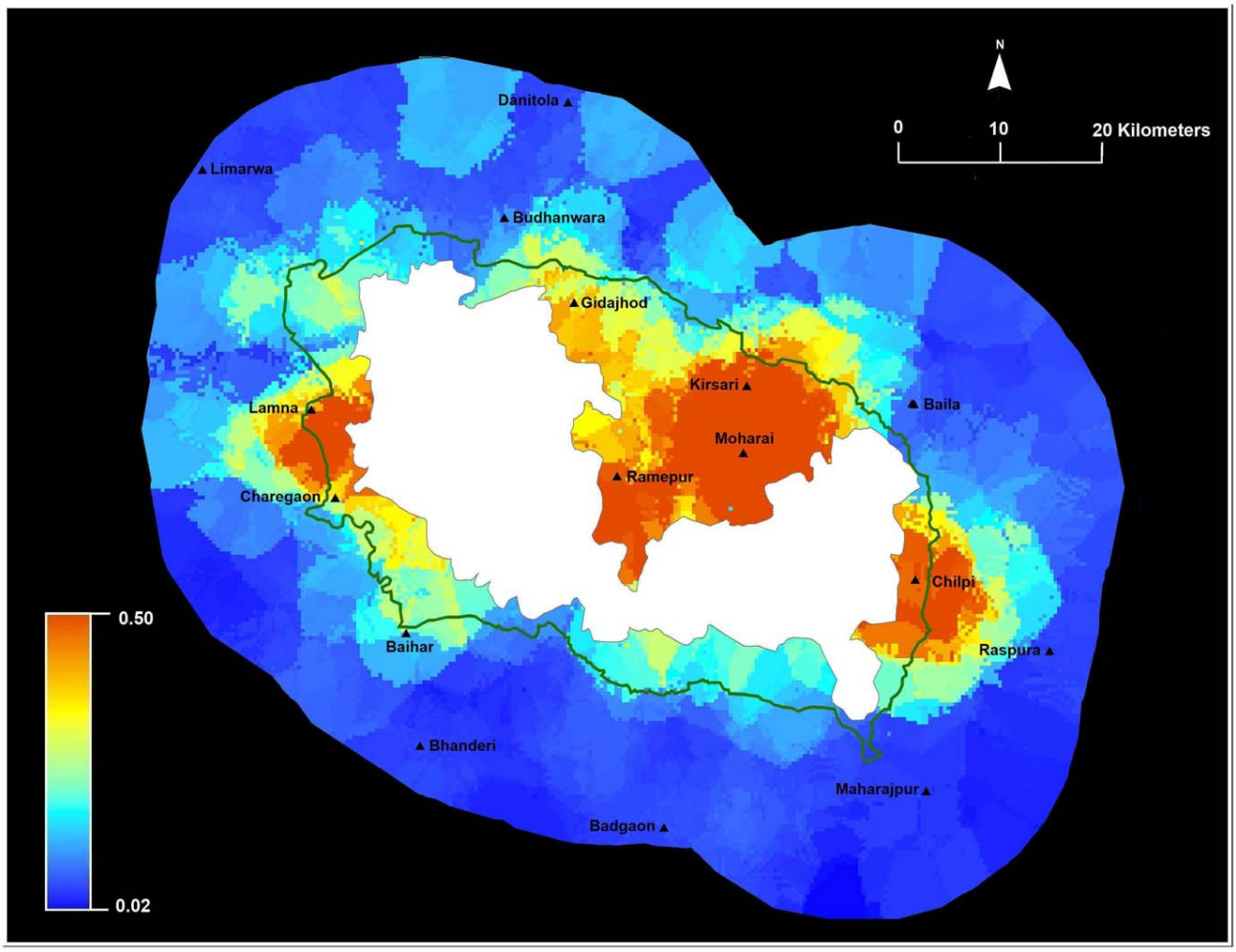


Figure 5. Predicted compensation distribution for households within 20 km around Kanha National Park. Household probabilities of receiving compensation in the landscape are generated by Kriging with blue areas depict low compensation and red areas depicting high compensation.
doi:10.1371/journal.pone.0050433.g005

considerable debate about people's ability to accurately report losses to authorities, and other studies have found mismatches to official records [2,14,32]. These reported losses determined by surveying people have to be complemented with field-based monitoring of households over time. We do find among surveyed households in the administrative buffer were twice as likely to report crop loss and thrice as likely to report livestock loss compared to outside (Table 6). Households in the administrative buffer were more likely to receive compensation from authorities, suggesting that the legal designation and management of an administrative buffer improves reporting of losses and receipt of compensation by households (Table 6).

Overall, many (87%) households around Kanha report experiencing some kind of conflict incident with wildlife and this is comparatively higher than other places in the world [2,33]. However, there appears to be some degree of acceptance among people that there will be some losses to wildlife (Karanth pers. obs 2011, Karanth et al. unpublished), and perhaps households have evolved coping strategies [11–12,17,34].

Unlike most Indian PAs, Kanha has a legal administratively designated management buffer [9]. We estimated higher risk of crop and livestock loss inside the buffer, and find higher compensation in the buffer. This corroborates with other studies that suggest delineation of buffers or other management strategies should be based on ecological and economic realities [9]. Additionally, in this landscape we found household distance to the PA was more relevant for livestock and crop loss compared to distance to water. Our results potentially suggest that some crop losses to wildlife from the PA but we know that wildlife naturally persists outside the PA and have overlapping resource use with people and domestic animals (forage and water, Karanth et al. unpublished). So this raises important questions about who is responsible for compensating local people for wildlife damage outside the jurisdiction of PAs and park authorities [6,35]. Our spatial risk maps (Figs. 3 and 4) clearly identified locations that are hotspots for crop and livestock loss and can guide managers in improving existing conflict prevention and mitigation practices.

Management Implications

Most conflict studies are characterized by poor spatial sampling and modeling [36–37]. To improve efficacy and efficiency of conservation actions, managers require surveying and modeling approaches that are spatially explicit and rigorous [5,7]. Our approach identified factors associated with crop and livestock loss

around Kanha National Park in India. We were able to link ground based survey data to landscape level data and spatially map conflict crop and livestock loss, as well as compensation distribution to identify conflict hotspots and focus ground based conservation efforts. For example, villages and households located in high-risk areas can be better educated about how to prevent and mitigate conflict and can report losses to authorities in an organized manner.

A monitoring system that systematically records and disperses information on conflict is much needed. Such an approach can guide the development of a risk database, and develop live warning and monitoring systems (such as those cell phone alert systems currently in use for Elephants in Valparai in southern India). Experiences from Uganda, Kenya and Sumatra suggest that establishing and long-term maintenance of monitoring systems in local communities remains a considerable challenge and is often difficult to sustain [15,30,38]. However, greater participation and involvement of multiple stakeholders (local households, PA authorities) can considerably decrease hostility towards wildlife species [29,35,39].

Kanha is atypical of most Indian PAs because the administrative buffer gives it a lesser hard edge than other PAs and livelihoods of people living in the buffer fall within the purview of park management [9]. As consideration is currently under way for designating buffers around other protected areas in India, our results from Kanha suggest that compensation is more likely to be distributed to those who suffer losses if the administrative buffer is designated to include more susceptible locations for crop and livestock loss. This conclusion may apply to other parks located in human-dominated landscapes, though cultural and ecological factors will vary.

Our study has some limitations. The analysis relied primarily on individuals' recall and self-reporting obtained through surveys and did not observationally monitor actual incidents of conflict. This raises possibilities for over-reporting or exaggeration of losses, although there are no indications of a systematic bias across so many households. The number of surveys is large relative to most published studies and field-based human-wildlife conflict surveys that reduced the potential biases (although we recognize the limitations of self-reported data). To complement the surveys and provide objective data on human wildlife conflicts, active location based monitoring of households to measure conflict as it happens along with assessment of mitigation measures and compensation distribution process is much needed. Extension of surveys across multiple years is also needed to assess additional factors that might influence conflict risk that may not be identified in a single year assessment. Overall, our approach for identifying associated factors and spatially mapping risk is easily extendable to other landscapes where people and wildlife co-occur, so that allocation of physical efforts and funds are more effectively targeted at conflict prevention and mitigation.

Supporting Information

Table S1 Correlation matrices for variable selection for crop loss and livestock loss.
(DOC)

Table S2 Models included in the model sets for crop and livestock loss.
(DOC)

Table S3 Models included in the model set for compensation distribution.
(DOC)

Acknowledgments

We thank Centre for Wildlife Studies and Wildlife Conservation Society-India program for support. We are grateful to the Madhya Pradesh forest department especially H. S. Pabla, J. S. Chauhan, H. S. Mohanta and R. Shukla for their support. We thank S. Dasgupta, R. Jhaveri, S. Kotian and N. Patil. We thank volunteers A. Sur, A. Dhanpal, A. Joshi, A. Kerketta, N. Kumar, A. Bade, A. Kodilkar, A. Mendis, C. V. Casiker, D. Prasad, D. Dixit, I. Majhgaonkar, K. P. Kothari, Mythri. S, N. Surve, N. Shitole, P. Ghaskadbi, P. Baro, P. Thatte, P.R. Khambe, P. D. Kamble, R. Chakravarty, R. Jha, S. Sawat, Shashank K, S. Yadav, V. Kumar, Vatsaraj, V. Shinde, and Y. S. Patil. Columbia University IRB approved all survey protocols.

Author Contributions

Conceived and designed the experiments: KKK AMG RD. Performed the experiments: KKK NB. Analyzed the data: KKK AMG RD. Contributed reagents/materials/analysis tools: KKK AMG RD NB. Wrote the paper: KKK AMG RD.

References

1. Treves A, Karanth KU (2003) Human-Carnivore conflict and perspectives on carnivore management worldwide. Conservation Biology 17: 1491–1499.
2. Dickman AJ (2010) Complexities of conflict: the importance of considering social factors for effectively resolving human-wildlife conflicts. Animal Conservation 13: 458–466.
3. Treves A, Naughton-Treves L (1999) Risk and opportunity for humans existing with large carnivores. Journal of Hum Evolution 36: 275–282.
4. Linkie M, Dinata Y, Nofrianto A, Leader-Williams N (2007) Patterns and perceptions of wildlife crop raiding in and around Kerinci Seblat National Park, Sumatra. Animal Conservation 10: 127–135.
5. MacDonald-Madden E, Baxter PWJ, Possingham HP (2008) Making robust decisions for conservation with restricted money and knowledge. Journal of Applied Ecology 45: 1630–1638. doi:10.1111/j.1365-2664.2008.01553.x.
6. Bulte EH, Rondeau D (2005) Research and management view point: Why compensating wildlife damages may be bad for conservation. Journal of Wildlife Management 69: 14–19.
7. Treves A, Martin KA, Wydeven AP, Weidenhoeft JA (2011) Forecasting environmental hazards and the application of risk maps to predator attacks on livestock. BioScience 61: 451–458.
8. Madden F (2004) Creating coexistence between humans and wildlife: Global perspectives on local efforts to address human–wildlife conflict. Human Dimensions of Wildlife 9: 247–257.
9. DeFries R, Karanth KK, Pareeth SJ (2010) Interactions between protected areas and their surroundings in human-dominated tropical landscapes. Biological Conservation 143: 2870–2880. doi:10.1016/j.biocon.2010.02.010.
10. Karanth KU, Madhusudan MD (2002) Mitigating human-wildlife conflicts in southern Asia. In: Terborgh JW, van Schaik C, Davenport L, Rao M, editors. Making Parks Work: Identifying key factors to implementing parks in the tropics. Covelo, CA: Island Press. 250–264.
11. Sekhar N (1998) Crop and livestock depredation caused by wild animals in protected areas: the case of Sariska Tiger Reserve, Rajasthan, India. Environmental Conservation 25: 160–171.
12. Karanth KK, Nepal S (2012) Local perceptions of benefits and losses of living around protected areas in India and Nepal. Environmental Management 49: 372–386. doi: 10.1007/s00267-011-9778-1.
13. Mateo-Tomas P, Olea PP, Sanchez-Barbudo IS, Mateo R (2012) Alleviating human-wildlife conflicts: identifying the causes and mapping the risk of illegal poisoning of wild fauna. Journal of Applied Ecology 49: 376–385. doi: 10.1111/j.1365-2664.2012.02119.x.
14. Naughton-Treves L (1997) Farming the forest edge: vulnerable places and people around Kebale National Park, Uganda. Geographical Review 871: 27–46.
15. Campbell-Smith G, Sembiring R, Linkie M (2012) Evaluating the effectiveness of human-orangutan conflict mitigation strategies. Journal of Applied Ecology. doi: 10.1111/j.1365-2664.2012.02109.x.
16. Sitati NW, Walpole MJ, Leader-Williams N (2005) Factors affecting susceptibility of farms to crop raiding by African elephants: using a predictive model to mitigate conflict. Journal of Applied Ecology 42: 1175–1182.

17. Treves A, Naughton-Treves L (2005) Evaluating lethal control in the management of human-wildlife conflict. People and Wildlife, Conflict or Coexistence? Pp 86–106. Cambridge University Press, Cambridge, UK.
18. Kolowski J, Holekamp K (2006) Spatial, temporal and physical characteristics of livestock depredations by large carnivores. Biological Conservation 128: 529–541.
19. Baruch-Mordo S, Breck SW, Wilson KR, Broderick J (2009) A tool box half full: How social science can help solve human-wildlife conflict. Human Dimensions of Wildlife 14: 219–223.
20. Gopalaswamy AM, Karanth KU, Kumar NS, Macdonald DW (2012) Estimating tropical forest ungulate densities from sign surveys using abundance models of occupancy. Animal Conservation doi:10.1111/j.1469-1795.2012.00565.x.
21. Buda A, Jarynowski A (2010) Life-time of correlations and its applications vol.1, Wydawnictwo Niezalezne.
22. Cohen J (1988) Statistical power analysis for the behavioral sciences.
23. Gelman A (2008) Scaling regression inputs by dividing by two standard deviations. Statistics in Medicine 27: 2865–2873.
24. Burnham KP, Anderson D (2002) Model Selection and Multimodal Inference. A practical information theoretic approach. Springer-Verlag, New York.
25. Mitas L, Mitasova H (1999) Spatial interpolation. In: Longley P, Good-child M, Maguire D, Rhind D, editors. Geographical Information Systems: Principles, Techniques, Management and Applications, vol. 1. London: Wiley. 481–492.
26. Akkala A, Devabhaktuni V, Kumar A (2010) Interpolation techniques and associated software for environmental data. Environ. Prog. Sustainable Energy 29: 134–141.
27. Boeringa M (2003) Kriging interpolator 3.2 for ArcView spatial analyst. Nieuwland Automatisering, Wageningen, The Netherlands. Available: http://www.nieuwland.nl. Accessed 2012 Jun 5.
28. Cressie NAC (1991) Statistics for Spatial Data. John Wiley & Sons. 900.
29. Gore ML, Knuth BA, Scherer CW, Curtis PD (2008) Evaluating a conservation investment design to reduce human-wildlife conflict. Conservation Letters 1: 136–145.
30. Graham MD, Ochieng T (2008) Uptake and performance of farm-based measures for reducing crop raiding by elephants Loxodonta africana among smallholder farms in Laikipia District, Kenya. Oryx 421: 76–82.
31. Gehring TM, VerCauteren KC, Provost ML, Cellar AC (2010) Utility of livestock-protection dogs for deterring wildlife from cattle farms. Wildlife Research 37: 715–721.
32. Boitani L, Ciucci P, Raganella-Pelliccioni E (2010) Ex-post compensation payments for wolf predation on livestock in Italy: a tool for conservation? Wildlife Research 37: 722–730. doi:10.1071/WR10029.
33. Woodroffe R, Thirgood S, Rabinowitz A (2005) People and Wildlife, Conflict or Coexistence? Cambridge University Press, Cambridge, UK.
34. Røskaft E, Bjerkec T, Kaltenborn B, Linnell JDC, Anderson R (2003) Patterns of self-reported fear towards large carnivores among the Norwegian public. Evolution & Human Behavior 24: 184–198.
35. Naughton-Treves L, Grossberg R, Treves A (2003) Paying for tolerance: The impact of livestock predation and compensation payments on rural citizens attitudes towards wolves. Conservation Biology 17: 1500–1511.
36. Peterson MN, Birckhead JL, Leong K, Peterson MJ, Peterson TR (2010) Rearticulating the myth of human–wildlife conflict. Conservation Letters 3: 74–82. doi:10.1111/j.1755-263X.2010.00099.x.
37. White PC, Ward AI (2011) Interdisciplinary approaches for the management of existing and emerging human-wildlife conflicts. Wildlife Research 37: 623–629.
38. Webber AD, Hill CM, Reynolds V (2007) Assessing the failure of a community-based human–wildlife conflict mitigation project in Budongo Forest Reserve, Uganda. Oryx 41: 177–184.
39. Treves A, Wallace RB, White S (2009) Participatory planning of interventions to mitigate human-wildlife conflicts. Conservation Biology 23: 1577–1587.

Appendix K

Site-occupancy modelling as a novel framework for assessing test sensitivity and estimating wildlife disease prevalence from imperfect diagnostic tests

Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2012, **3**, 339–348 doi: 10.1111/j.2041-210X.2011.00156.x

Site-occupancy modelling as a novel framework for assessing test sensitivity and estimating wildlife disease prevalence from imperfect diagnostic tests

Shelly Lachish[1]*, Arjun M. Gopalaswamy[2,3,4], Sarah C. L. Knowles[1,5] and Ben C. Sheldon[1]

[1]*Edward Grey Institute, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK;* [2]*Wildlife Conservation Research Unit (WildCRU), Department of Zoology, University of Oxford, Recanati-Kaplan Centre, Tubney House, Tubney OX13 5QL, UK;* [3]*Wildlife Conservation Society: India Program, 1669 31st Cross 16th Main, Banashankari 2nd Stage, Bengaluru 560070, India;* [4]*Centre for Wildlife Studies, 26-2, Aga Abbas Ali Road, Ulsoor, Bengaluru 560 042, India; and* [5]*Institute of Evolutionary Biology, University of Edinburgh Ashworth Labs, King's 12 Buildings, West Mains Road, Edinburgh EH9 3JT, UK*

Summary

1. Reliable assessments of infection status and population prevalence are critical for epidemiological modelling and disease management, but can be greatly biased when disease state is determined from imperfect diagnostic tests. Available statistical methods to adjust test-based prevalence estimates by correcting for test accuracy demand that many stringent requirements and assumptions be met (knowledge about underlying population prevalence or multiple diagnostic methods), limiting their utility for wildlife disease surveys.

2. In this paper, we present site-occupancy modelling as a flexible approach to derive estimates of population prevalence and test sensitivity under imperfect pathogen detection without a need for restrictive requirements or assumptions. We extend the utility of the standard site-occupancy framework for pathogen detection data by novel application of abundance-induced heterogeneity (AIH) models (Royle & Nichols 2003) that allow test sensitivity to vary with host pathogen load or infection intensity.

3. We demonstrate the utility of this approach for wildlife disease studies by applying site-occupancy models to a data set consisting of replicate quantitative (q)PCR diagnoses of malaria parasites (*Plasmodium* spp.) in blood samples from wild blue tits (*Cyanistes caeruleus*).

4. Model selection revealed that *Plasmodium* detection rates by qPCR were strongly dependent on host parasite load. Estimates of parasite detection rates revealed the qPCR assay to be highly sensitive, with accordingly, a very low probability of false negative diagnosis for the majority of infected hosts in our population and little bias in naive estimates of population prevalence, although this will be a system-specific result.

5. Our results demonstrate the utility of a site-occupancy approach for deriving estimates of population prevalence under imperfect pathogen detection and reveal that accounting for host variation in pathogen load allows a more accurate assessment of diagnostic test sensitivity.

6. By identifying factors that influence pathogen detection rates, and revealing optimal protocols for obtaining unbiased prevalence estimates, while minimising the probability of false negative diagnoses, we also show that this approach can enhance both diagnostic accuracy and cost-efficiency in wildlife disease surveys.

Key-words: *Cyanistes caeruleus*, imperfect diagnostic tests, pathogen detection, pathogen load, *Plasmodium*, population prevalence, qPCR, site-occupancy modelling, test sensitivitywildlife disease

Introduction

Understanding the epidemiology and dynamics of wildlife diseases requires accurate estimation of disease parameters, particularly pathogen prevalence and infection status (Jennelle

*Correspondence author: E-mail: shelly.lachish@zoo.ox.ac.uk
Correspondence site: http://www.respond2articles.com/MEE/

© 2011 The Authors. Methods in Ecology and Evolution © 2011 British Ecological Society

et al. 2007; McClintock *et al.* 2010). Reliable assessments of individual infection status and population prevalence are also critical for effective disease monitoring and risk analysis, for understanding the ecological and evolutionary implications of pathogens in host populations and for conservation endeavours involving selective culling operations, captive breeding programmes or population insurance strategies (Woodroffe, Frost, & Clifton-Hadley 1999; Jones *et al.* 2007; Poulin 2007; Williams & Moffitt 2010). In recent years, there has been a growing awareness that differences in the detection probability (i.e. capture or sighting probability) of infected and uninfected individuals in the wild can bias estimates of population prevalence and lead to spurious inferences regarding the epidemiological nature of disease (Senar & Conroy 2004; Jennelle *et al.* 2007; Conn & Cooch 2009). This recognition has led to the increasing adoption of sophisticated mark–recapture methods by wildlife disease ecologists to obtain parameter estimates while accounting for variability in detection rates with infection status (Senar & Conroy 2004; Conn & Cooch 2009).

In contrast, however, the potential for imperfect detection of disease status *per se* (i.e. the failure to detect pathogens when present, or the failure to diagnose correctly) has received considerably less attention by both wildlife disease ecologists and wildlife epidemiologists (but see Smith *et al.* 2001; Thompson 2007; Gomez-Diaz *et al.* 2010; McClintock *et al.* 2010). Failure to account for such uncertainty in disease-state assignment can significantly bias parameter estimates and lead to erroneous inferences about diseases in natural systems (Smith *et al.* 2001; McClintock *et al.* 2010). For example, failure to diagnose infection in a subsample of individuals will result not only in prevalence estimates that are biased low and inaccurate estimates of virulence but also in significant underestimation of pathogen transmission rates and hence flawed predictions regarding infection dynamics and epidemic spread (Senar & Conroy 2004).

Problems associated with estimating disease prevalence and establishing true infection status in the presence of imperfect diagnostic tests are well-known in the veterinary and medical fields and have led to a variety of statistical methods that attempt to adjust apparent (test-based) prevalence estimates by correcting for the accuracy of the tests used (reviewed in Enoe, Georgiadis, & Johnson 2000; Greiner & Gardner 2000; Toft, Jorgensen, & Hojsgaard 2005). Most of these methods involve assessing the accuracy of a diagnostic test in comparison with a reference test that reflects true infection status (Greiner & Gardner 2000; Drewe *et al.* 2009). In the absence of such a reference test, latent class methods (in which the unknown, or latent, variable is the true infection status) can be used to estimate corrected prevalence with imperfect tests (Enoe, Georgiadis, & Johnson 2000). Implementation of these latent class methods involves either the application of two or more conditionally independent tests (where the outcomes of the diagnostic tests are independent of each other) or a comparison between two or more populations with different prevalence of infection (Enoe, Georgiadis, & Johnson 2000). These approaches thus assume not only that several diagnostic methods are available (to satisfy assumptions of conditional independence), but also that there is some knowledge of underlying population prevalence, and more importantly, that the accuracy of the tests used remains constant in different populations (Enoe, Georgiadis, & Johnson 2000; Toft, Jorgensen, & Hojsgaard 2005).

The requirements and assumptions above will rarely be met in studies of wildlife disease. Wildlife disease diagnoses are generally obtained from only one diagnostic method or even simply from visual inspection when no diagnostic tool is available, or when costs or capture of individuals are prohibitive (Faustino *et al.* 2004; Lachish, Jones, & McCallum 2007). Knowledge of population prevalence prior to sampling is also almost always unavailable for wildlife systems, as this is often the goal of such studies. Moreover, it is plausible that pathogen detection probabilities will vary between populations as a result of variation in the intrinsic properties of wild hosts (genetic and physiological traits, immune function) and their pathogens (virulence, pathogenicity) (Toft, Jorgensen, & Hojsgaard 2005; Acevedo-Whitehouse & Cunningham 2006; Lively 2006). Therefore, existing statistical methods to estimate prevalence while accounting for imperfect pathogen detection are inadequate for most wildlife disease systems.

A further consideration is that the accuracy of many diagnostic tests is known to vary with the intensity of infection (Greiner & Gardner 2000). For example, the sensitivity of the serological assay used to detect *Mycobacterium bovis* (TB) infection in cattle increases with the time elapsed since experimental inoculation (Drewe *et al.* 2009). Similarly, the probability of detecting pathogens by quantitative PCR technologies increases with parasite density within the host (DNA copy number, Kriger, Hero, & Ashton 2006; Knowles *et al.* 2011), while the visual diagnosis of facial tumour disease in captured Tasmanian devils (*Sarcophilus harrisii*) improves when symptoms are more severe (S. Lachish, unpublished data). The importance of this source of variation for confidence in the assignment of infection status or precision in estimates of prevalence, however, is commonly overlooked or relegated to cursory treatments (such as grouping individuals with differing clinical symptoms, Williams & Moffitt 2010). Hence, a robust method for estimating disease prevalence accounting for imperfect pathogen detection should clearly allow for detection probabilities to vary as a function of pathogen load (or disease severity) as well as for greater flexibility in the detection/diagnosis methods utilised.

In this paper, we present site-occupancy modelling (MacKenzie *et al.* 2002; Royle & Nichols 2003) as a novel framework to derive estimates of population prevalence and test sensitivity under imperfect pathogen detection, and when test sensitivity varies with host pathogen load, without the need for knowledge about underlying population prevalence or multiple diagnostic methods. To demonstrate the utility of site-occupancy modelling to obtain reliable estimates of disease prevalence and pathogen detection probabilities in wildlife disease studies, we applied site-occupancy models to a data set consisting of replicate quantitative (q)PCR diagnoses of avian malaria parasites (*Plasmodium* spp.) in blood samples obtained from wild blue tits (*Cyanistes caeruleus*). We aimed to

(i) evaluate factors contributing to variation in *Plasmodium* detection probability, (ii) quantify the influence of host parasite load on *Plasmodium* detection probabilities, (iii) determine the probability of false negative diagnoses and (iv) obtain robust estimates of *Plasmodium* prevalence in this population accounting for variation in detection probabilities. We show that occupancy modelling provides a flexible alternative framework for the estimation of wildlife disease prevalence from imperfect diagnostic tests and discuss future model advancements (for example, to allow for false positive diagnoses) that can be incorporated to overcome current limitations of this modelling framework. Finally, we show that occupancy modelling can also serve as a tool to refine and optimise diagnostic protocols for disease detection and thus enhance cost-efficiency in wildlife disease surveys.

Methods

FIELD METHODS

Individuals were sampled as part of an ongoing study of avian malaria conducted in Wytham Woods (51°46′N, 1°20′W), a 385-ha mixed deciduous woodland near Oxford, UK, where 250–450 pairs of blue tits breed in nest boxes every year (Wood *et al.* 2007). Adult birds were captured at nest boxes when their broods were between 6 and 15 days post-hatch. Hence, all hosts were sampled at a standardised point in their annual cycle, controlling for seasonal variation in parasite prevalence at this location (Cosgrove *et al.* 2008). The data set analysed here relates to blood samples collected from 156 individuals in 2007.

MOLECULAR METHODS FOR PLASMODIUM DIAGNOSES

Blood samples were taken by brachial or jugular venepuncture and stored in SET buffer. DNA was extracted using a standard ammonium acetate method, involving thoroughly mixing and vortexing (homogenising) the samples, and then diluted to a standard concentration of 2 ng μL^{-1}. A qPCR assay was employed for the detection and quantification of *Plasmodium* spp. (lineages pSGS1, pGRW11, pTURDUS1, pBT7; Palinauskas *et al.* 2007) in blood samples, using the primers L9 50-AAACAATTCCTAACAAAACAGC-30 and NewR 50-ACATCCAATCCATAATAAAGCA-30, which target a 188-bp region of the mitochondrial cytochrome *b* gene. Complete details of the DNA extraction protocols, reaction mixes and standard curve generation methods for DNA quantification and thermal-cycling conditions for application of the qPCR are given in Knowles *et al.*(2011). We included multiple negative and positive controls in each PCR plate, and as results were only used when all negative controls on a PCR plate were clean (no contamination), we are confident that the occurrence of false positives is negligible. For each blood sample in this data set, six replicate qPCR were performed using the same DNA extraction. These replicate qPCR were run as two batches of triplicate assays (i.e. over two PCR plates). Parasite load was determined from the mean *Plasmodium* DNA copy number across all six qPCR. Note that we have not yet undertaken the necessary comparative molecular and microscopic methods to translate qPCR DNA copy number to numerical measures of parasite abundance.

APPLICATION OF SITE-OCCUPANCY MODELS TO PATHOGEN DETECTION DATA

In traditional applications of site-occupancy models, a number of sites are surveyed multiple times for the presence of a particular species to estimate the proportion of all sites that are occupied by the species, with the explicit recognition that the species may be missed, even if present during a survey (Fig. 1; MacKenzie *et al.* 2002). In our application of site-occupancy modelling to pathogen detection data, 'sites' can be interpreted either as individuals or as biological samples obtained from individuals that a pathogen can occupy (Fig. 1). It is assumed that the sampling of individuals from within populations adheres to a rigorous sampling scheme designed to minimise

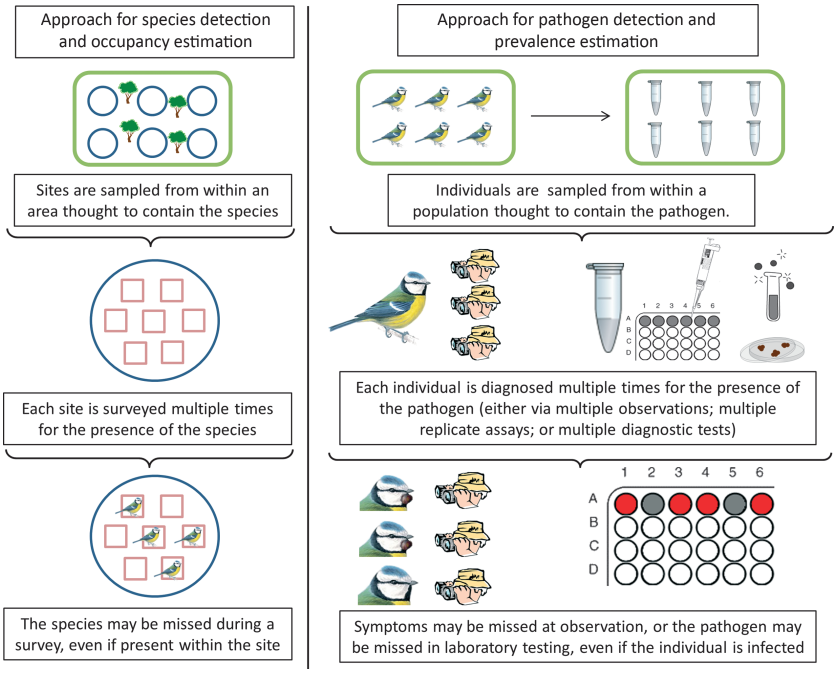


Fig. 1. Application of a site-occupancy modelling approach in the traditional ecological context for estimating occupancy under imperfect species detection and in the current context for estimating disease prevalence under imperfect diagnostic testing.

stochastic variation in the sample statistic; the proportion of infected individuals in the sample (see for example sampling designs detailed in Thompson 2002).

'Surveys' then constitute the replicate diagnoses of infection status for each individual or sample obtained from diagnostic tests, assays or through visual inspections (or a mix of these; Fig. 1). This sampling regime results in a pathogen detection history composed of a vector of 1's and 0's for each individual (positive and negative diagnoses, much like a capture history in mark–recapture data). From this detection history, the probability of pathogen detection (p, conditional on pathogen presence) is estimated along with an estimate of the true proportion of individuals infected by the pathogen (ψ, prevalence), corrected for detection probability. All positive pathogen diagnoses are regarded as true (false positive diagnoses are assumed not to occur, a point to which we return later: see Discussion), but a negative diagnosis may reflect either a true or a false absence (false negative).

Models are defined using a series of probabilistic statements describing the probability of each possible pathogen detection history (H). Assuming that the detection histories for all individuals (U) are independent, and that individuals cannot change infection status between diagnoses (i.e. individuals are either always infected or always uninfected during testing), the product of all probabilistic terms creates the model likelihood for the observed set of data:

$$L(\psi, p|H_1, \ldots, H_U) = \prod_{i=1}^{U} \Pr(H_i)$$

which can be maximised to obtain estimates of the parameters, p and ψ. Missing results for some individuals are easily accommodated within this framework by disregarding the detection probability associated with that particular diagnostic test (MacKenzie *et al.* 2002). The parameters (ψ, p) can also be easily modelled as functions of measured covariates using the logistic equation:

$$\theta = \frac{e^{(XB)}}{1 + e^{(XB)}}$$

where the parameter of interest is denoted θ, the covariate X and the vector of model parameters B. For example, if multiple diagnostic methods are used, or if tests are conducted in different laboratories, or if visual inspections are conducted by different observers, then 'method', 'laboratory' or 'Observer ID' can be included as covariates on detection rates to assess the relative importance of these factors within a model selection framework (Thompson 2007; Adams *et al.* 2010). Likewise, if diagnoses are obtained for several species, or in multiple populations, these factors can be included as covariates for corrected prevalence, and their importance assessed within a model selection framework (see for example Gomez-Diaz *et al.* 2010).

However, pathogen detection rates also frequently vary between individuals as a function of pathogen load. In most circumstances, such variation cannot be accounted for by measured covariates, either because hosts are asymptomatic or because pathogen load/severity measures are obtained only for individuals that are known to be infected (i.e. they are conditional on pathogen detection). This results in significant unaccounted for heterogeneity in detection probabilities among samples, violating the assumptions of the standard site-occupancy model (MacKenzie *et al.* 2002). This problem was recognised by Royle & Nichols (2003) who proposed an extension to the MacKenzie *et al.* (2002) model, which allows detection probabilities to be modelled as a function of the underlying, but unknown, species abundance in each sample:

$$p_{ij} = 1 - (1 - r)^{N_{ij}}$$

where p_{ij} (in this case) is the probability of detecting at least one parasite in diagnostic test j, of individual i; r is the probability of detecting an individual parasite; and N_i is the pathogen load of individual i.

The generalised likelihood for the Royle & Nichols (2003) abundance-induced heterogeneity (AIH) model takes the form:

$$L(w_i) = \sum_{k=0}^{\infty} \binom{T}{w_i.} p_k^{w_i.} (1 - p_k)^{T - w_i.} f_k$$

where (in this case) w_i is the number of detections observed in T repeated diagnostic tests of individual i and detection probability (p) varies as a function of the unknown sample-specific pathogen abundance, 'k' (such that $p_k = p_i$ when $N_i = k$; and $f_k = Pr(N_i = k)$). The marginal likelihoods of the detection histories in the AIH model are obtained by integrating the binomial likelihood over all possible values of pathogen abundance for each individual (i.e. by summation to a value of k that, ideally, is sufficiently large to include the range of possible abundances), assuming that pathogen load across individuals follows some specified statistical distribution (f_k; see Appendix S1 for full details of the parameterisation and implementation for the specific AIH likelihood models utilised in this study). Estimates of detection probabilities (p) and prevalence (ψ) are obtained as derived parameters from this model (see Appendix S1). While this model was originally and strictly designed for temporal replication of surveys at sites, recent work has demonstrated its relevance to spatially replicated surveys (Gopalaswamy *et al.*, under review), under a restricted range of conditions, and thus its applicability to both repeated visual disease diagnoses and repeated diagnostic testing of single biological samples was obtained from hosts. We recognise that this model may not represent an exact solution to this particular sampling situation because the potential variation in parasitaemia abundance between each 5-μL aliquot of an individual's DNA sample is unaccounted for in our current modelling. However, we expect this variation to be negligible as a result of homogenisation and standardisation of DNA extracts prior to assay (see Molecular Methods earlier). Furthermore, the potential to use a negative binomial distribution (with the overdispersion parameter) to describe variation in pathogen abundance among individuals provides a means to accommodate for this extra source of variation (if any exist) that a more complex, exact model would permit.

FITTING SITE-OCCUPANCY MODELS TO PLASMODIUM INFECTION DATA IN BLUE TITS

Detection histories for individual blue tits were obtained according to the infection status (presence/absence of *Plasmodium* spp.) recorded for each of the six replicate qPCR performed, in the order that they were run. It is important to note that under the sampling protocol utilised in this study, the probability of detecting an individual parasite (r) will actually be a product of: (i) the probability that an individual parasite was sampled when blood was initially obtained from the host, and (ii) the probability that at least one copy of parasite DNA was subsequently detected by qPCR. This sampling protocol does not provide us with sufficient information to estimate these two probabilities separately. However, as low-intensity chronic *Plasmodium* infections are detected in blood films made from < 10 μL of blood (Valkiunas *et al.* 2008), the volume of blood obtained from hosts in

this study (~10–20 μL) should be sufficient to detect very low concentration of circulating parasites, and hence, we assume that the probability of sampling parasites during bleeding is high (i.e. close to 1·0).

We fitted a number of probabilistic models for estimating population prevalence and detection rates of *Plasmodium* to our data set of individual *Plasmodium* detection histories (MacKenzie *et al*. 2002; Royle & Nichols 2003). First, to determine whether *Plasmodium* detection varied among qPCR replicates or between batches run on different PCR plates, we fitted standard occupancy models (MacKenzie *et al*. 2002) in which p varied with replicate ('rep', a 6-level factor) or batch ('b', a 2-level factor indicating in which PCR plate the sample was run), or remained constant ('.') across all replicates. Next, to assess the influence of heterogeneity in parasite load on *Plasmodium* detection (and obtain improved estimates of prevalence), we developed and fitted a series of AIH models (Royle & Nichols 2003) to our data set. Note that because values of parasite load (DNA copy numbers) were only obtained for individuals that we detected as infected (i.e. they were conditional on pathogen detection), they could not be included as covariates for pathogen detection in the standard occupancy model. We fitted four parameterisations of the AIH models each with a different statistical distribution describing the variation in the abundance of parasites among hosts: an AIH model in which abundance of parasites varied among hosts according to a random Poisson process (commonly used for modelling animal distributions); an AIH model with an underlying negative binomial distribution, which is likely to better characterise the highly aggregated, or overdispersed, distributions common to many parasite species (Shaw, Grenfell, & Dobson 1998); and two variants of AIH models to account for an inflation in the number of uninfected individuals in our sample, which could occur from sampling excessively in areas with zero probability of infection (see Appendix S1 for details). In fitting these AIH models, we encountered difficulties during numerical optimisation in obtaining the maximum likelihood estimates of the parameters when the upper bound on parasite abundance, K, was set to a very high value. We therefore set the upper bound to the highest value that was computationally tractable and in our analysis, $K = 100$. With a larger data set containing more replicate diagnoses per individual, we would expect to be able to successfully run our models using higher values of K, which is preferable. Standard occupancy models were run in programme MARK (White & Burnham 1999). All AIH models were implemented and run using the statistical package R (http://cran.r-project.org; code provided in Appendix S2).

All models were ranked based on AIC (Burnham & Anderson 2002). We used the most parsimonious model identified (model with lowest AIC) to explore the relationship between host parasite load and *Plasmodium* detection rate, to determine the probability of a 'false negative' diagnosis at different parasite loads and to assess bias between naive and corrected prevalence estimates, as a function of increasing number of qPCR per individual. To achieve this last task, we created new detection history data sets based on increasing numbers of qPCR diagnoses from our original data set (four new data sets from 2 to 5 replicate reactions) and then fitted the preferred model to these new data sets. Bias was calculated as the underestimation of naive prevalence (the number of individuals with at least one positive diagnosis in any qPCR divided by the total number of individuals sampled) relative to the best estimate of corrected prevalence (from the data set with six qPCR).

Results

Model selection revealed unequivocal support for all AIH models over standard occupancy models, indicating that the probability of detecting *Plasmodium* infection in a blood sample by qPCR strongly depends on host parasite load (Table 1). The best supported of the AIH models allowed parasite abundance to vary according to a negative binomial distribution, which typifies the overdispersed distribution of parasite load per host individual that is common to many parasites (Shaw, Grenfell, & Dobson 1998; Table 1). Estimates from this model appear reliable as the computed correlation coefficients between parameters were acceptable (Table 2). Moreover, simulations undertaken by Royle & Nichols (2003) revealed that the negative binomial model produces unbiased estimation of prevalence under conditions commensurate with those of this study ($T = 5$; $R = 200$; $\mu = 1{\cdot}61$, $a = 0{\cdot}54$ and $r = 0{\cdot}2$). In addition, their simulation results also showed that estimator bias diminishes at higher levels of overdispersion and with increasing number of replicate diagnoses (conditions pertinent to this study), and thus, we are confident of the parameter estimates obtained from this model. Although models with an underlying zero-inflated negative binomial distribution also appeared to provide an adequate fit to the data (Table 1), the estimated parameters (r and μ) appeared non-identifiable, indicated by extremely high correlation coefficients (−0·99) between the two parameters. Model selection revealed no

Table 1. Model selection results for candidate occupancy models for the detection of *Plasmodium* by qPCR in blood samples obtained from wild blue tits

Model*	K†	AIC	ΔAIC	w_i‡
[r, μ, a] (AIH; Negative binomial)	3	688·255	0·00	0·728
[r, μ, a, π] (AIH; Zero-inflated negative binomial)	4	690·255	2·00	0·268
[r, λ, π] (AIH; Zero-inflated Poisson)	3	696·749	8·494	0·010
[r, λ] (AIH; Poisson)	2	710·402	22·147	0·000
ψ(.) p(.)	2	739·959	51·704	0·000
ψ(.) p(batch)	3	741·953	53·698	0·000
ψ(.) p(replicate)	7	746·705	58·450	0·000

*AIH, abundance-induced heterogeneity models. Parameters r, λ, μ and a are as defined in Royle & Nichols 2003 (and see Appendix S1); π is the zero-inflation parameter defined for both the negative-binomial and the Poisson AIH models; ψ = corrected prevalence; p = detection probability per qPCR (standard occupancy model; MacKenzie *et al*. 2002).
†Number of parameters.
‡Model weight.

Parameter	MLE (SE)	Correlation coefficients		
r	0·443 (0·073)	−0·789		
μ	2·430 (0·910)			0.639
a	0·329 (0·095)	−0·748		
ψ	0·503 (0·038)	–		

Table 2. Parameter estimates (and standard errors) from the best identified occupancy model for the detection of *Plasmodium* by six replicate qPCR performed on blood samples obtained from wild blue tits

r, probability of detecting an individual parasite; μ, the average parasite abundance, *a*, the overdispersion parameter and ψ, corrected prevalence (see Methods and Appendix S1 for details). Also shown are the correlation coefficients between model parameters.

evidence of variation in pathogen detection rates among qPCR replicates and little indication that detection rate varied between batches.

Estimates of the detection rate of an individual parasite(r) from the best model were high (Mackenzie & Royle 2005; Table 2), indicating that the qPCR assay utilised in this study is a very sensitive and efficient means of detecting *Plasmodium* infections in wild birds. This is revealed in far greater clarity by examining the relationship between parasite load and the *Plasmodium* detection probability (p) for a single qPCR replicate (i.e. one qPCR; Fig. 2a). In this study, a single qPCR was able to detect *Plasmodium* in infected blood samples with $>90\%$ probability, for all blood samples with a parasite load of greater than four DNA copies, and with almost complete accuracy for samples with >10 DNA copies (Fig. 2). This is a substantially improved assessment of qPCR sensitivity than that provided by the standard site-occupancy model for which p was simply estimated as 0·756 (model [psi(.) p(.)] from Table 1).

Given such high pathogen detection rates, it is not surprising that the probability of a false negative diagnosis was negligible for individuals with a parasite load above 10 DNA copies (Fig. 2b). Moreover, although false negative diagnoses were significantly more likely in hosts with very low parasite load, increasing the number of qPCR replicates performed substantially reduced this probability, such that with three replicate reactions per sample the probability of a false negative diagnosis became negligible ($<0{\cdot}03\%$) for individuals with a parasite load >2 DNA copies, and with four replicate reactions a false negative diagnosis would be highly improbable (Fig. 2b). As the number of infections missed after six qPCR was found to be very low (very little difference between corrected and naive estimates of prevalence, see Table 2 and discussion later), we suggest that very few infections in this population will go undiagnosed by this diagnostic technique.

Naïve estimates of population disease prevalence were not greatly biased, provided at least three qPCR per individual were performed (3% bias relative to the best estimate of corrected population prevalence), with negligible bias ($<2\%$) achieved with four reactions per individual (Table 3). The preferred negative binomial AIH model could not be fitted to the detection history data set containing only two replicate qPCR per individual because of over-parameterisation relative to the sample size (Table 3). In addition, this model when applied to the detection history data set containing three replicate qPCR per individual gave an apparent overestimate of the corrected population prevalence. These results indicate that sufficient diagnostic testing per individual will be necessary to support

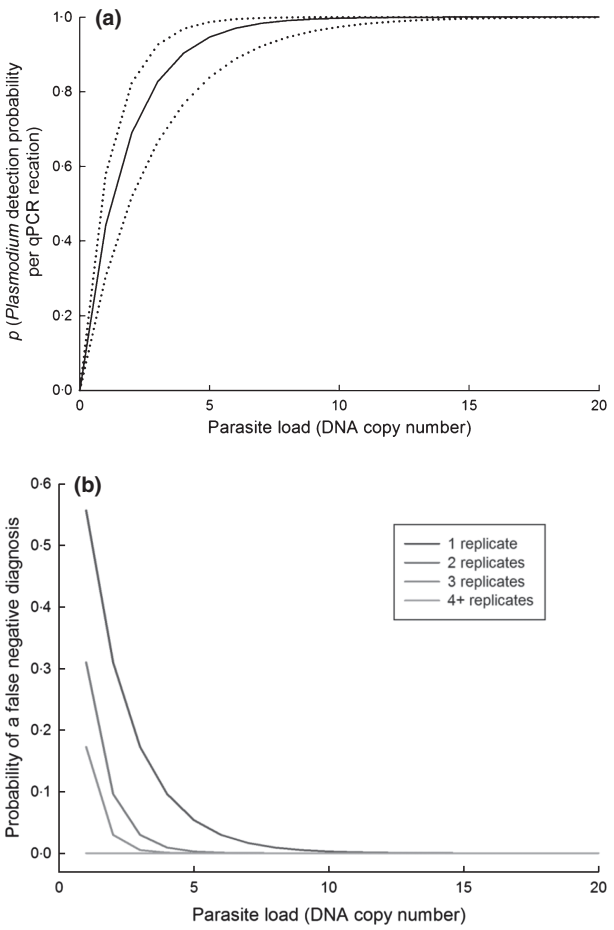


Fig. 2. (a) Relationship between parasite load and the individual-specific detection probability of *Plasmodium* infection by a single qPCR in blue tits. Individual-specific detection probability is calculated as $p = 1 - (1 - r)^{N_i}$, where N_i = host parasite load (DNA copy number in this case) and r (the per parasite detection probability) = 0·443. 95% CI are shown as dotted lines; (b) Probability of a false negative diagnosis as a function of the estimated sample parasite load (DNA copy number) and the number of replicate qPCR performed.

Table 3. Naive and corrected (ψ) estimates of *Plasmodium* prevalence (with standard errors) in a wild population of blue tits, as a function of the increasing number of qPCR performed. Corrected prevalence estimates were obtained by application of the preferred negative binomial AIH model to each of the data sets (see Methods for details)

No. qPCR	Naive prevalence	Corrected prevalence (SE)	Bias*
1	0·372	Not estimable	–
2	0·448	Not estimable	0·055
3	0·490	0·567 (0·081)	0·013
4	0·495	0·516 (0·096)	0·008
5	0·498	0·517 (0·074)	0·005
6	0·500	**0·503 (0·067)**	0·003

AIH, abundance-induced heterogeneity.
*Bias (underestimation) in naive estimates relative to the best estimate of corrected prevalence (shown in bold).

an assumption of a negative binomial distribution for parasite abundance in these AIH models. Consequently, a Poisson distribution (with fewer parameters) may be preferred when diagnostic tests performed per individual are limited.

Discussion

SITE-OCCUPANCY MODELLING AS A FLEXIBLE FRAMEWORK FOR ASSESSING TEST SENSITIVITY AND ESTIMATING WILDLIFE DISEASE PREVALENCE FROM IMPERFECT DIAGNOSTIC TESTS

Available statistical methods to account for imperfect diagnosis in the estimation of population disease prevalence and test sensitivity demand that a number of stringent requirements and assumptions be met (Greiner & Gardner 2000), limiting their utility in wildlife disease surveys. In this paper, we have shown that site-occupancy modelling provides a flexible alternative framework for deriving estimates of population prevalence and assessing test sensitivity under imperfect pathogen detection, without the need to make restrictive assumptions. Moreover, our novel application of AIH models (Royle & Nichols 2003) to pathogen detection data reveals that accounting for host variation in pathogen load substantially improves model fit and allows a more accurate assessment of diagnostic test sensitivity.

The results of our occupancy analyses of qPCR diagnosis of *Plasmodium* parasites obtained from wild blue tits revealed pathogen detection rates to be strongly dependent on host parasite load, supporting the findings of previous analyses in this population (Knowles *et al*. 2011). Because *Plasmodium* detection rate did not vary neither among qPCR replicates nor between different batches, we believe that this variation in detection rates by qPCR is indicative of the sensitivity of qPCR to the relative concentrations of parasite and host DNA within the extract, rather than the reproducibility of qPCR *per se*(Koru *et al*. 2007). Similarly, strong relationships between host parasite load and pathogen detection rates by qPCR have been observed in other wildlife disease studies (Kriger, Hero, & Ashton 2006; Grodio *et al*. 2008), including one study that employed an occupancy modelling approach (Gomez-Diaz *et al*. 2010). Unlike this study, Gomez-Diaz *et al*. (2010) were able to include host parasitaemia as a covariate for pathogen detection probability as values of parasite load were apparently not conditional on pathogen detection and could be assigned to all individuals. However, in most wildlife disease studies, variation in pathogen load, infection intensity or disease severity among hosts will not be quantifiable or, as in this study, may only be obtained for a subset of the population. The AIH modelling approach introduced in this study (adapted from Royle & Nichols 2003) allows pathogen detection rates to vary as a function of the underlying, but unknown, sample-specific parasite abundance. Our application of these AIH models greatly extends the utility of standard occupancy models for pathogen detection data, to offer a far more flexible and widely applicable approach than that utilised in recent applications of occupancy models (Gomez-Diaz *et al*. 2010; McClintock *et al*. 2010).

Our occupancy modelling results also revealed that the qPCR assay utilised in this study is a highly sensitive means of detecting *Plasmodium* infections in blood samples of wild birds. Parameter estimates indicated that the infection status of a large majority of individuals in our population could be determined with a very high degree of accuracy using just a single qPCR. Superior detection rates are a well-recognised and valued feature of a qPCR approach to pathogen detection, evidenced by the increasing use of qPCR for diagnosing early, pre-clinical infections and sub-patent infections that are otherwise impossible to detect (Grodio *et al*. 2008; Pollock *et al*. 2010). Nevertheless, the sensitivity of qPCR can decline rapidly when protocols or primers are not adequately optimised, when published protocols are adapted for use with different pathogens, or even when published protocols are used in different laboratories (Farcas *et al*. 2004). Moreover, the direct influence of parasite load on qPCR detection rates observed in this and other studies implies that the sensitivity of qPCR methods will also be reduced in populations or in species that harbour lower parasite loads (as a result of superior population health or host immunity). Hence, although there was little difference between naive and corrected estimates of *Plasmodium* prevalence in our study population, this is clearly a system-specific and assay-specific result and substantially greater bias will result when pathogen detection rates are lower. In these circumstances, the application of occupancy models to pathogen detection data will result in substantially more accurate estimation of infection status.

SITE-OCCUPANCY MODELS AS A TOOL FOR OPTIMISATION OF DIAGNOSTIC METHODS AND PROTOCOLS

All disease surveys will be subject to trade-offs between increased sampling effort to improve parameter estimates and increased expenditure involved in processing samples. This will be particularly true for surveys of wildlife and of non-zoonotic diseases, for which allocated research funds are generally more limited. The consequences of these trade-offs for experimental protocols and study designs include substantially limited sampling (at the population or individual levels) or the use of cheaper but less efficient diagnostic methods. For example, the slow uptake of qPCR technology by researchers studying chytridiomycosis infection in amphibians was recently attributed to the high costs of laboratory analysis via this method compared to traditional microscopy (Kriger, Hero, & Ashton 2006). The results of this present study highlight several ways in which occupancy modelling can serve as a statistically rigorous tool to refine and optimise diagnostic protocols for disease detection and thus enhance cost-efficiency in wildlife disease surveys. The direct evaluation of the influence of disease severity as well as other potential covariates on pathogen detection rates within a model selection framework allows for the clear identification of factors that limit or enhance pathogen detection. Once identified, the influence of these factors on pathogen detection can then be mitigated or promoted before further use (e.g. via laboratory optimisation), or be statistically controlled for in subsequent analyses (as will be necessary in our study system to control for the effect of host parasite load on pathogen detection rates). In addition, where pathogen detection rates of less than one can be tolerated, adopting an occupancy modelling approach will still enable more robust inferences to be made.

Occupancy modelling can also be used to maximise cost-efficiency in diagnostic protocols by identifying the optimal number of replicate diagnoses needed for accurate estimation of population prevalence. In this study, for example, we suggest that three qPCR per sample would be most cost-efficient to obtain minimally biased estimates of *Plasmodium* prevalence, but that four reactions are necessary to eliminate false negative diagnoses for hosts with very low parasite loads (< 2 DNA copies) and ensure accurate identification of disease state for all hosts. The possibility for disagreement between the optimal number of diagnoses needed for accurate prevalence estimation and the optimal number needed to accurately determine host infection status highlights that optimisation of diagnostic procedures will vary depending on the specific goals of the study. For example, studies that assess broad patterns of disease prevalence in wild populations will require only accurate estimates of population prevalence, whereas conservation breeding programmes or wildlife vaccination trials clearly require accurate identification of every infected individual in the population. In our broader study of the dynamics of endemic avian malaria infection in wild blue tits, individuals with very low parasite loads (those for which false negative diagnoses are likely) are likely less influential as mediators of disease spread and transmission because they probably harbour fewer transmission stage parasites to transmit to mosquitoes(Ross, Killeen, & Smith 2006). Hence, they are unlikely to be key drivers of disease dynamics in this population. Very few infected individuals in our blue tit-malaria system possessed parasite loads of < 2 DNA copies (although the true proportion of host harbouring such low parasitaemia values in our population cannot be known with certainty), and the probability of a false negative diagnosis with three qPCR for these individuals was not high ($< 20\%$). Thus, adopting a protocol of three qPCR per sample should provide improved cost-efficiency without severely compromising inferences regarding disease dynamics and host–pathogen relationships in this system.

An alternative solution to achieve both precision in estimates of population prevalence and the minimisation of false negatives may be to adopt a novel survey design known as a removal sampling scheme, in which disease diagnosis stops once the pathogen is detected (or once the maximum replicates have been performed; Mackenzie & Royle 2005). A recent simulation study found that a removal design could be more efficient than a standard design provided that neither the detection rate nor the occurrence of the species (pathogen prevalence) were particularly low ($< 0{\cdot}3$; Mackenzie & Royle 2005). The data yielded by a removal study, however, provides less flexibility for modelling and for exploring potential sources of variation in detection probability (e.g. for assessing the influence of pathogen load on pathogen detection rate) and should thus be implemented only after preliminary investigations of diagnostic protocols (Mackenzie & Royle 2005). The simulation study conducted by Mackenzie & Royle (2005) also revealed that when either pathogen detection probability (r) or pathogen prevalence (ψ) is very low ($< 0{\cdot}1$), then the number of replicate diagnoses needed will be very large. When population prevalence and pathogen detection rates are both low, then number of replicate diagnoses needed can be so large that it may become logistically or financially impossible to obtain an accurate estimate of prevalence or host infection status (Mackenzie & Royle 2005).

LIMITATIONS OF THIS APPROACH

One current limitation of the occupancy modelling approach proposed in this paper to obtain prevalence estimates corrected for imperfect diagnostic tests is the assumption that false positive diagnoses cannot occur. We are confident that this assumption was met in the present study as adherence to strict laboratory procedures ensured false positives results were negligible: results were only used when all negative controls on a PCR plate were clean and qPCR product melting temperature confirmed all positive detections as *Plasmodium* spp. (Knowles *et al.* 2011). Indeed, very high specificity (near 100%) has been observed in other applications of qPCR for pathogen detection (Farcas *et al.* 2004), indicating that the occupancy modelling approach proposed in this paper will be readily applicable where pathogen diagnosis occurs via qPCR. In addition, we can envisage other situations in which the possibility of a false positive diagnosis will be rare. For example, when diagnosis

occurs via visual inspection, disease symptoms are severe and when infected individuals are defined conservatively, there may be little likelihood of mistaking infected and uninfected individuals (e.g. chronic wasting disease, Williams *et al*. 2002 and devil facial tumour disease, S. Lachish unpublished data). However, when the probability of false positive diagnoses is appreciable, then the results obtained from the occupancy models described in this paper must be interpreted with caution, as the presence of false positives will cause positive bias in estimates of both p and ψ. This was revealed by Royle and Link (2006) who suggested a possible approach to accommodate false positives (which they considered to be misclassifications of species rather than positive detections in the absence of species occurrence) that relies on the relative magnitude of false positive and false negative rates being known. qPCR procedures provide an estimate of parasite load via the DNA copy number detected. If we observe unusually high detection rates at low parasite loads, such occurrences might be because of false positives. Adapting the occupancy models described in this paper to accommodate false positive diagnoses is the focus of this work.

The bias introduced to estimates of disease parameters as a result of imperfect diagnostic tests presents a pervasive problem for both epidemiology and wildlife disease management. This study has highlighted the value of an occupancy modelling approach, which allows the pathogen detection and infection processes to be explicitly integrated into models, over other statistical methods that attempt to correct prevalence estimates for imperfect diagnosis. An occupancy modelling approach for disease prevalence estimation does not require any *a priori* knowledge of population prevalence. A wide variety of diagnostic methods can be utilised to assess host infection status (from visual inspections to sophisticated molecular methods), with the only caveat being that the true infection status of the host remain unchanged between diagnoses. Importantly, the influence of heterogeneity in pathogen load among hosts on test sensitivity (pathogen detection probability) can be explicitly accounted for, without the need for specific information on pathogen abundance or disease severity *per se*. Finally, models with varying assumptions can be directly compared, and the influence of covariates on model parameters evaluated, within a model selection framework. This flexibility renders occupancy modelling an ideal framework for assessing population prevalence and test sensitivity in the face of pathogen detection bias arising from imperfect diagnostic tests in wildlife disease studies.

Acknowledgements

The work was funded by NERC grant NE/F005725/1 to BCS and by a NERC studentship to SCLK. We are grateful to numerous people for field assistance and to Jim Nichols and an anonymous referee for comments.

References

Acevedo-Whitehouse, K. & Cunningham, A.A. (2006) Is MHC enough for understanding wildlife immunogenetics? *Trends in Ecology & Evolution*, **21**, 433–438.

Adams, M.J., Chelgren, N.D., Reinitz, D., Cole, R.A., Rachowicz, L.J., Galvan, S., McCreary, B., Pearl, C.A., Bailey, L.L., Bettaso, J., Bull, E.L. & Leu, M. (2010) Using occupancy models to understand the distribution of an amphibian pathogen, Batrachochytrium dendrobatidis. *Ecological Applications*, **20**, 289–302.

Burnham, K.P. & Anderson, D.R. (2002) *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*. Springer-Verlag, New York.

Conn, P.B. & Cooch, E.G. (2009) Multistate capture-recapture analysis under imperfect state observation: an application to disease models. *Journal of Applied Ecology*, **46**, 486–492.

Cosgrove, C.L., Wood, M.J., Day, K.P. & Sheldon, B.C. (2008) Seasonal variation in *Plasmodium* prevalence in a population of blue tits *Cyanistes caeruleus*. *Journal of Animal Ecology*, **77**, 540–548.

Drewe, J.A., Dean, G.S., Michel, A.L. & Pearce, G.P. (2009) Accuracy of three diagnostic tests for determining Mycobacterium bovis infection status in live-sampled wild meerkats (Suricata suricatta). *Journal of Veterinary Diagnostic Investigation*, **21**, 31–39.

Enoe, C., Georgiadis, M.P. & Johnson, W.O. (2000) Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Preventive Veterinary Medicine*, **45**, 61–81.

Farcas, G.A., Zhong, K.J.Y., Mazzulli, T. & Kain, K.C. (2004) Evaluation of the RealArt malaria LC real-time PCR assay for malaria diagnosis. *Journal of Clinical Microbiology*, **42**, 636–638.

Faustino, C.R., Jennelle, C.S., Connolly, V., Davis, A.K., Swarthout, E.C., Dhondt, A.A. & Cooch, E.G. (2004) *Mycoplasma gallisepticum* infection dynamics in a house finch population: seasonal variation in survival, encounter and transmission rate. *Journal of Animal Ecology*, **73**, 651–669.

Gomez-Diaz, E., Doherty, P.F., Duneau, D. & McCoy, K.D. (2010) Cryptic vector divergence masks vector-specific patterns of infection: an example from the marine cycle of Lyme borreliosis. *Evolutionary Applications*, **3**, 391–401.

Greiner, M. & Gardner, I.A. (2000) Epidemiologic issues in the validation of veterinary diagnostic tests. *Preventive Veterinary Medicine*, **45**, 3–22.

Grodio, J.L., Dhondt, K.V., O'Connell, P.H. & Schat, K.A. (2008) Detection and quantification of Mycoplasma gallisepticum genome load in conjunctival samples of experimentally infected house finches (Carpodacus mexicanus) using real-time polymerase chain reaction. *Avian Pathology*, **37**, 385–391.

Jennelle, C.S., Cooch, E.G., Conroy, M.J. & Senar, J.C. (2007) State-specific detection probabilities and disease prevalence. *Ecological Applications*, **17**, 154–167.

Jones, M.E., Jarman, P., Lees, C., Hesterman, H., Hamede, R., Mooney, N., Mann, D., Pukk, C.E., Bergfeld, J. & McCallum, H. (2007) Conservation management of Tasmanian Devils in the context of an emerging, extinction-threatening disease: Devil Facial Tumor Disease. *EcoHealth*, **4**, 326–337.

Knowles, S.C.L., Wood, M.J., Alves, R., Wilson, G.J., Bensch, S. & Sheldon, B.C. (2011) Molecular epidemiology of malaria prevalence and parasitaemia in a wild bird population. *Molecular Ecology*, **20**, 1062–1076.

Koru, O., Qvarnstrom, Y., Slemenda, S.B., Xayavong, M., Johnston, S.P. & Da Silva, A.J. (2007) Evaluation of real-time PCR protocols for laboratory diagnosis of malaria. *American Journal of Tropical Medicine and Hygiene*, **77**, 232–233.

Kriger, K.M., Hero, J.M. & Ashton, K.J. (2006) Cost efficiency in the detection of chytridiomycosis using PCR assay. *Diseases of Aquatic Organisms*, **71**, 149–154.

Lachish, S., Jones, M. & McCallum, H. (2007) The impact of disease on the survival and population growth rate of the Tasmanian devil. *Journal of Animal Ecology*, **76**, 926–936.

Lively, C.M. (2006) The ecology of virulence. *Ecology Letters*, **9**, 1089–1095.

Mackenzie, D.I. & Royle, J.A. (2005) Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology*, **42**, 1105–1114.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. & Langtimm, C.A. (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**, 2248–2255.

McClintock, B.T., Nichols, J.D., Bailey, L.L., MacKenzie, D.I., Kendall, W.L. & Franklin, A.B. (2010) Seeking a second opinion: uncertainty in disease ecology. *Ecology Letters*, **13**, 659–674.

Palinauskas, V., Kosarev, V.V., Shapoval, A.P., Bensch, S. & Valkiūnas, G. (2007) Comparison of mitochondrial cytochrome b lineages and morphospecies of two avian malaria parasites of the subgenera *Haemamoeba* and *Giovannolaia* (*Haemosporida: Plasmodiidae*). *Zootaxa*, **1626**, 39–50.

Pollock, F.J., Morris, P.J., Willis, B.L. & Bourne, D.G. (2010) Detection and Quantification of the Coral Pathogen Vibrio coralliilyticus by Real-Time PCR with TaqMan Fluorescent Probes. *Applied and Environmental Microbiology*, **76**, 5282–5286.
Poulin, R. (2007) *Evolutionary Ecology of Parasites*. Princeton University Press, Princeton.
Ross, A., Killeen, G. & Smith, T. (2006) Relationships between host infectivity to mosquitoes and asexual parasite density in Plasmodium falciparum. *American Journal of Tropical Medicine and Hygiene*, **75**, 32–37.
Royle, J.A. & Link, W.A. (2006) Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, **87**, 835–841.
Royle, J.A. & Nichols, J.D. (2003) Estimating abundance from repeated presence-absence data or point counts. *Ecology*, **84**, 777–790.
Senar, J.C. & Conroy, M.J. (2004) Multi–state analysis of the impacts of avian pox on a population of Serins (*Serinus serinus*): the importance of estimating recapture rates. *Animal Biodiversity and Conservation*, **27**, 133–146.
Shaw, D.J., Grenfell, B.T. & Dobson, A.P. (1998) Patterns of macroparasite aggregation in wildlife host populations. *Parasitology*, **117**, 597–610.
Smith, G.C., Cheeseman, C.L., Wilkinson, D. & Clifton-Hadley, R.S. (2001) A model of bovine tuberculosis in the badger Meles meles: the inclusion of cattle and the use of a live test. *Journal of Applied Ecology*, **38**, 520–535.
Thompson, S.K. (2002) *Sampling*, 2nd edn. Wiley, New York.
Thompson, K.G. (2007) Use of site occupancy models to estimate prevalence of Myxobolus cerebralis infection in trout. *Journal of Aquatic Animal Health*, **19**, 8–13.
Toft, N., Jorgensen, E. & Hojsgaard, S. (2005) Diagnosing diagnostic tests: evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. *Preventive Veterinary Medicine*, **68**, 19–33.
Valkiunas, G., Iezhova, T.A., Krizanauskiene, A., Palinauskas, V., Sehgal, R.N.M. & Bensch, S. (2008) A comparative analysis of micropscopy and PCR-based detection methods for blood parasites. *Journal of Parasitology*, **94**, 1395–1401.
White, G.C. & Burnham, K.P. (1999) Program MARK: survival estimation from populations of marked animals. *Bird Study*, **46**, 120–138.
Williams, C.J. & Moffitt, C.M. (2010) Estimation of fish and wildlife disease prevalence from imperfect diagnostic tests on pooled samples with varying pool sizes. *Ecological Informatics*, **5**, 273–280.
Williams, E.S., Miller, M.W., Kreeger, T.J., Kahn, R.H. & Thorne, E.T. (2002) Chronic wasting disease of deer and elk: a review with recommendations for management. *Journal of Wildlife Management*, **66**, 551–563.
Wood, M.J., Cosgrove, C.L., Wilkin, T.A., Knowles, S.C.L., Day, K.P. & Sheldon, B.C. (2007) Within-population variation in prevalence and lineage distribution of avian malaria in blue tits, *Cyanistes caeruleus*. *Molecular Ecology*, **16**, 3263–3273.
Woodroffe, R., Frost, S.D.W. & Clifton-Hadley, R.S. (1999) Attempts to control tuberculosis in cattle by removing infected badgers: constraints imposed by live test sensitivity. *Journal of Applied Ecology*, **36**, 494–501.

Received 31 March 2011; accepted 17 August 2011
Handing Editor: Robert Freckleton

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Likelihood functions of the AIH models (Royle & Nichols 2003) with Poisson, negative binomial, zero-inflated Poisson and zero-inflated negative binomial processes for pathogen load (abundance).

Appendix S2. R code to obtain likelihood based parameter estimates for the AIH site-occupancy models (Royle & Nichols 2003), utilising Poisson, negative binomial and ZIP distributions of pathogen abundance.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be reorganized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Appendix L

Sinks as Saviours: Why flawed inference cannot assist tiger recovery

LETTER

Sinks as saviors: Why flawed inference cannot assist tiger recovery

A recent study of tigers in Chitwan, Nepal (1) stirred controversy by challenging the "source-sink" approach that underlies current global tiger conservation strategies (2). The observed lack of difference in tiger density estimates inside the protected area compared with a multiple-use area outside is offered as evidence. Based on this result, the study questions the relevance of strictly protected tiger reserves involving regulation of extractive uses and relocation of human settlements. The study offers an alternate vision of sustainable, syntopic "coexistence" of tigers and humans as a solution to increasing human resource demands on tiger habitats.

Long-term substantive studies contradict this alternative view. Tiger densities in better-protected reserves are three- to five-times higher than in areas with poorer protection and extractive human uses (2, 3). At other ecologically comparable sites, such as Kaziranga and Corbett (2), which curtail human uses, tiger densities are five-times higher (17–19 tigers per 100 km^2). Even in Chitwan, tiger densities were much higher historically, under stricter regulation of extractive uses (4).

Curiously, all evidence reported in this study itself strengthens the prevalent "source-sink" view. Naive tiger densities (the number of individual tigers divided by the area of the camera trap array) were 1.33- to 3-times higher inside the protected reserve. Additionally, between 2010 and 2011 the number of individual tigers photo-trapped within the reserve increased from 12 to 18, but it decreased from 6 to 4 outside. At a "fine scale" of measurement (combined trapping area ~80 km^2), this difference in tiger density is obvious. Furthermore, results from the small cell (1 km^2) occupancy analysis indicate that tiger densities are higher inside the reserve because detection probability is monotonically related to density at such cell sizes. Astonishingly, despite all contrary evidence, the authors conclude that estimated densities do not differ solely based on statistical insignificance. This anomaly likely arises because the spatially explicit capture-recapture models used are confronted with sparse data (only four to six tigers were photo-captured outside) in very small areas (50 km^2 inside, 30 km^2 outside), generating unreliable estimates of tiger densities (5).

Based on inference likely rooted in survey design flaws (5), the authors boldly attack a "prevailing belief" that tigers cannot spatially overlap with humans at fine spatial scales. Past studies show that tigers do coexist at "fine scales" in human-dominated landscapes, but with unacceptable levels of conflict, including predation on humans, as in Chitwan (4). Furthermore, although prescribing a radical reversal of current "source-sink" strategy focusing on protected areas, the authors' two-season study ignores multisite, long-term studies of tiger population dynamics that do support this approach (2–4).

Global spatial analyses show >70% of wild tigers now live and reproduce in 6% of source habitats in protected reserves, and their survival in surrounding landscape sinks depends on replenishment from sources (2–4). The sober reality is that the species has been extirpated from 93% of its former range because of conflict-ridden coexistence with humans. Hence, the authors' prescription of tiger population sinks over sources as a future recovery strategy is doomed to failure.

K. Ullas Karanth[a,b,1], Arjun M. Gopalaswamy[b,c], Krithi K. Karanth[b,d], John Goodrich[e], John Seidensticker[f], and John G. Robinson[a]

[a]*Global Conservation Program, Wildlife Conservation Society, Bronx, NY 10460;* [b]*Centre for Wildlife Studies, Bangalore, 560042, India;* [c]*Wildlife Conservation Research Unit, Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom;* [d]*Nicholas School of Environment, Duke University, Durham, NC 27708;* [e]*Tiger Program, Panthera, New York, NY 10018; and* [f]*Smithsonian Conservation Biology Institute, Washington, DC 20013*

1. Carter NH, Shrestha BK, Karki JB, Pradhan NMB, Liu J (2012) Coexistence between wildlife and humans at fine spatial scales. *Proc Natl Acad Sci USA* 109(38):15360–15365.
2. Walston JS, et al. (2010) Bringing the tiger back from the brink—The six percent solution. *PLoS Biol* 8(9):e1000485.
3. Karanth KU, et al. (2011) Monitoring carnivore populations at the landscape scale: Occupancy modelling of tigers from sign surveys. *J Appl Ecol* 48(4):1048–1056.
4. Barber-Myer SM, et al. (2012) Influence of prey depletion and human disturbance on tiger occupancy in Nepal. *J Zool*, 10.1111/j.1469-7998.2012.00956.x.
5. Sollman R, Gardner B, Lenot JL (2012) How does spatial study design influence density estimates from spatial capture-recapture models? *PLoS ONE* 7(4):e34575.

Author contributions: K.U.K., A.M.G., and K.K.K. designed research; K.U.K., A.M.G., K.K.K., J.G., J.S., and J.G.R. performed research; A.M.G. contributed new reagents/analytic tools; K.U.K. and A.M.G. analyzed data; and K.U.K., A.M.G., K.K.K., J.G., J.S., and J.G.R. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: ukaranth@gmail.com.

Appendix M

Patterns of human-wildlife conflicts and compensation: Insights from Western Ghats protected areas

Biological Conservation 166 (2013) 175–185

Contents lists available at SciVerse ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Patterns of human–wildlife conflicts and compensation: Insights from Western Ghats protected areas

Krithi K. Karanth [a,b,c,*], Arjun M. Gopalaswamy [a,d,e], Parvathi K. Prasad [a], Shreya Dasgupta [a]

[a] Centre for Wildlife Studies, 224, Garden Apartments, Vittal Malya Road, Bengaluru 560 001, India
[b] Nicholas School of Environment, Duke University, Durham, NC 27708, USA
[c] Wildlife Conservation Society, New York, 2300 Southern Boulevard, Bronx, NY 10460, USA
[d] Wildlife Conservation Research Unit, Department of Zoology, The Recanati-Kaplan Centre, University of Oxford, Tubney, Abingdon OX13 5QL, UK
[e] Wildlife Conservation Society – India Program, 1669, 31st Cross, 16th Main, Banashankari 2nd Stage, Bengaluru 560 070, India

ARTICLE INFO

Article history:
Received 7 May 2013
Received in revised form 22 June 2013
Accepted 23 June 2013

Keywords:
Compensation
Crop raiding
India
Livestock predation
Mitigation
Park
People
Wildlife

ABSTRACT

Preventing and mitigating human–wildlife conflicts are a top conservation priority, particularly in India where wildlife and high densities of people co-occur. We surveyed 1972 households from 1371 villages in a 7449-km^2 area surrounding five reserves in the Western Ghats. Our observational study modeled self-reported crop and livestock loss and compensation access by households. Crop loss was reported by 64% of households and associated with growing cotton, sugarcane, coffee and rice. Livestock loss was reported by 15% of households, and associated with grazing animals inside reserves. Losses incurred by households varied across reserves, averaging INR Rs 23,010 for crop loss and Rs 5423 for livestock loss. Compensation receipt was reported by 31% of households, and associated with reporting loss to authorities and elephant related incidents. Overall, landscape estimated probability of crop loss was 0.91, livestock loss was 0.19 and compensation was 0.29. Common mitigation measures for crop protection were night watching (46%), fencing (34%) and scare devices (34%); and for livestock protection were closer watch on animals (7%), guard animals (3%) and fencing (2%). Among 13 reported mitigation measures, no individual measure appeared to be associated with lowering crop or livestock loss. Unexpectedly, reported losses were similar across all reserves with higher losses incurred by households closer to the reserves. We identified conflict hot spots and influential factors across five reserves to improve current management efforts directed at conflict prevention and mitigation, and this approach is extendable to other human-dominated wildlife rich landscapes.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Addressing human–wildlife conflicts are among the most challenging issues that conservation practitioners face today. Despite decades of research and significant financial resources invested, we still lack a fundamental understanding of which ecological and social factors drive human–wildlife conflicts (Dickman, 2010). With conflict incidents gaining increasing attention, and in some cases resulting in retaliation against threatened species, it is imperative that we develop effective preventive and mitigation strategies (Naughton-Treves, 1998; Woodroffe et al., 2005; Linkie et al., 2007; Treves et al., 2009). Identifying commonalities across sites, gathering information on perceived conflict and compensation paid to affected people are all critical steps in this prioritization process (Madden, 2004; Baruch-Mordo et al., 2009; Dickman et al., 2011; Karanth et al., in press). Understanding influential factors and developing predictive risk maps are vital to developing preventive action against loss and improving compensation distribution to affected people (Treves et al., 2011; White and Ward, 2011).

In India, people exhibit some tolerance for wildlife despite experiencing substantial losses to different species (Sekhar, 1998; Mishra et al., 2003; Karanth and Nepal, 2012; Karanth et al., in press). Several species are reported to cause crop and livestock loss (Karanth and Madhusudan, 2002; Karanth et al., 2012). As documented elsewhere, the species involved in conflict seem to affect people' attitudes, tolerance, retaliation and reporting. For instance elephants *Elephas maximums* and tigers *Panthera tigris* related incidents are more likely to be reported than pigs, smaller canids or felids (Treves and Naughton-Treves, 2005). The co-occurrence of wildlife and people in many places and the potential for conflict between them provides an ideal context in which to examine these critical issues.

Our study aimed to examine patterns of conflict loss and compensation around five important reserves in Karnataka's

* Corresponding author at: Centre for Wildlife Studies, 224, Garden Apartments, Vittal Malya Road, Bengaluru 560 001, India. Tel.: +91 9900902041.
E-mail addresses: krithi.karanth@gmail.com (K.K. Karanth), arjungswamy@gmail.com (A.M. Gopalaswamy), parvathi68@gmail.com (P.K. Prasad), shreya.dg@gmail.com (S. Dasgupta).

0006-3207/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.biocon.2013.06.027

Western Ghats. The diversity of people and wildlife in this large 7500 km^2 landscape (Table 1) provided an ideal context to examine general and site-specific factors associated with conflict and compensation (Karanth et al., 2012). We examined which socio-economic and environmental factors are associated with an individual household's reporting of crop and livestock loss, and compensation access. We modeled landscape and household related environmental and agricultural factors to assess how losses and compensation are associated with conflict type, species and reporting to authorities. We expected households in closer proximity to reserves, surrounded by tree cover, and growing a diversity of palatable crops to be more prone to loss (Naughton-Treves, 1997; Hoare, 1999; Sitati et al., 2005). We expected households using mitigation measures (lighting, fencing, guarding of animals), and with more resources (land size, people to guard) to be less prone to loss (Karanth et al., 2012). We expected compensation access to be influenced by species (tiger, leopard *Panthera pardus*, elephant incidents more likely to be compensated), conflict type (human loss or injury better compensated than livestock and crop loss), respondent characteristics (age, land size) and reporting ability of households (Dickman et al., 2011). Lastly, we expected higher incidents of conflict and compensation for households around Bandipur, Nagarahole and Biligiri Rangan Temple reserves compared to Bhadra and Dandeli-Anshi, as these reserves have higher densities of wildlife (Karanth et al., in preparation).

2. Methods

2.1. Study sites

We selected five wildlife reserves in India's Western Ghats (Fig. 1 and Table 1). These reserves support globally significant populations of tiger *P. tigris*, leopard *P. pardus*, wild dog *Cuon alpinus*, elephant *Elephas maximus*, sambar *Cervus unicolor*, chital *Cervus axis*, and gaur *Bos gaurus* (Karanth et al., in preparation). Human population densities ranged between 135 and 443 people/km^2, and livestock densities ranged between 97 and 445 animals/km^2 in districts adjoining these reserves (Table 1).

2.2. Household surveys and landscape covariates

We surveyed 1972 individual households (Table 1) selected from 1371 surrounding villages in Karnataka state within a 10 km radius of each reserve (Fig. 1). We systematically sampled an area of 7449 km^2 by placing 573 grid cells (13 km^2 in size, Table 1, Karanth et al., 2012). This 10 km radius was chosen based on ecological reasons, specifically activity and movement patterns of wildlife in the area (Gopalaswamy et al., 2012), and policy reasons (the Karnataka government is establishing buffers around these reserves). The 13 km^2 grid size permits systematic sampling of the villages and wildlife species found in this landscape (Gopalaswamy et al., 2012). We digitized 2015 villages using Survey of India topographic maps and Google Earth imagery (v6.1). We randomly sampled at least 60% of villages in every cell, and in each village we opportunistically surveyed individual households (ensuring that households were located as far apart as possible). Our structured surveys with few open-ended questions were conducted in Kannada from February to June 2012. Teams of trained volunteers accompanied by research assistants conducted the surveys.

Individual adult male and female respondents were questioned specifically about conflict incidents (crop raiding, livestock predation and human injury or death) experienced in 2011–2012. We restricted questioning to the most recent year to minimize recall issues and confusion of incidents occurring over multiple years, as well as compensation reported by households to authorities and received. Additionally, we asked about demographic and socio-economic characteristics (types of crops grown, harvests, land area, etc.) and mitigation measures (fencing, lighting, guard animals etc.). For every household, we derived distance to reserve and water, and percentage of forest cover within 3 km of each household (this distance functions as reasonable estimate for daily activity ranges of ungulates in tropical forests, Gopalaswamy et al., 2012 and citations therein).

2.3. Variable and model selection

We selected variables that we hypothesized to be associated with crop and livestock loss and compensation reported by

Table 1
Details on five protected reserves in the Western Ghats.

Characteristics	Dandeli-Anshi	Bhadra	Nagarahole	Bandipur	BRT
Area	1303	492	644	880	540
Dominant vegetation	Tropical evergreen and semi-evergreen forests	Tropical moist deciduous, and evergreen forests	Tropical moist and dry deciduous forests	Tropical moist and dry deciduous forests	Tropical dry and moist deciduous, and evergreen forests
Rainfall in mm*	1250–4000	2000–2540	900–1500	625–1250	600–3000
Total area surveyed in km^2	1859	1807	1391	1313	1079
Total grid cells	245	174	119	122	119
Grid cells surveyed	143	139	107	101	83
Villages surveyed	304	316	274	262	215
Total HH Surveyed	468	471	400	370	263
Average HH members (Range)	7 (0–55)	5 (0–30)	5 (0–50)	5 (0–60)	6 (0–16)
Education < 8th grade	50%	63%	58%	47%	64%
Average livestock (Range)	5 (0–38)	4 (0–153)	3 (0–40)	4 (0–35)	3 (0–85)
Human population density in adjoining districts (persons per km^2)[a]	Belgaum: 356 Uttar Kannada: 140	Chikmagalur: 158 Devanagere: 326 Shimoga: 207	Kodagu: 135 Mysore: 443	Chamarajanagar: 180 Mysore: 443	Chamarajanagar: 180
Livestock density in adjoining districts (total livestock per km^2)[b]	Belgaum: 445 Uttar Kannada: 97	Chikmagalur: 182 Devanagere: 371 Shimoga: 203	Kodagu: 84 Mysore: 337	Chamrajnagar: 194 Mysore: 337	Chamrajnagar: 194

Note: Only grid cells with villages and people could be surveyed. HH refers to household.
Sources
* Nagarahole, Bandipur and Bhadra: DDTP Final CWS report, 2007. AD: Gururaja, K.V. & Ramachandra, T.V. (2012). Anuran Diversity and Distribution in Dandeli Anshi Tiger Reserve, ENVIS Technical Report 27, Centre for Ecological Sciences, Indian Institute of Science. BRT: Kumara, HN. Rathnakumar, S., Sasi, R. & Singh, M. (2012). Conservation status of wild mammals in Biligiri Rangaswamy Temple Wildlife Sanctuary, the Western Ghats, India, Current Science 103: 933–940.
[a] Census of India 2011.
[b] Livestock Census of India 2007.

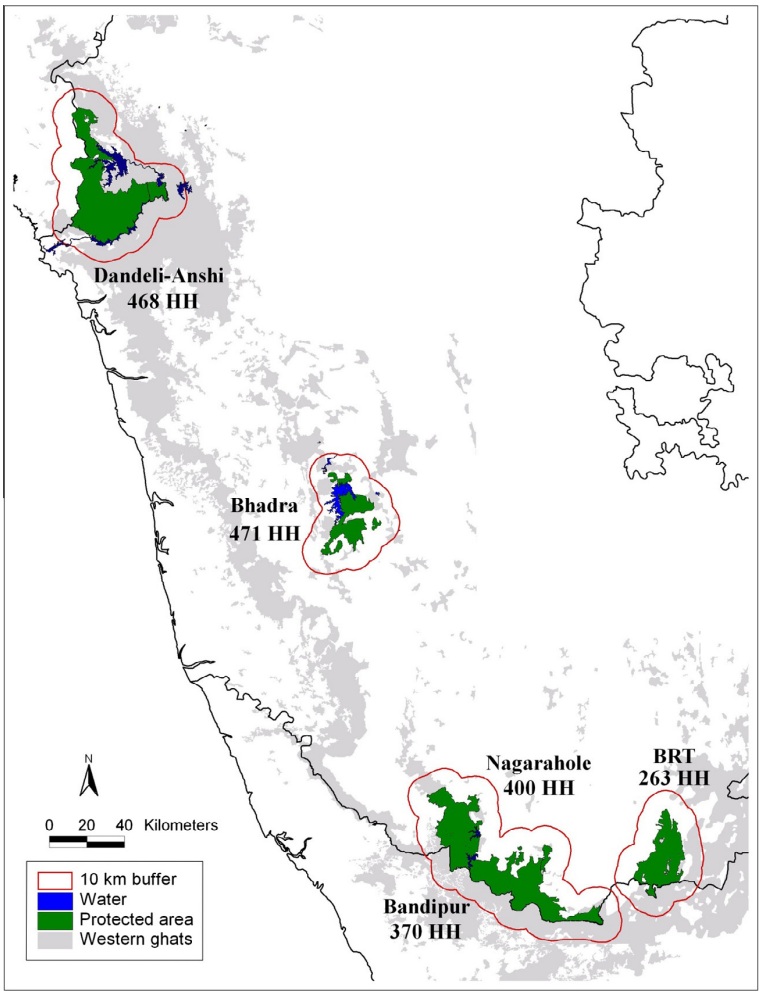


Fig. 1. Location of five reserves in the Western Ghats of Karnataka, India.

households (Table 2, Appendix 1a–c). We computed Pearsons' correlation coefficient to eliminate collinearity among variables of interest (Appendix 1a–c). Along with survey-based variables, we classified our five reserves based on known animal densities into high (Nagarahole, Bandipur and BRT), medium (Bhadra) and low (Dandeli-Anshi).

We modeled crop and livestock loss, and compensation access for each household across all reserves as a function household characteristics, agricultural and environmental factors. Since responses were binary, we fit logistic regression models to model individual household level loss and compensation. This allowed us to tease apart household to landscape level factors associated with losses and compensation access. We defined a priori model set, assessed model fit and identified those variables significantly associated with crop and livestock loss using Corrected Akaike's Information Criterion (AICc, Burnham and Anderson, 2002). The AICc weights represented the relative measure of appropriateness of a given model relative to the entire model set and ranked the models (Karanth et al., 2012). The top models (a subset of models accounting for >95% of the model uncertainty) included variables that provided the most optimal trade-off between parsimony and fit.

We constructed 28 crop loss, 19 livestock loss and 24 compensation access models based on factors that we hypothesized to be most relevant to loss and compensation (Appendix 2–4). We estimated crop loss, livestock loss and compensation access probabilities for every household using model averaging across all the candidate models (Burnham and Anderson, 2002). The weighted average was used to derive overall probability estimates for each households' crop loss, livestock loss and compensation access (for more details see Karanth et al., 2012).

2.4. Mapping loss and compensation

We applied spatial kriging to estimate crop loss, livestock loss and compensation to estimate and interpolate loss or compensation in non-sampled areas. In our case, the technique permits inferences about parameters of interest (crop loss, livestock loss and compensation access) in non-sampled areas in the entire landscape based on information in sampled locations (individual household estimates of crop loss, livestock loss and compensation for surveyed households) and accounting for uncertainties). Spatial kriging (without nugget effects) generated landscape level maps of hotspots for crop loss, livestock loss and compensation distribution probabilities for our entire 7500 km^2 landscape of interest (Karanth et al., 2012). We fit spherical models to the observed data based on AIC in Kriging Interpolator 3.2 for Spatial Analyst in ArcView to generate maps (Boeringa, 2003).

Table 2
Details on conflict and compensation reported by households surveyed in the Western Ghats.

Characteristics	Dandeli-Anshi	Bhadra	Nagarahole	Bandipur	BRT
HH reporting crop loss	67%	59%	58%	73%	67%
Average crop loss in Rs (range)	9,934 (0–60,000)	38,692 (0–800,000)	21,521 (0–700,000)	22,015 (0–450,000)	22,888 (0–700,000)
Number of crop raiding incidents	0–5	0–7	0–6	0–5	0–6
Top-ranked crop raiders (% of HH reporting)	Wild pig 43%	Wild pig 43%	Elephant 45%	Wild pig 53%	Wild pig 56%
	Elephant 19%	Elephant 15%	Wild pig 44%	Elephant 42%	Elephant 28%
	Gaur 12%	Bonnet macaque 13%	Chital 4%	Chital 3%	Bonnet macaque 5%
Top crop loss mitigation measures (% of HH reporting)	Night watching 55%	Scare devices 37%	Night watching 31%	Night watching 58%	Night watching 57%
	Fencing 44%	Fencing 36%	Scare devices 27%	Fencing 35%	Scare devices 36%
	Scare devices 44%	Night watching 35%	Fencing 25%	Scare devices 25%	Fencing 27%
HH reporting livestock loss	14%	16%	19%	15%	9%
Average livestock loss in Rs (range)	2954 (0–75,000)	7370 (0–500,000)	12352 (0–1,000,000)	2190 (0–50,000)	2249 (0–70,000)
Number of livestock predation incidents	0–3	0–2	0–2	0–2	0–3
Top-ranked predators	Tiger 65%	Tiger 64%	Leopard 62%	Leopard 62%	Leopard 69%
	Leopard 22%	Leopard 21%	Tiger 30%	Tiger 37%	Fox 17%
	Fox 8%	Fox 12%	Wild dog 4%	Fox 2%	Leopard 10%
Livestock loss Top mitigation measures	Closer eye on animals 9%	Closer eye on animals 9%	Night watching 7%	Night watching 5%	Closer eye on animals 2%
	Removal of waste or dead animals 4%	Use of guard animals 5%	Use of guard animals 4%	Closer eye on animals 3%	Removal of waste or dead animals 1%
	Use of guard animals 1%	Reduced use of public land 2%	Scare devices 4%	Fencing 3%	Scare devices 1%
					Night watching 1%
Households reporting loss to authorities:					
Crop loss	56%	47%	70%	66%	42%
Livestock loss	65%	60%	71%	70%	35%
Human death and injury	83%	17%	89%	77%	86%

Note: HH refers to household and Rs refers to Indian Rupees.

Table 3
Top ranked models with beta coefficients for crop loss, livestock loss and compensation.

Models	Crop loss	Livestock loss		Compensation			
	26 $w_i = 1$	16 $w_i = 0.59$	17 $w_i = 0.38$	2 $w_i = 0.60$	4 $w_i = 0.26$	1 $w_i = 0.08$	3 $w_i = 0.04$
Intercept	0.97 (0.07)	−1.92 (0.07)	−1.93 (0.07)	−3.19 (0.17)	−3.19 (0.17)	−3.19 (0.17)	−3.19 (0.17)
Distance to park	−0.52 (0.16)	−0.76 (0.14)	−0.76 (0.14)	NA	NA	NA	NA
Distance to water	0.23 (0.15)	0.06 (0.13)	0.06 (0.13)	NA	NA	NA	NA
Forest cover	−0.08 (0.18)	NA	NA	NA	NA	NA	NA
Rice	1.01 (0.17)	NA	NA	NA	NA	NA	NA
Areca nut	0.08 (0.21)	NA	NA	NA	NA	NA	NA
Banana	0.61 (0.26)	NA	NA	NA	NA	NA	NA
Coffee	1.11 (0.21)	NA	NA	NA	NA	NA	NA
Ragi	0.59 (0.22)	NA	NA	NA	NA	NA	NA
Cotton	1.32 (0.25)	NA	NA	NA	NA	NA	NA
Maize	0.75 (0.23)	NA	NA	NA	NA	NA	NA
Turmeric	0.04 (0.31)	NA	NA	NA	NA	NA	NA
Sugarcane	1.14 (0.37)	NA	NA	NA	NA	NA	NA
Vegetable	0.37 (0.34)	NA	NA	NA	NA	NA	NA
Any mitigation measure against crop loss	2.98 (0.15)	NA	NA	NA	NA	NA	NA
Any mitigation measure against livestock loss	NA	1.64 (0.17)	1.65 (0.17)	NA	NA	NA	NA
Total livestock	NA	0.40 (0.13)	0.38 (0.13)	NA	NA	NA	NA
Grazing in park by cattle	NA	0.67 (0.16)	0.66 (0.16)	NA	NA	NA	NA
Park wildlife density	0.11 (0.18)	0.26 (0.14)	0.27 (0.14)	NA	NA	NA	NA
Total number of people in household	NA	NA	0.12 (0.11)	NA	NA	NA	NA
Crop loss reported to forest department	NA	NA	NA	2.40 (0.28)	2.40 (0.28)	2.40 (0.28)	2.40 (0.28)
Livestock loss reported to forest department	NA	NA	NA	0.10 (0.30)	0.09 (0.30)	0.10 (0.30)	0.09 (0.30)
Human injury reported to forest department	NA	NA	NA	NA	−0.24 (0.40)	NA	−0.25 (0.40)
Crop loss by elephant	NA	NA	NA	1.68 (0.22)	1.69(0.22)	1.65 (0.22)	1.66 (0.22)
Crop loss by pig	NA	NA	NA	0.30 (0.23)	0.28 (0.23)	0.30 (0.23)	0.28 (0.23)
Crop loss by chital	NA	NA	NA	NA	NA	0.16 (0.23)	0.17 (0.23)
Crop loss by gaur	NA	NA	NA	NA	NA	−0.28 (0.28)	−0.29 (0.28)
Crop loss by bonnet macaque	NA	NA	NA	NA	NA	−0.28 (0.37)	−0.28 (0.37)
Livestock loss by tiger	NA	NA	NA	0.20 (0.30)	0.21 (0.30)	0.23 (0.30)	0.24 (0.30)
Livestock loss by leopard	NA	NA	NA	0.24 (0.30)	0.24 (0.28)	0.23 (0.28)	0.23 (0.28)
Model AICc	1449.59	1506.49	1507.4	1020.501	1022.157	1024.49	1026.109
Δ AICc	0	0	0.91	0	1.66	3.99	5.61

Note: Standard errors in brackets and top-ranked models are shown, wi is the AIC model weight ΔAIC is the difference in values between lowest AIC model and each model.

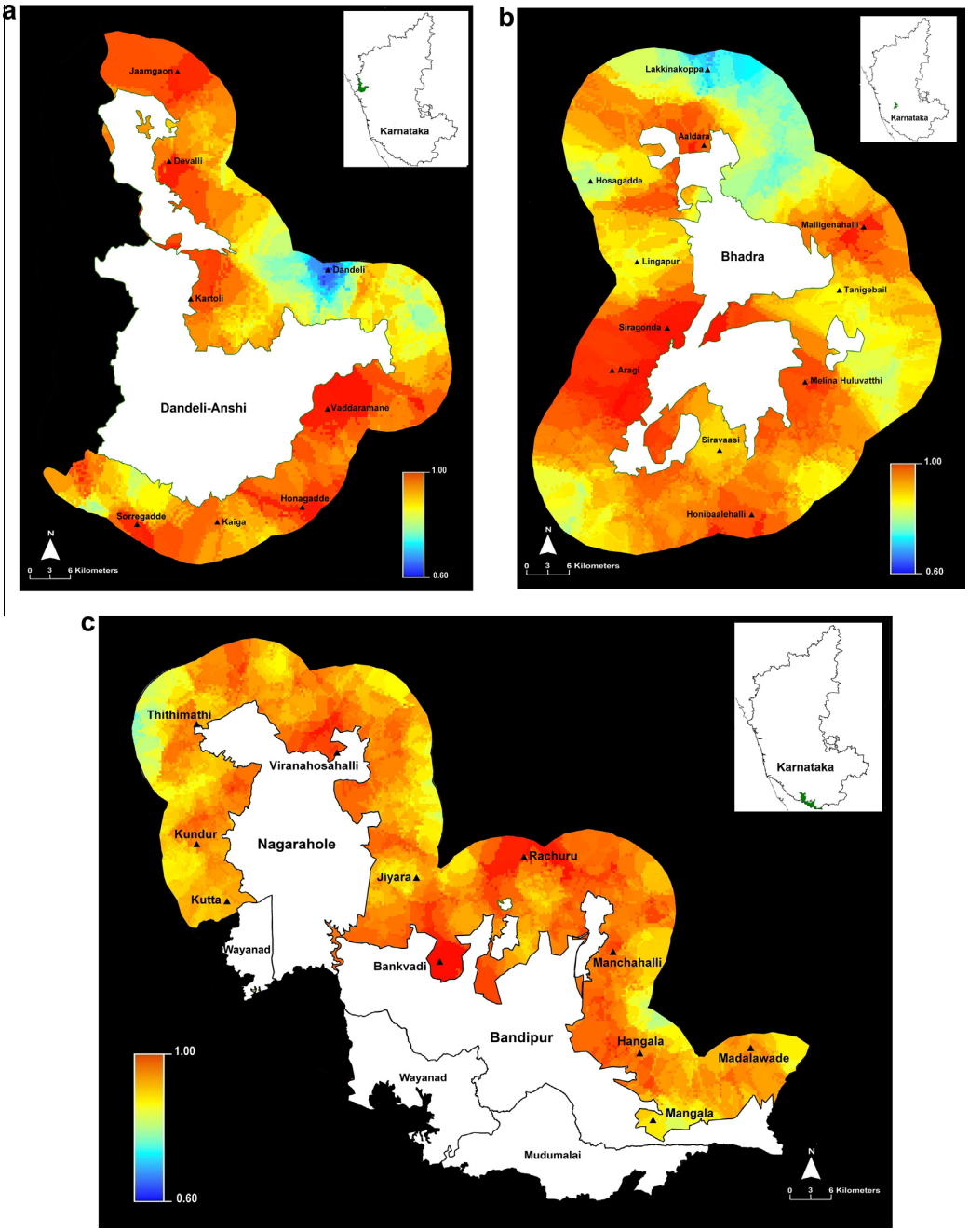


Fig. 2. Estimated crop loss around: (a) Dandeli-Anshi; (b) Bhadra; (c) Nagarahole and Bandipur; (d) BRT.

3. Results

3.1. Crop loss

We surveyed 1972 households around five reserves. Most (76%) households were primarily agricultural (ranging from 73% in BRT and Bhadra to 83% in Bandipur). The average size of agricultural land owned was 9 acres (ranging from 4 acres in Dandeli-Anshi, Bandipur and BRT to 22 acres in Bhadra). The region supported high crop diversity with 49 crops reportedly grown by households (ranging from 27 crops in Nagarahole to 47 crops in BRT). The major crops grown by households were rice (33%), ragi (20%), maize (17%), areca nut (16%), coffee (14%), cotton (13%) and banana (10%). Crop raiding incidents were experienced throughout the year, and were slightly higher during October through December months.

Sixty-four percent of surveyed households reported crop loss, ranging from 58% in Nagarahole to 73% in Bandipur. Crop loss was attributed to 19 species, with top ranked species across all reserves being wild pig (57%), elephant (37%) and chital (8%, details

in Table 2). The number of reported incidents averaged 1 per household annually across all reserves (ranging from 0 to 7, Table 2).

We modeled factors associated with household reported crop loss and found a single top ranked model (Table 3, Appendix Table 2). As predicted, distance to reserve was associated with lower crop loss (negative β coefficient in Table 3), while distance to water was not strongly associated with loss. The presence of forest cover in proximity to a household was weakly associated with loss but high standard errors suggest caution while interpreting results. Among individual crops, households growing cotton, sugarcane, coffee and rice were associated with higher loss (positive β coefficient in Table 3). Overall, the general use of mitigation measures appeared important, but no individual measure (fencing, lighting, night watching) appeared to be strongly associated to lowering crop loss (Table 3). We did not find any other factors associated with crop loss (Appendix 2).

We estimated and mapped probabilities of crop loss for surveyed households around all reserves (Fig. 2a–d). We found crop loss was high in most places, average estimated probability of crop loss was 0.91 (SE = 0.003), and similar across reserves (ranging from 0.90 in Bhadra to 0.93 in Bandipur, Table 4). Spatial modeling based on ordinary kriging suggested that households located closer to the PA experienced higher crop loss. The vulnerable villages and hot spots of crop loss around all reserves are in Fig. 2a–d.

3.2. Livestock loss

Fifteen percent of surveyed households reported livestock losses, ranging from 9% in BRT to 19% in Nagarahole. Losses were attributed to nine species, with tiger (8%) and leopard (7%) being top ranked predators (Table 2). A parallel study (Karanth et al., in preparation) assessed wildlife occurrence around these reserves, using photographs and behavioral descriptions to confirm species identification abilities of people. Reported predation incidents ranged from 0 to 3 per year, and were similar across all reserves during all months of the year (Table 2).

We identified two top-ranked models from modeling factors associated with self-reported livestock loss (cumulative AICc > 0.95, Table 3, Appendix Table 3). As predicted distance to reserve was associated with lower livestock loss (negative β coefficient in Table 3), while distance to water was not strongly associated with loss. Overall, the general use of mitigation measures appeared important but no individual measure (closer eye on animals, night watching, guard animals etc.) appeared to be strongly associated with lowering livestock loss (Table 3). We found that livestock loss to be positively associated with grazing animals inside and to a lesser extent the number of livestock owned and the individual reserve itself (Table 3).

We estimated and mapped livestock loss for households around these selected reserves (Fig. 3a–d). We found that the average estimated probability of livestock was lower 0.19 (SE = 0.003), and similar across reserves (ranging from 0.18 in Dandeli-Anshi, Bhadra and BRT to 0.20 in Bandipur, Table 4). Spatial kriging identified hot spots and vulnerable villages around these reserves (Fig. 3a–d).

3.3. Compensation reporting and receipt by households

We found that 64% of surveyed households reported crop loss and 15% reported livestock loss to us. However, 57% households surveyed reported crop loss (ranging from 42% in BRT to 70% to Nagarahole) and 64% reported livestock loss to authorities (ranging from 35% in BRT to 71% in Nagarahole, Table 2). Seventy-seven percent of households reported human injury or death to authorities (ranging from 17% in Bhadra to 89% in Nagarahole, Table 2). Reported losses incurred by households averaged Rs 23,010 (1 US

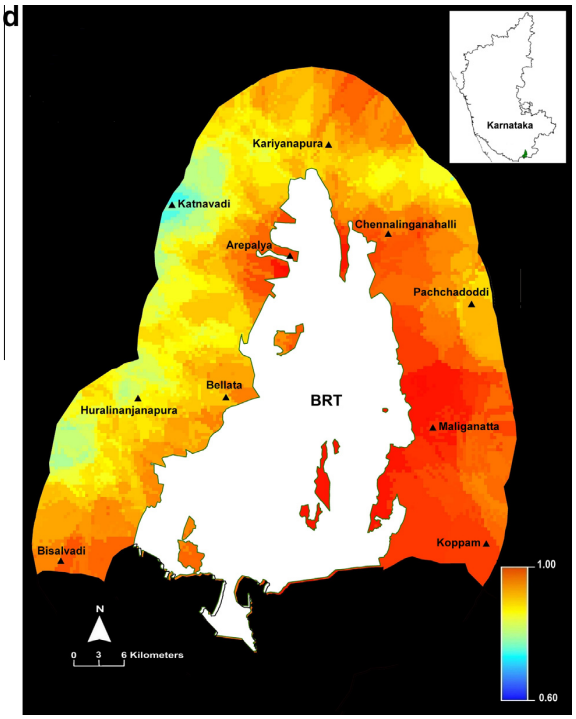


Fig. 2 (*continued*)

$ = INR Rs 50) for crop loss (ranging from Rs 0–8,00,000, Table 2) and Rs 5423 livestock loss (ranging INR Rs 0–1,00,000, Table 2). Reported crop losses varied across reserves ranging from INR Rs 9934 in Dandeli-Anshi to INR Rs 38,692 in Bhadra. Similarly, reported livestock losses varied across reserves ranging from INR Rs 2190 in Bandipur to INR Rs 12,352 in Nagarahole.

We modeled factors associated with reported access to compensation for surveyed households and found four top-ranked models (Table 3, Appendix Table 4). We found strong associations between households reporting losses to authorities and receiving compensation (positive β coefficients in Table 3). We expected species involved to influence compensation receipt by households and found strong positive associations for elephant related incidents (positive β coefficients in Table 3), and weaker associations for pig, leopard and tiger incidents (although for the latter three high standard errors suggest caution in interpreting these results).

We estimated, extrapolated and mapped compensation reported and received by surveyed households. The probabilities averaged 0.29 (SE = 0.007) and varied across reserves (ranging from 0.20 in Bhadra to 0.40 in Bandipur). The spatial models suggested that compensation was higher for some villages around each of the reserves (Fig. 4a–d and Table 4).

3.4. Mitigation measures

Households reported 13 mitigation measures used to protect crops, property and lives. The most common measures to protect crops were night watching (46%), fencing (34%) and scare devices (34%). To protect livestock households engaged in closer watch on animals (7%), use of guard animals (3%) and fencing (2%, Table 2). Our models suggested that although mitigation measures were generally positively associated with lowering loss no single measure was strongly associated with lowered crop or livestock loss (Table 3, Appendix 2 and 3).

Table 4
Estimated probabilities of crop loss, livestock loss and compensation for 1972 households surveyed in the Western Ghats (standard errors in brackets).

Type		Dandeli-Anshi	Bhadra	Nagarahole	Bandipur	BRT	All Reserves
Crop Loss	Average	0.91 (0.006)	0.90 (0.006)	0.91 (0.005)	0.93(0.005)	0.91(0.007)	0.91 (0.003)
	Range	0.57–1.00	0.62–1.00	0.62–1.00	0.63–1.00	0.61–1.00	0.57–1.00
Livestock loss	Average	0.18 (0.008)	0.18 (0.006)	0.19 (0.007)	0.20 (0.007)	0.18 (0.008)	0.19 (0.003)
	Range	0.04–0.76	0.06–0.98	0.07–0.79	0.07–0.75	0.05–0.89	0.04–0.98
Compensation	Average	0.26 (0.03)	0.20 (0.012)	0.33 (0.02)	0.40 (0.02)	0.25 (0.02)	0.29 (0.01)
	Range	0.04–0.84	0.04–0.85	0.04–0.85	0.04–0.85	0.04–0.85	0.04–0.085

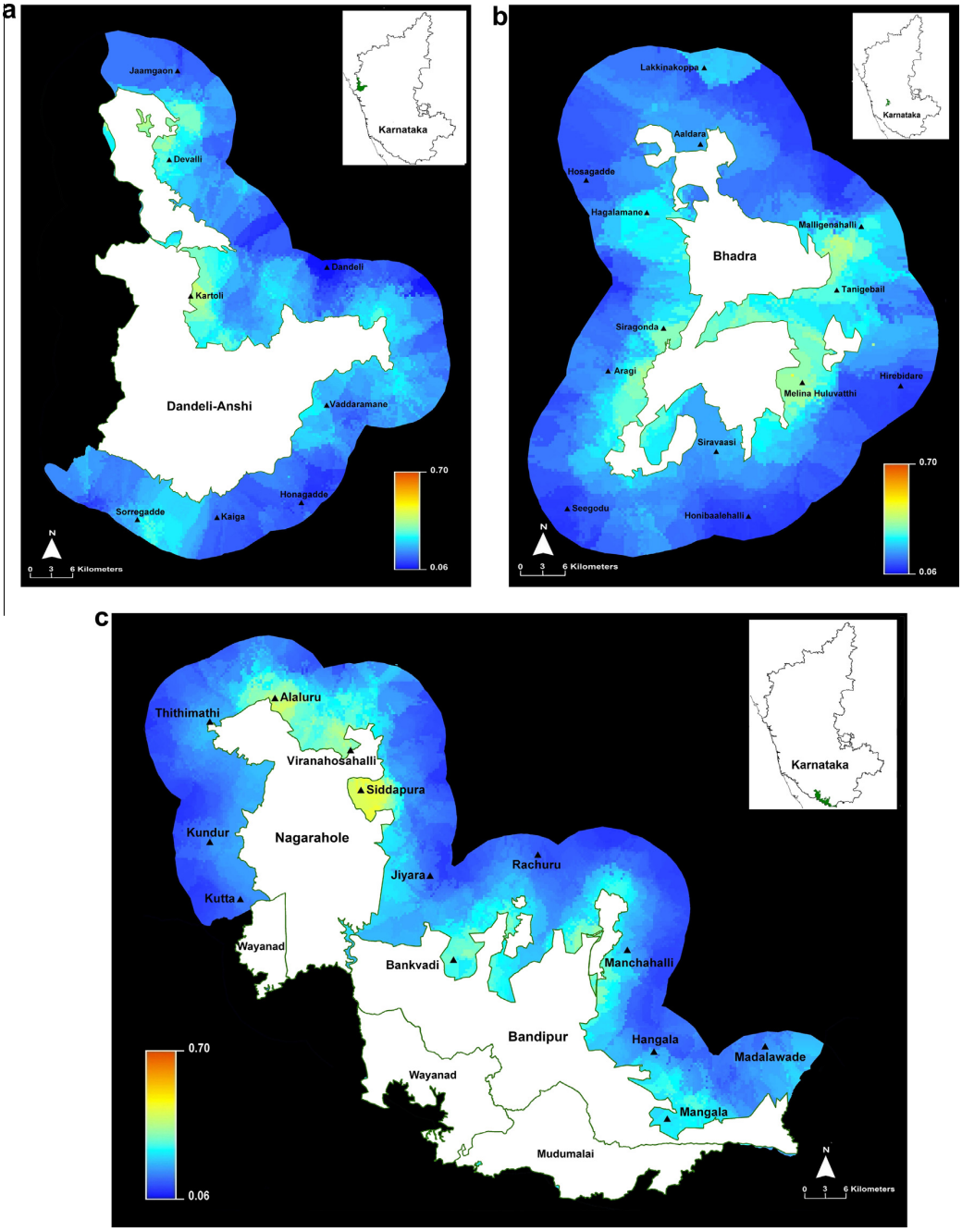


Fig. 3. Estimated livestock loss around: (a) Dandeli-Anshi; (b) Bhadra; (c) Nagarahole and Bandipur; (d) BRT.

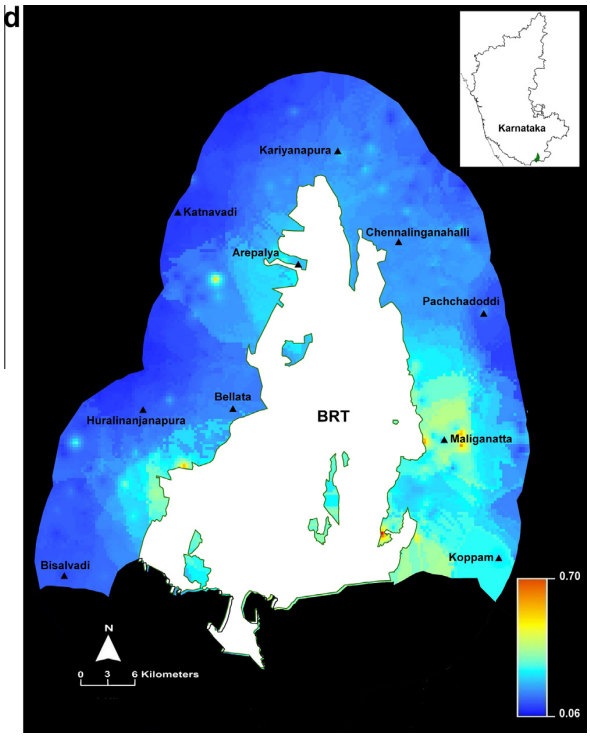


Fig. 3 (*continued*)

4. Discussion

Our assessment across five reserves in a large area (7500 km^2) offered several useful insights on conflict loss and compensation at the regional scale. Overall, the modeled and extrapolated for the study area was 91% for crop loss, which was higher than 19% for livestock loss. Our estimates for crop loss around reserves in the Ghats were comparable to losses in Central India, but livestock losses were much lower in the Ghats compared to Central India (Karanth et al., 2012). Unexpectedly, reported losses were similar across all reserves despite our expectation of higher losses around Nagarahole, Bandipur and BRT which support higher densities of wildlife, although spatial mapping suggests higher losses closer to all the reserves (Fig. 2a–d). Both crop and livestock loss were associated with proximity to reserve, which suggests that losses may be partially attributed to wildlife from the reserves, but there may be some species such as pigs and leopards that occur outside as well (Naughton-Treves, 1998; Karanth et al., in press). Additionally, crop loss was associated with growing cotton, sugarcane, coffee and rice. It is interesting that crops such as cotton and coffee that are less attractive to wildlife compared to sugarcane and rice (this was stated by respondents as well, Karanth personal observation 2012) were also raided. Our results suggest that some proportion of households will remain vulnerable simply due to their geographical location, so identifying mitigation measures that work for households on the edges of reserves is imperative for long-term success. Our results suggest that crop loss is associated with wildlife in the reserves and interestingly in some of these reserves successful conservation policies and protection have led to increase in densities of wildlife in the past 30 years making it imperative that hard edges around these reserves have to be managed wisely. In contrast, livestock loss was associated grazing animals inside which is common practice around many Indian reserves, and often difficult to restrict or modify (Karanth and Nepal, 2012; Karanth et al., 2012).

Many households reported experiencing loss to us, but a smaller proportion reported losses to authorities (highest for Nagarahole and lowest for BRT, Table 2) and reporting is higher than in Central India (Karanth et al., 2012). The compensation process requires households to file official documentation supported by evidence such as photographs of damage incurred. Field verifications are conducted by officials to assess damage, but often limited to cases where reported crop loss is high or involves death or injury to livestock and people. Many other crop loss incidents involving less damage are rarely examined. Typically, the compensation is provided immediately for human deaths and injuries but in other cases often takes more than a year to reach affected people. Non-reporting of losses has been corroborated by other studies in India (Sekhar, 1998; Mishra et al., 2003; Karanth et al., in press). Our respondents explained to us that the effort required to report an incident was both time consuming and costly, corroborating findings from other previous studies. Therefore, people were only inclined to report losses for high value species such as elephants and tigers, compared to others such as pigs and leopards (Karanth et al., personal observation 2012; Dickman, 2010).

The estimates of crop and livestock loss reported by households are comparable to official records of compensation provided to households in these areas (Karnataka Forest Department 2010). In Karnataka state for 2000–2010. In this period, the government provided compensation for almost 115,000 incidents of crop loss, property damage, livestock loss, human injury and death. The majority (>93%) of incidents were related to crop loss and livestock loss (4%). Our results indicated a need to improve the process of filing, verification and processing of claims, but also accommodating transaction costs that people incur during the claims process (Nyhus et al., 2003; Hemson et al., 2009).

Although these predictive models and maps have to be interpreted with caution, we have recommendations to improve existing conflict prevention and mitigation management practices. First, park and revenue authorities and non-governmental institutions can use our maps to target preventive action in the most vulnerable conflict prone villages around each of these reserves. There are examples where such live monitoring interventions are being tested elsewhere in India, for example with elephants in Valparai. Second, extending support to people in reporting conflict incidents and improving the processing of compensation claims is needed (Wilson et al., 2006; Kaartinen et al., 2009; Treves et al., 2009; Agarwala et al., 2010). Third, although generally mitigation measures were associated with lower losses we were unable to identify specific mitigation measures that reduced losses similar to our findings from Central India (Karanth et al., 2012). Therefore, rather than blindly investing in mitigation measures as it is often practiced today (Gore et al., 2008; Dickman et al., 2011), these measures have to be scientifically evaluated and monitored for effectiveness (Sitati et al., 2005; Gehring et al., 2010). Fourth, our results suggested that improving compensation process was clearly needed, but raised concerns about if compensation promotes tolerance and lowers retaliation against wildlife in the long term or raises expectations and hostility when requests are poorly dealt with (Ogra and Badola, 2008; Naughton-Treves et al., 2003; Mishra et al., 2003; Bulte and Rondeau, 2005; Nelson, 2009).

Our surveys suggested that individual respondent attitudes towards wildlife and these reserves vary widely and making it challenging to easily implement general strategies in such a diverse landscape (Webber et al., 2007; Campbell-Smith et al., 2012; Karanth and Nepal, 2012). The use of financial incentives and instruments (whether they be compensation, insurance or revenue sharing for example from tourism) must be applied with caution at scales that can be observed and monitored locally (Ferraro and

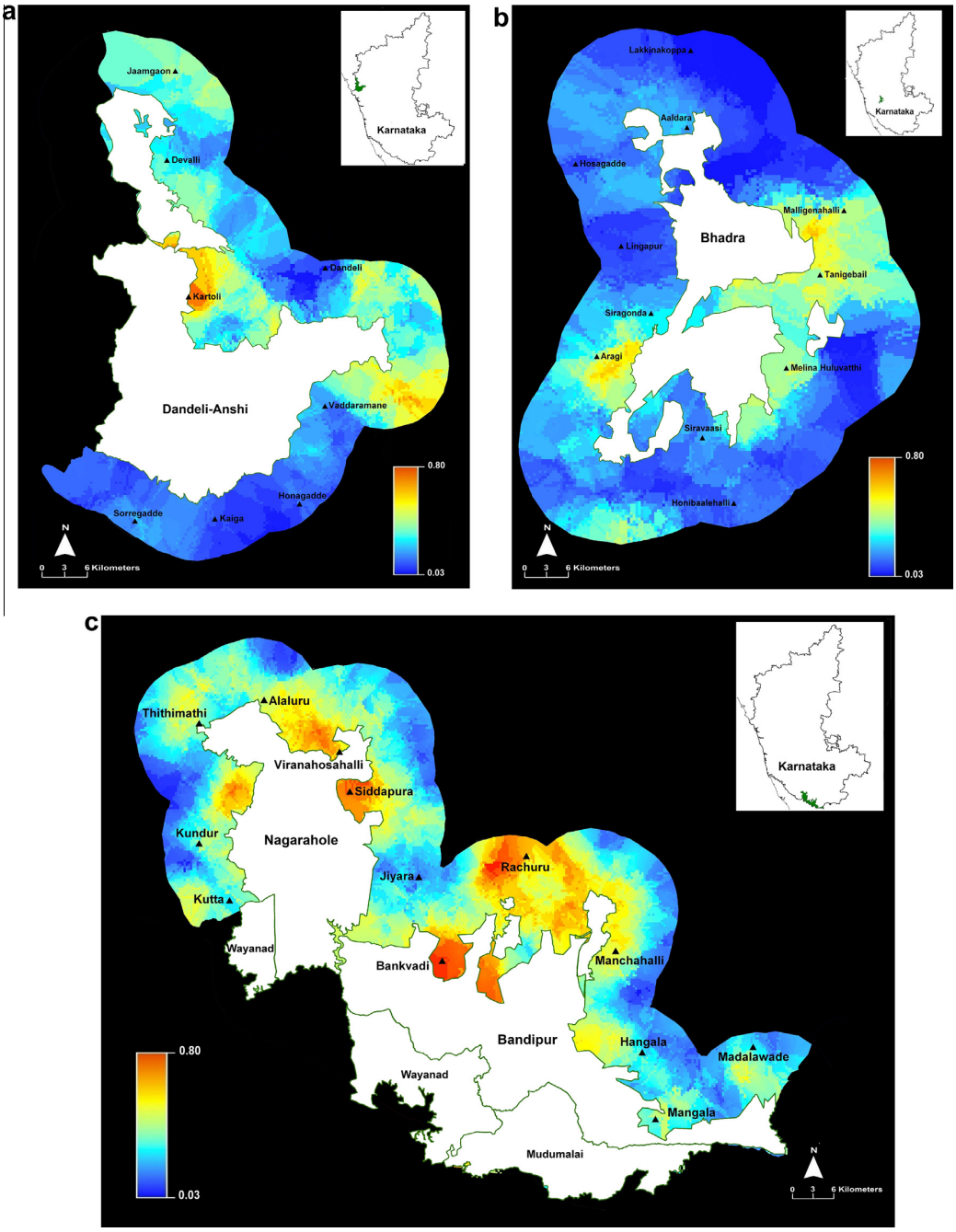


Fig. 4. Estimated Compensation access around: (a) Dandeli-Anshi; (b) Bhadra; (c) Nagarahole and Bandipur; (d) BRT.

Simpson, 2002; Zabel and Engel, 2010; Dickman et al., 2011). Some successful examples of such efforts in the US, Kenya, Sweden and India do exist (Mishra et al., 2003; Nyhus et al., 2003; Zabel and Holm-Müller, 2008; Maclennan et al., 2009).

Some limitations of our study include reliance on respondents' self-reporting and not directly monitoring actual incidents of conflict. Therefore, raising the potential for exaggeration of losses reported by people (Treves and Naughton-Treves, 2005). However, the large number of household surveys lowers the potential for widespread systematic bias, and comparable to similar work from Central India (Karanth et al., 2012). Further, follow up studies must be extended to timely, direct monitoring of temporal and spatial patterns of wildlife loss incidents.

We were able to survey almost 2000 households in a large area of 7500 km^2 independently to assess their perceptions and reporting of conflict loss and compensation in a critically important conservation region- the Western Ghats that supports a rich array of large mammal species that are living alongside very high densities

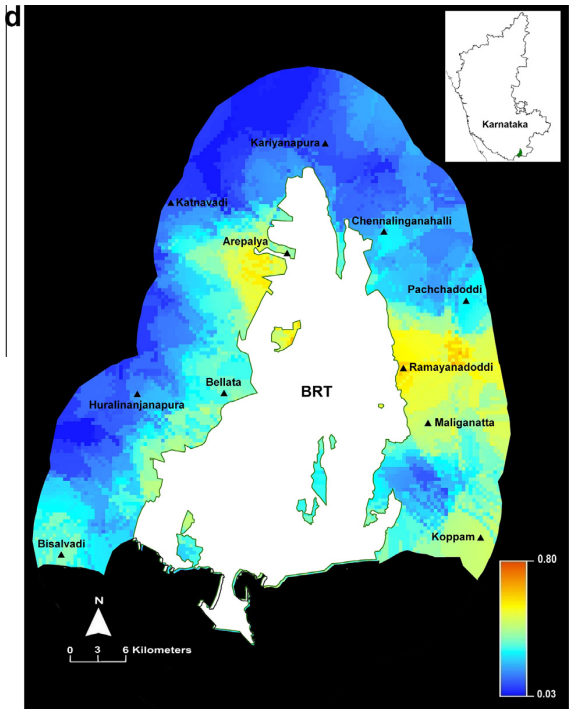


Fig. 4 (*continued*)

of people. The scale of our study and spatial rigor is much larger than all existing published studies that assess patterns of human–wildlife conflicts and compensation. The robustness of the sampling design, modeling and mapping approaches can easily be adapted to other landscapes and species across the world. Our approach permits easy understanding of basic patterns and spatially maps conflict loss and compensation across a large landscape so that we can better target conflict reporting, mitigation and compensation in places shared by wildlife and people.

Acknowledgements

We thank Natasha Ballal, PM Muthanna, DV Girish, U Karanth, NS Kumar, K Varma, Centre for Wildlife Studies and Wildlife Conservation Society-India program. We are grateful to the Karnataka forest department especially B.K Singh. We thank volunteers Abhimanyu, Adarsh, Aditya, Ajai, Akash, Akshatha, Amar, Amith, Anaga, Anand, Aparna, Aravind, Arjun, Arun, Avinash, Beeresh, Chetan, Chetana, Dayanand, Gagan, Hareesh, Hari, Harika, Jayalakshmi, Karthik, Kenneth, Kiran, Krithika, Kumar, Kumaraswamy, Madhu, Madhukar, Mahananda, Manjunath, Mohan, Mrunmayee, Mustak, Mythreyi, Nayana, Nihal, Nitin, Pavan, Phanindra, Pooja, Poornima, Pragna, Prakash, Pramod, Rachana, Rahul, Ramesh, Sampada, Shraddha, Shreenath, Sumanas, Sumanth, Swathi, Tarsh, Tarun, Thejaswi,Varsha, Vasundara, Venkatesh, Vinay K,Vinay M, Vinay P, Vinod, Vishnupriya, Vishwanath. DST Ramanujan Fellowship, National Geographic Society and Rufford grants supported Karanth.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.biocon.2013.06.027.

References

Agarwala, M., Kumar, S., Treves, A., Naughton-Treves, L., 2010. Paying for wolves in Solapur, India and Wisconsin, USA: Comparing compensation rules and practice to understand goals and politics of wolf conservation. Biol. Conserv. 143, 2945–2955. http://dx.doi.org/10.1016/j.biocon.2010.05.003.

Baruch-Mordo, S., Breck, S.W., Wilson, K.R., Broderick, J., 2009. A tool box half full: How social science can help solve human–wildlife conflict. Hum. Dimens. Wildl. 14, 219–223.

Boeringa, M., 2003. Kriging Interpolator 3.2 for ArcView Spatial Analyst. Nieuwland Automatisering, Wageningen, The Netherlands. <http://www.nieuwland.nl>.

Bulte, E.H., Rondeau, D., 2005. Research and management view point: why compensating wildlife damages may be bad for conservation. J. Wildlife Manage. 69, 14–19.

Burnham, K.P., Anderson, D., 2002. Model Selection and Multimodal Inference. A practical information theoretic approach. Springer-Verlag, New York.

Campbell-Smith, G., Sembiring, R., Linkie, M., 2012. Evaluating the effectiveness of human–orangutan conflict mitigation strategies. J. Appl. Ecol.. http://dx.doi.org/10.1111/j.1365-2664.2012.02109.x.

Dickman, A.J., 2010. Complexities of conflict: the importance of considering social factors for effectively resolving human–wildlife conflicts. Anim. Conserv. 13, 458–466.

Dickman, A., Macdonald, E., Macdonald, D., 2011. A review of financial instruments to pay for predator conservation and encourage human-carnivore coexistence. Proc. Natl. Acad. Sci. USA 108, 13937–13944.

Ferraro, P.J., Simpson, R.D., 2002. The cost-effectiveness of conservation payments. Land. Econ. 78, 339–353.

Gehring, T.M., VerCauteren, K.C., Provost, M.L., Cellar, A.C., 2010. Utility of livestock-protection dogs for deterring wildlife from cattle farms. Wildlife Res. 37, 715–721. http://dx.doi.org/10.1071/WR10023.

Gopalaswamy, A.M., Karanth, K.U., Kumar, N.S., Macdonald, D.W., 2012. Estimating tropical forest ungulate densities from sign surveys using abundance models of occupancy. Anim. Conserv.. http://dx.doi.org/10.1111/j.1469-1795.2012.00565.x.

Gore, M.L., Knuth, B.A., Scherer, C.W., Curtis, P.D., 2008. Evaluating a conservation investment design to reduce human–wildlife conflict. Conserv. Lett. 1, 136–145.

Hemson, G., Maclennan, S., Mills, G., Johnson, P., Macdonald, D., 2009. Community, lions, livestock and money: a spatial and social analysis of attitudes to wildlife and the conservation value of tourism in a human-carnivore conflict in Botswana. Biol. Conserv. 142, 2718–2725.

Hoare, R.E., 1999. Determinants of human–elephant conflict in a land-use mosaic. J. Appl. Ecol. 36, 689–700.

Kaartinen, S., Luoto, M., Kojola, I., 2009. Carnivore-livestock conflicts: Determinants of wolf (Canis lupus) depredation on sheep farms in Finland. Biodivers. Conserv. 18, 3503–3517.

Karanth, K.U., Madhusudan, M.D., 2002. Mitigating human–wildlife conflicts in southern Asia. In: Terborgh, J., van Schaik, C.P., Daven-port, L.C., Rao, M. (Eds.), Making Parks Work: Identifying Key Factors to Implementing Parks in the Tropics, Island Press, pp. 250–264.

Karanth, K.K., Nepal, S., 2012. Local perceptions of benefits and losses of living around protected areas in India and Nepal. Environ. Manage. 49, 372–386.

Karanth, K.K., Gopalaswamy, A.M., DeFries, R., Ballal, N., 2012. Assessing patterns of conflict risk and compensation distribution around a Central Indian protected area. PLoS One, 7(12), e50433. <http://dx.plos.org/10.1371/journal.pone.0050433>.

Karanth, K.K., Naughton-Treves, L., DeFries, R., Gopalaswamy, A.M. in press. Living with wildlife and mitigating conflicts around three Indian protected areas. Environ. Manage.

Linkie, M., Dinata, Y., Nofrianto, A., Leader-Williams, N., 2007. Patterns and Perceptions of wildlife crop raiding in and around Kerinci Seblat National Park. Sumatra Anim. Conserv. 10, 127–135.

Maclennan, S.D., Groom, R.J., Macdonald, D.W., Frank, L.G., 2009. Evaluating a compensation scheme to bring about pastoralist tolerance of lions. Biol. Conserv. 142, 2419–2427.

Madden, F., 2004. Creating coexistence between humans and wildlife: Global perspectives on local efforts to address human–wildlife conflict. Hum. Dimens. Wildl. 9, 247–257.

Mishra, C., Allen, P., Maccarthy, T., Madhusudan, M.D., Bayarjargal, A., Prins, H.H.T., 2003. The role of incentive programs in conserving the snow leopard. Conserv. Biol. 17, 1512–1520.

Naughton-Treves, L., 1997. Farming the forest edge: vulnerable places and people around Kebale National Park, Uganda. Geogr. Rev. 871, 27–46.

Naughton-Treves, L., 1998. Predicting patterns of crop damage by wildlife around Kibale National Park, Uganda. Conserv. Biol. 12, 156–168.

Naughton-Treves, L., Grossberg, R., Treves, A., 2003. Paying for tolerance: the impact of livestock predation and compensation payments on rural citizens attitudes towards wolves. Conserv. Biol. 17, 1500–1511.

Nelson, F., 2009. Developing payments for ecosystem service approaches to carnivore conservation. Hum. Dimens. Wildl. 14, 381–392.

Nyhus, P., Fischer, F., Madden, F., Osofsky, S., 2003. Taking the bite out of wildlife damage: the challenge of wildlife compensation schemes. Conserv. Prac. 4, 37–40.

Ogra, M., Badola, R., 2008. Compensating human–wildlife conflict in Protected Area Communities: ground level perspectives from Uttarkhand, India. Human. Ecol. 36, 717–729.

Sekhar, N., 1998. Crop and livestock depredation caused by wild animals in protected areas: the case of Sariska Tiger Reserve, Rajasthan, India. Environ. Conserv. 25, 160–171.
Sitati, N.W., Walpole, M.J., Leader-Williams, N., 2005. Factors affecting susceptibility of farms to crop raiding by African elephants: using a predictive model to mitigate conflict. J. Appl. Ecol. 42, 1175–1182.
Treves, A., Naughton-Treves, L., 2005. Evaluating lethal control in the management of human–wildlife conflict. In: Woodroffe, R., Thirgood, S., Rabinowitz, A. (Eds.), People and Wildlife: Conflict or Coexistence? Cambridge University Press, Cambridge, UK, pp. 86–106.
Treves, A., Wallace, R.B., White, S., 2009. Participatory planning of interventions to mitigate human–wildlife conflicts. Conserv. Biol. 23, 1577–1587.
Treves, A., Martin, K.A., Wydeven, A.P., Weidenhoeft, J.A., 2011. Forecasting environmental hazards and the application of risk maps to predator attacks on livestock. Bioscience 61, 451–458.
Webber, A.D., Hill, C.M., Reynolds, V., 2007. Assessing the failure of a community-based human–wildlife conflict mitigation project in Budongo Forest Reserve, Uganda. Oryx 41, 177–184.
White, P.C., Ward, A.I., 2011. Interdisciplinary approaches for the management of existing and emerging human–wildlife conflicts. Wildl. Res. 37, 623–629.
Wilson, S.M., Madel, M.J., Mattson, D.J., Graham, J.M., Merrill, T., 2006. Landscape conditions predisposing grizzly bears to conflicts on private agricultural lands in the western USA. Biol. Conserv. 130, 47–59.
Woodroffe, R., Thirgood, S., Rabinowitz, A., 2005. People and Wildlife, Conflict or Coexistence? Cambridge University Press, Cambridge, UK.
Zabel, A., Engel, S., 2010. Performance payments: a new strategy to conserve large carnivores in the tropics? Institute for environmental decisions, Zurich. USA. Biol. Conserv. 130, 47–59.
Zabel, A., Holm-Müller, K., 2008. Conservation performance payments for carnivore conservation in Sweden. Conserv. Biol. 22, 247–251.

Appendix N

Living with Wildlife and Mitigating Conflicts around three Indian Protected Areas

Environmental Management
DOI 10.1007/s00267-013-0162-1

Living with Wildlife and Mitigating Conflicts Around Three Indian Protected Areas

Krithi K. Karanth · Lisa Naughton-Treves · Ruth DeFries · Arjun M. Gopalaswamy

Received: 17 December 2012 / Accepted: 30 August 2013
© Springer Science+Business Media New York 2013

Abstract Crop and livestock losses to wildlife are a concern for people neighboring many protected areas (PAs) and can generate opposition to conservation. Examining patterns of conflict and associated tolerance is important to devise policies to reduce conflict impacts on people and wildlife. We surveyed 398 households from 178 villages within 10 km of Ranthambore, Kanha, and Nagarahole parks in India. We compared different attitudes toward wildlife, and presented hypothetical response scenarios, including killing the problem animal(s). Eighty percent of households reported crop losses to wildlife and 13 % livestock losses. Higher crop loss was associated with more cropping months per year, greater crop variety, and more harvest seasons per year but did not vary with proximity to the PA, suggesting that PAs are not necessarily "sources" for crop raiders. By contrast, complaints of "depredating carnivores" were associated with people-grazing animals and collecting resources from PAs. Many households (83 %) engaged in mitigation efforts. We found that only fencing and guard animals reduce crop losses, and no efforts to lower livestock losses. Contrary to our expectations, carnivores were not viewed with more hostility than crop-raiding wildlife. Households reported greater inclination to kill herbivores destroying crops or carnivores harming people, but not carnivores preying on livestock. Our model estimated crop loss was 82 % across surveyed households (highest in Kanha), while the livestock loss experienced was 27 % (highest in Ranthambore). Our comparative study provides insights into factors associated with conflict loss and tolerance, and aids in improving ongoing conservation and compensation efforts.

Keywords Crop raiding · India · Livestock predation · Mitigation · Park · People · Tolerance · Wildlife

K. K. Karanth (✉) · R. DeFries
Ecology, Evolution, and Environmental Biology, Columbia University, New York, NY 10027, USA
e-mail: krithi.karanth@gmail.com

R. DeFries
e-mail: rd2402@columbia.edu

K. K. Karanth · A. M. Gopalaswamy
Centre for Wildlife Studies, 224, Garden Apartments, Vittal Malya Road, Bengaluru 560001, India
e-mail: arjungswamy@gmail.com

K. K. Karanth
Wildlife Conservation Society, New York, NY 10460, USA

L. Naughton-Treves
Department of Geography, University of Wisconsin, Madison, WI 53706, USA
e-mail: lnaughto@wisc.edu

A. M. Gopalaswamy
Wildlife Conservation Research Unit (WildCRU), The Recanati-Kaplan Centre, Department of Zoology, University of Oxford, Tubney House, Abingdon Road, Tubney, Abingdon OX13 5QL, UK

Introduction

Crop damage and livestock predation by wildlife around protected areas (PAs) are a major conservation problem in many regions, and particularly in India. Crop and livestock losses, and the threat of human injury or death, may increase hostility of the affected people toward wildlife (Treves and Karanth 2003; Sillero-Zuberi et al. 2007). Local resentment is often intensified by conservation regulations that impede local citizens' capacity to cope with losses to wildlife (Dickman 2010). Such resentment may

Published online: 13 September 2013

lead to retaliation against species (Gadd 2005; Treves and Naughton-Treves 2005; Linkie et al. 2007; Treves 2009). Due to high densities of people living around PAs in Asia, human–wildlife conflicts can affect many people (Nyhus and Tilson 2004; Fernando et al. 2005; Johnson et al. 2006; Spiteri and Nepal 2008). Ongoing land use change and habitat fragmentation in and around PAs are associated with more conflict incidents as many species (particularly wide-ranging large mammals such as elephants, leopards, wolves, and tigers) move well beyond boundaries of limited PAs (Sekhar 1998; Distefano 2005, Karanth et al. 2012; Athreya et al. 2013; Suryavanshi et al. 2013; Karanth et al. 2013). Developing appropriate strategies first requires assessing conflict levels and patterns, as well as factors associated with tolerance for wildlife (Karanth and Madhusudan 2002).

Despite facing loss and damage to property and lives, cultural and religious practices in India have been associated with high tolerance, meaning low rates of lethal defense strategies (Mishra et al. 2003; Rangarajan 2001; Ogra and Badola 2008; Dickman 2010). Such tolerance for species has played a role in the persistence of some large mammals in some parts of India (Karanth et al. 2010). Evidence of tolerance includes reverence, and consequently protection for monkeys; some ungulates such as *Boselaphus tragocamelus* (Nilgai) and *Gazella bennettii* (Chinkara, Rangarajan 2001); and *Uncia uncia* (snow leopards, Mishra et al. 2003) despite crop and livestock losses. Tolerance is not universal, and in some instances, people kill *Elephas maximus* (elephants) and ungulates (Karanth and Madhusudan 2002; Bagchi et al. 2004), *Canis lupus* (Tibetan wolf, Mishra et al. 2003), *Panthera tigris* (tigers), and *Panthera pardus* (leopards, Treves and Karanth 2003). Nonetheless, compared with other Asian countries, tolerance for wildlife in India is generally high (Athreya et al. Athreya et al. 2010, 2013; Karanth et al. 2012, 2013; Karanth and Nepal 2012). Wildlife survival outside park boundaries may be encouraging to biologists concerned about species isolation, particularly in India where habitat has been greatly reduced. But, even though local communities may tolerate wildlife, they may suffer food insecurity, opportunity costs (e.g., time lost to guarding crops and livestock), and even "diminished psychosocial wellbeing" due to human–wildlife conflict (Barua et al. 2013).

In this article, we examined factors associated with human–wildlife conflict losses and attitudes toward wildlife of households living outside Ranthambore, Kanha, and Nagarahole National Parks, three important PAs in northern, central, and southern India. Indian PAs are surrounded by high densities of people who commonly graze livestock and collect forest resources from inside the PA (Madhusudan 2003; DeFries et al. 2010; Karanth and DeFries 2010; Karanth et al. 2013). Most PAs are not fenced, and people, livestock, and wildlife move across PA boundaries. Attempts by park management to impose restrictions have been met with resistance from local people (Karanth and Nepal 2012). Although in India PAs classified as National Parks do not legally permit any human resource collection, grazing, and other related activities, the laws are difficult to enforce, and illicit resource use within parks is widespread.

Here we define "conflict" as reported incidents of crop raiding and livestock predation, and "loss" as the likelihood an individual household reported crop raiding or livestock predation in the last year (Hill 2004; Sitati et al. 2005; Wang and Macdonald 2006; Spiteri and Nepal 2008). We define "tolerance" as a household opting not to kill wildlife under hypothetical scenarios of crop damage, injury, or killing of livestock or people. By comparing households in three PAs that differ ecologically, socioeconomically, and culturally, we identify general and PA-specific factors that influence conflict loss, tolerance for these incidents, and effectiveness of mitigation measures.

Ecologically, the three PAs differ in vegetation characteristics (dry scrub in Ranthambore to moist deciduous forests in Nagarahole), size (Ranthambore is half the size of Nagarahole, and one third the size of Kanha), rainfall (in Ranthambore it is significantly lower than in Kanha and Nagarahole), and the zones of interaction with the surrounding landscape (DeFries et al. 2010). Human population densities range from 67 to 242 people/km^2. Most people cultivate crops and/or raise livestock (>60 %), but they differ in land ownership and caste diversity (Karanth and Nepal 2012). Known cultural and religious practices lead us to expect differences in tolerance across PAs with the lowest tolerance in southernmost Nagarahole due to a tradition of hunting and eating wildlife, and tolerance increasing northward toward Ranthambore where people are largely vegetarian and attach religious–cultural value to protecting some species (Karanth and Madhusudan 2002; Karanth et al. 2010; Chhangani et al. 2010; Velho et al. 2012).

In this study, we first compared reported patterns of conflict with key physical variables and land use practices identified in previous studies. Specifically, we examined variations between the three PAs at varying distances from each PA (proximity to the PA is associated with conflict frequency or intensity related to wildlife densities). We recorded household activities including where livestock are grazed, total number of crops planted per year, size of farms, and use of protection measures (e.g., fencing, night guards). Finally, we explored influence of indirect factors including form of land ownership and years of farming experience. We then examined the variations in people's tolerance and attitudes toward carnivores and herbivores.

We expected the most crop damage to wildlife in households that are close to PAs, have more abundant

crops and livestock, and use fewer protection measures to experience more wildlife damage (Treves 2009). Less frequent losses were expected for households having more years of farming experience, and larger landholdings. We expected people to be less tolerant to carnivores than herbivores based on fear and/or the higher cost of losing livestock than crops (Treves and Naughton-Treves 2005). Yet, we also anticipated patterns of tolerance to play out differently across the three PAs due to cultural or religious factors (Kleiven et al. 2004; Treves 2006).

Methods

Study Sites

Ranthambore, Kanha, and Nagarahole National Parks (Fig. 1, RNP, KNP, and NNP from here onward) are located in diverse ecological and socioeconomic settings, and harbor many threatened species such as tigers, leopards, *Cuon alpinus* (wild dogs), elephants, and *Bos gaurus* (gaur), Table 1, 2). All the three were designated National Parks in 1955 and as Project Tiger Reserves in the 1970s, designations intended to spur stricter protection and attract higher levels of funding and public interest (Karanth and DeFries 2010). All three PAs attract tourists and have experienced an explosion of tourism and tourist facilities (DeFries et al. 2010; Karanth and DeFries 2011).

All three PAs are typical of many Indian PAs, surrounded by high-density human settlements and anthropogenic land use particularly cultivation and pastoralism (DeFries et al. 2010; Karanth et al. 2013). These PAs are not fenced, and therefore, people and livestock move freely in and out of the PAs (after the study period, park authorities attempted to fence certain sections of Kanha, Karanth, pers. obs., 2009). Ranthambore is located in an arid landscape within a mosaic of scrub and agriculture fields with little tree cover. Degraded forest patches interspersed among agricultural fields surround Kanha. Nagarahole is surrounded by coffee plantations and irrigated agriculture, with households growing crops up to 9 months a year. All PAs have limited habitat connectivity, with Kanha perhaps the least isolated (DeFries et al. 2010). The majority (81 %) of local residents depend at least partly on the PAs for natural resources (wood, plant parts, and nontimber forest products), and many graze their livestock inside the PAs (Karanth and Nepal 2012). Typically, livestock are herded into the PAs by a few people who may take responsibility for managing livestock belonging to several families. The average herd size was four animals/households comprising mainly not only cattle but also buffalos, sheep, and goats. They are left to graze and herded back to the villages at night (Karanth, pers. obs., 2009, 2011). Herding and grazing occur during the day (Karanth, pers obs., 2009, 2011; Karanth and Nepal 2012).

The compensation mechanisms in place are similar across PAs with individuals reporting losses by providing evidence to forest or revenue officials of the area where their household is located. The compensation is provided once damage (either area, numbers, and type of crops destroyed, or number and breed of livestock killed) is assessed by officials. Essentially, there are no differences in compensation amounts and policies for losses inside the PA versus outside. However, Karanth et al. (2012) found higher compensation in Kanha for households located within the legal buffer versus outside.

Social Surveys and Interviews

Structured and open-ended interviews were conducted with 398 households spread between the three PAs, and opportunistically selected from 178 villages surrounding the three PAs (Karanth et al. 2012). When we approached respondents, we explained that the purpose of the study was to understand conflict loss and compensation experienced by local people living around these PAs. Most respondents were adult men (>85 %) as women were either unavailable or reluctant to be interviewed. We visited villages at various distances around each PA (ranging from 0.01 to 10.50 km), choosing 10 km as an upper estimate within which most conflicts occur (Karnataka, Madhya Pradesh, and Rajasthan Forest Department 2009; Shukla, pers comm., 2009). Survey questions asked respondents about household's demographic characteristics, resource use, attitudes toward wildlife, and mitigation measures (Karanth and Nepal 2012). Respondents were specifically asked to recall conflict incidents experienced by their households during the past year (2008–2009), and provide information on compensation received. Local field assistants conducted surveys during June–August 2009 in the local languages Hindi (RNP and KNP) and Kannada (NNP), and responses were translated to English. Each household's geographic location was mapped and distance to PA was calculated.

Modeling Household Risk to Crop-raiding, Livestock Predation, and Tolerance for Wildlife

To model patterns of reported crop raiding and livestock predation, we fit a priori logit models (as responses were in binary form). We identified factors associated with reported levels of crop raiding and livestock predation using model selection. Here, the focus is on assessing the model fit from an a priori set of candidate models, allowing us to identify the variables significantly associated with conflict



Fig. 1 Location of Ranthambore, Kanha, and Nagarahole National Parks in India (selected parks in black and other parks are gray)

loss based on the best fit using corrected Akaike's information criteria (AICc) as the model selection tool (Burnham and Anderson 2002; Karanth et al. 2012).

Accordingly, we fit separate models for crop loss and livestock loss, representing a different hypothesis about important factors potentially influencing a household's loss. We arrived at a candidate model set based on a priori hypotheses about the influence of individual factors (See Table 8 in Appendix). The models were ranked based on AICc values (Burnham and Anderson 2002). AICc weights (Burnham and Anderson 2002) calculated for each model represent the relative measure of appropriateness of a given model relative to the model set. In such multimodel selection problems, Burnham and Anderson (2002) recommend benchmarking against a "global model" that includes all potentially important variables of interest. The top model (based on AIC) is expected to be a model that includes a subset of these variables, which has the most

Table 1 Details on Ranthambore, Kanha, and Nagarahole National Parks in India

Characteristic	Ranthambore	Kanha	Nagarahole
Size	392 km^2	940 km^2	644 km^2
Location	25°54′–26°12′ N	22°7′–22°27′ N	11°5′–12°15′ N
	76°23′–76°39′ E	80°26′–81°3′ E	76°0′–76°15′ E
State	Rajasthan	Madhya Pradesh	Karnataka
Established	1955	1955	1955
Tourists/year	159,110 (2007–2008)	132,601 (2007–2008)	67,841 (2007–2008)
Vegetation	Dry scrub and deciduous forests	Sal and mixed deciduous forest	Dry and moist deciduous forests, teak plantations
Rainfall	800 mm	1371 mm	1500 mm
	(June–September)	(June–September)	(June–September)
Key species	Tiger, Leopard, Blackbuck, Sambar	Tiger, Leopard, Dhole, Barasingha, Sambar, Chital	Tiger, Leopard, Dhole, Elephant, Gaur, Sambar, Chital
Human population density within 10 km	128/km^2	91/km^2	681/km^2

Information is from DeFries et al. 2010, Karanth and DeFries 2011. KNP area excludes buffer

Table 2 Characteristics of surveyed households around Ranthambore, Kanha, and Nagarahole Parks

Characteristics	Ranthambore	Kanha	Nagarahole
Interviewed	98	117	186
Villages	51	49	78
Male	87 %	93 %	86 %
Age (in years)	38	43	40
Average household size (# members)	9	6	5
Born locally	92 %	91 %	80 %
Hindu	96 %	99 %	98 %
Caste groups	24	13	21
Households owning land	91 %	77 %	84 %
Agricultural crops as primary income	75 %	88 %	61 %
Average land size (acres)	6.1	4.6	4.4
Years of farming	35	40	21
Average distance to field (km) from house	1.1	1.9	1.0
Share crop (yes)	71 %	28 %	11 %
Average number of crops grown	12	13	23
Percentage of households owning livestock	32 %	27 %	27 %
Distance to park (km)	2.9 (0.0–8.1)	3.5 (0.0–9.6)	2.2 (0.0–1.4)
Crop raiding in the last year (yes)	74 %	86 %	79 %
Top crop raiders	Pig, Nilgai, Sambar	Pig, Chital	Elephant, Pig
Livestock predation in the last year (yes)	18 %	8 %	13 %
Animals lost (average)	1–3	1–5	1–5
Top predators blamed for livestock losses	Tiger, Leopard	Tiger, Leopard	Tiger, Leopard

optimal trade-off between model fit and parsimony (Burnham and Anderson 2002).

In developing our candidate model set, we removed obviously correlated variables (number of crops grown versus number of subsistence or commercial crop types) and tested for other correlations among variables by computing Pearsons' correlation coefficients. High correlation (>0.5) resulted in removal of three correlated variables (household practicing cultivation, land ownership, and subsistence versus commercial cultivation). We scaled the regression inputs by dividing by two standard deviations, as suggested by (Gelman 2008), to permit direct comparison of estimated coefficient values. We used the AICc to select the best models to assess model fit (Karanth et al.

2012, 2013). For closely competing models ($\Delta AIC < 2$), we applied model-averaging procedures for parameter estimation, while simultaneously accounting for a large part of multicollinearity (Freckelton 2011). Chi-squared tests of significance were used to test for differences among households across PAs.

We modeled household crop loss as a function of environmental factors (distance to PA edge and PA itself), household resource practices (use of fuel wood, water, and grazing inside the PA), socioeconomic and agricultural factors (number of harvests, number of crops and cropping months, farming experience, land area etc.), and mitigation factors (lighting, fencing, guard animals etc., See Table 8 in Appendix). Similarly, we modeled household risk to livestock predation as a function of multiple factors (See Table 8 in Appendix). For the global "crop loss" model we selected 23 variables and constructed 20 models. For the global "livestock loss" model, we selected 26 variables and constructed 21 models for livestock loss representing our hypotheses. We calculated the weighted average of these individual model estimates to generate overall estimates for probabilities of crop and livestock loss for individual households and around each PA.

To examine and compare tolerance and attitudes toward carnivores and herbivores, we presented households with hypothetical conflict situations and response scenarios, and asked them to select their preferred scenario based on Manfredo et al. (1998). The conflict situations refer to an animal coming outside the PA, damaging crop or property, injuring or killing livestock, and injuring or killing humans. Response scenarios include taking no action, capturing, and relocating animal, frightening or deterring animal, and killing the animal.

Results

Crop Raiding, Compensation, and Modeling Household Loss

About 80 % of households across all PAs reported crop damage in the previous year, ranging from 74 % in RNP to 86 % in KNP ($n = 398$). Wildlife named by respondents included eight herbivores, eight carnivores, monkeys (macaques and langurs), and peacocks (Fig. 2; see Table 2 for details). Wild pigs were reported to be the most frequent raiders across all three PAs. In all PAs, 96 % percent of households reported no injury or death due to confrontations with wildlife and 99 % of households did not report killing problem animals.

In RNP and KNP households experiencing crop damage was reported highest during the months of September to December. In NNP crop damage was experienced year round, but highest from May to August. In terms of intensity, more NNP households reported losing their entire crop to raiding by wildlife (partly attributable to elephants only found in NNP). Average reported per household annual loss of income from crop loss was US $155 in KNP, US $264 in RNP and US $546 in NNP (1 US$ = 52 Indian Rupees, these losses are significant particularly in areas where annual household incomes averaged for households we surveyed $300), but these estimates have to be verified with official records. Most households (89 %) reported that they did not receive any compensation from authorities for crop losses, but reporting efficacy varies by species. In another study, in the Western Ghats, households were more likely to report elephants compared to other species (Karanth et al. 2013). Overall, the majority of respondents

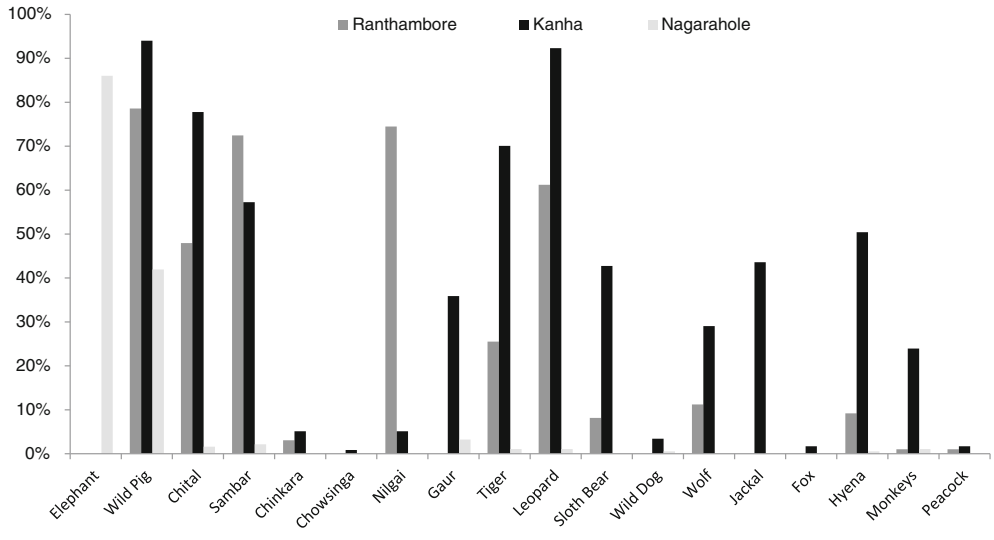


Fig. 2 Reported wildlife species in conflict with households in Ranthambore, Kanha, and Nagarahole National Parks in India

Table 3 Best models and beta coefficients for predicting crop loss around Ranthambore, Kanha, and Nagarahole National Parks

	Crop loss				
Models	8	4	16	6	3
	$w_i = 0.64$	$w_i = 0.18$	$w_i = 0.07$	$w_i = 0.05$	$w_i = 0.03$
Intercept	8.20	6.73	6.69	6.28	6.76
	(1.71)	(1.10)	(1.13)	(1.02)	(1.12)
Distance to park	_	_	_	_	−0.01
					(0.75)
Number of harvests	2.29	1.88	1.91	1.42	1.87
	(1.25)	(1.01)	(1.03)	(0.93)	(1.01)
Land area	_	_	_	_	0.58
					(1.16)
Share crop	2.37	2.36	2.34	_	2.31
	(1.39)	(1.28)	(1.29)		(1.26)
Number of crops per year	4.90	3.27	3.29	3.02	3.21
	(1.53)	(1.07)	(1.07)	(0.99)	(1.10)
Cropping months per year	15.23	13.02	12.90	12.77	13.07
	(3.33)	(2.11)	(2.28)	(2.00)	(2.13)
Fencing	−1.60	_	_	_	_
	(0.97)				
Night watching	0.55	_	_	_	_
	(0.92)				
Guard animals	−1.98	_	_	_	_
	(1.79)				
Lighting	3.26	_	_	_	_
	(1.44)				
Any of the mitigation measure	_	_	0.12	_	_
			(0.84)		
Model AICc	66.64	69.18	71.22	71.68	73.07
ΔAICc	0	2.54	4.59	5.04	6.42

said they had reported crop damage to authorities (80 %), and those receiving compensation were paid US $96 on average (range $45–454).

In modeling factors associated with household risk to crop raiding, we found three top-ranked models with substantial weights (Table 3). As predicted, factors related to forage availability for wildlife- the number of harvests per year, share cropping (growing crops on someone else's land), number of crops per year and cropping months in a year all increased risk (positive β coefficients in Table 3). Although four mitigation factors appeared in the top-ranked model only two (fencing and guard animals) appeared to decrease risk (negative β coefficients in Table 3). We suspect the positive correlation to night watching and lighting is due to these mitigation measures being in place as result of households experiencing past incidents of conflict. For crop loss, the strongest predictor variables were the number of cropping months, number of crops and number of harvests (Table 3). The number of cropping months was five times as influential as variety of crops, and seven times as influential as number of harvests and share cropping (Table 3). Importantly, we found that the individual PA and distance to PA area were not significant in all three PAs. This may suggest crop raiders may not be limited to animals coming out from the PA, and perhaps some raiders may naturally reside outside the PAs.

We estimated probabilities of crop loss for each household and compared across these PAs. We find that average estimated crop loss across all households within 10 km of these reserves was 0.76 in Ranthambore to 0.86 in Kanha, and these estimated probabilities are similar to reported losses (Table 5).

Livestock Predation, Compensation, and Modeling Household Loss

Livestock ownership did not vary across the PAs, with most common animals being cattle, sheep and goats (Table 2). Livestock losses were reported by 13 % of households, ranging from 8 % in RNP to 18 % in KNP. Animals blamed for livestock loss included tiger, leopard, *Hyaena hyaena* (Hyena), dhole (*Cuon alpinus*), *Canis*

Table 4 Best models and beta coefficients for predicting livestock loss around Ranthambore, Kanha, and Nagarahole National Parks

Models	Livestock loss					
	13	9	10	2	4	16
	$w_i = 0.60$	$w_i = 0.23$	$w_i = 0.07$	$w_i = 0.02$	$w_i = 0.02$	$w_i = 0.02$
Intercept	−2.17	−2.16	−2.18	−2.05	−2.06	−2.04
	(0.18)	(0.18)	(0.18)	(0.16)	(0.17)	(0.16)
Distance to park	_	0.12	0.15	0.24	_	_
		(0.33)	(0.33)	(0.31)		
Park-Nagarahole	0.85	0.82	0.71	_	_	_
	(0.96)	(0.97)	(0.97)			
Park-Ranthambore	1.76	1.76	1.82	_	_	_
	(0.90)	(0.90)	(0.92)			
Park-Kanha	0.32	0.35	0.36	_	_	_
	(0.85)	(0.85)	(0.85)			
Wood collected from park	_	_	0.03	_	−0.42	_
			(0.37)		(0.34)	
Water collected from park	_	_	0.81	_	0.68	_
			(0.61)		(0.57)	
Graze livestock in park	1.85	1.84	1.91	1.50	1.64	1.54
	(0.40)	(0.40)	(0.41)	(0.36)	(0.37)	(0.36)
Any mitigation measure	_	_	_	_	_	0.04
						(0.39)
Model AICc	283.04	284.99	287.31	289.70	290.08	290.29
ΔAICc	0	1.94	4.27	6.65	7.04	7.24

aureus (jackal), Mongoose species, *Viverricula indica (civet)*, *Sus scrofa* (wild pig), and *Macaca mulata* and *radiata* (macaques). Leopards were identified as the most damaging species in all PAs. The majority of households (99 %) reported not receiving any compensation from authorities for losses. The compensation mechanisms in place are similar across PAs with individuals reporting losses by providing evidence to forest or revenue officials. Very few of the households reporting losses (less than 1 %) reported about having received any compensation (average $ 87/household for households receiving compensation).

Six top-ranked models of factors associated with reported household livestock loss (Table 4) indicate that as predicted, factors such as grazing, collecting water from the PA, and the individual PA increased household risk (positive β coefficients in Table 4). Although distance to PA and collecting water and wood from the PA appeared in the top models, these factors were insignificant as evidenced by the high standard errors of the β coefficients. Other socioeconomic, demographic, and mitigation factors were not significant. The strongest predictor variable in all models was grazing domestic animals inside the PAs (Table 4). Households in RNP are apparently twice as vulnerable to livestock predation compared with KNP, and five times as vulnerable compared with NNP households (Table 4). This is because RNP households perhaps, have fewer choices of grazing areas outside the PA compared with KNP and NNP. Our results indicate that households' decision to enter the PA to graze and collect forest resources varies across the PAs.

Table 5 Estimated probabilities of crop and livestock losses for 398 households surveyed around three parks in India

Type		RNP	KNP	NNP
Crop loss	Average	0.76	0.86	0.81
	Range	0.0002–1	0.0001–1	0.0002–1
Livestock loss	Average	0.38	0.16	0.27
	Range	0.35–0.79	0.13–0.49	0.19–0.65

We modeled estimated probabilities of livestock loss for each household and compared across PAs. We find average estimated loss across all households within 10 km of these reserves ranged between 0.16 in Kanha to 0.38 in Ranthambore, which are higher than directly reported losses (Table 5).

Resident Tolerance of Crop Raiders and Livestock Predators

Households were presented with hypothetical scenarios about carnivores' and herbivores' behaviors and asked to choose among a range of actions based on Manfredo et al.

Table 6 Comparison of respondents' attitudes toward hypothetical scenarios for carnivores and herbivores, and potential action to be taken (1–4) in Ranthambore, Kanha, and Nagarahole National Parks

Hypothetical Scenario	Carnivore				Herbivore			
	RNP	KNP	NNP	Chi squared Test	RNP	KNP	NNP	Chi Squared Test
Comes outside park	3 (57 %)	1 (71 %)	1 (41 %)	$\chi^2 = 88.92$	3 (70 %)	1 (95 %)	3 (40 %)	$\chi^2 = 174.75$
Damages crops or property	3 (66 %)	2 (34 %)	3 (65 %)	$\chi^2 = 70.57$	3 (71 %)	2 (40 %)	4 (35 %)	$\chi^2 = 75.84$
Injures or kills domestic animals	3 (44 %)	3 (37 %)	3 (57 %)	$\chi^2 = 65.82$	3 (47 %)	3 (35 %)	1 (36 %)	$\chi^2 = 120.51$
Injures or kills humans	3 (55 %)	4 (85 %)	2 (40 %)	$\chi^2 = 40.04$	3 (61 %)	4 (92 %)	3 (34 %)	$\chi^2 = 152.69$

The most preferred option is stated under the park name and differences among parks tested using Chi squared tests ($P < 0.0001$ in all cases)
1 Take no action and monitor situation, *2* Capture and relocate animal, *3* Frighten or deter animal, *4* Kill the animal

(1998). The behaviors include an animal coming outside the PA, damaging crop or property, injuring or killing livestock, and injuring or killing humans. Potential action choices provided include taking no action, capturing and relocating animal, frightening or deterring animal, and killing the animal. Contrary to our expectations, carnivores were not viewed with more hostility than herbivores. Instead, respondents reported greater inclination to kill an herbivore (Table 6). Residents of KNP were also significantly more likely to favor killing wildlife, unexpected—given that KNP has the lowest reported livestock losses to wildlife (8 %, Table 2). Across the three PAs, we found that RNP households preferred to frighten or deter both carnivores and herbivores irrespective of the situation and conflict experience (Table 6) despite the largest number of households (18 %) experiencing livestock predation (Table 2). For KNP households, we found the expected pattern of decreasing tolerance as the severity of the problem increased, with injury or loss of human life being the least acceptable (Table 6). In NNP, households preferred to frighten or deter animals or take no action.

Mitigation Measures

Households were questioned about practices followed to protect their property, domestic animals, and lives. The majority (83 %) of households reported taking measures, most commonly adding or improving night watching (66 %), fencing (61 %), enclosed structures (51 %), and lighting (49 %, Table 7). The only protection mechanism that did not vary across PAs was fencing, which was lower for all three PAs possibly because of high costs of building and maintenance (Table 7). Our analysis suggests that across all PAs, despite the use of many mitigation measures, only fencing and guarding animals appear to be effective in lowering losses.

Discussion

Logistic models indicate unsurprisingly that crop loss is associated with factors related to forage availability (a greater diversity of crops available for longer periods increase the risk of crop raiding). Importantly, the percentage of household reporting crop loss did not vary markedly across PAs (Table 5). We find that regardless of ecological or geographic attributes households universally perceive high losses. Our models suggest that only fencing and guard animals lowered crop loss similar to Sitati et al. (2005) and Linkie et al. (2007), and other mitigation factors had little effect. The fact that distance to PA showed no significant association with reported conflict across a 10-km survey region suggests that crop raiders may not be coming entirely from the PAs alone and may be residing outside park boundaries. Although respondents tended to blame parks as "sources", at least some wildlife might be foraging entirely outside the PAs. Therefore, compensating people for losses becomes particularly challenging in areas where wildlife naturally occurs outside PAs, and losses may not be attributable to wildlife from the PAs (Naughton et al. 2003; Bulte and Rondeau 2005). Other studies from India (Athreya et al. 2013; Banerjee et al. 2013; Karanth et al. 2012, 2013; Suryavanshi et al. 2013) report similar results making it imperative for the government to enact policies that are effective and efficient to ensure wildlife in India are able to persist outside PAs as well.

In contrast, livestock loss was more sensitive to individual practices, such as grazing animals inside the PA, which are often carried out in parallel with collecting fuel wood and water (Karanth and Nepal 2012, Karanth et al. 2012). Respondents reported that livestock losses occurred inside the PAs during the day, and fewer incidents occurred at night when the livestock were corralled near their homes or villages. PA differences also emerged with RNP households reporting

Table 7 Comparison of protection mechanisms and mitigation measures used by households around Ranthambore, Kanha, and Nagarahole National Parks

Protection/mitigation measures	RNP (%)	KNP (%)	NNP (%)	All parks (%)	Chi squared test
Added or improved fencing	69	58	59	61	$\chi^2 = 3.45$, $P = 0.17$
Added or improved night watching	74	87	47	66	$\chi^2 = 55.42$, $P < 0.0001$
Added or improved physical structures	70	86	19	51	$\chi^2 = 150.28$, $P < 0.0001$
Added or improved use of guard animals (e.g., dogs)	36	26	6	19	$\chi^2 = 40.50$, $P < 0.0001$
Added or improved lighting	63	69	28	49	$\chi^2 = 58.25$, $P < 0.0001$
Added or improved scare devices	54	34	40	42	$\chi^2 = 9.19$, $P < 0.01$
Added or improved removal of waste or dead animals	40	66	1	30	$\chi^2 = 153.21$, $P < 0.0001$
Added or improved removal of brush or forest	24	6	2	8	$\chi^2 = 44.12$, $P < 0.0001$
Changed breeds or types of animals owned or managed	15	14	1	8	$\chi^2 = 25.69$, $P < 0.0001$
Started allowing hunters on land	4	5	1	3	$\chi^2 = 6.53$, $P = 0.03$
Started keeping a closer eye on animals	70	87	2	43	$\chi^2 = 252.50$, $P < 0.0001$
Stopped or reduced allowing hunters on land	65	82	1	40	$\chi^2 = 232.63$, $P < 0.0001$
Stopped or reduced feeding other wild animals	67	85	1	41	$\chi^2 = 245.17$, $P < 0.0001$
Stopped or reduced use of public lands with wildlife	26	6	1	8	$\chi^2 = 53.76$, $P < 0.0001$

the highest losses. Livestock losses could be reduced if grazing in PAs were avoided, but this simple finding is realistically hard to manage or enforce around Indian PAs. This is particularly so in RNP and NNP where grazing areas outside are limited compared with KNP, and there is some evidence to suggest that households are increasing stall-feeding (Karanth et al. 2012). In the absence of alternate grazing areas, households in closer proximity to the PAs are likely to graze and collect resources inside (particularly in RNP). We also find that mitigation factors other than fencing and guarding animals did not decrease livestock loss, a finding similar to that in Kolowski and Holekamp (2006). These findings are important particularly when increasing conservation attention and monies are being invested in mitigation practices worldwide, and highlight the need to carefully identify mitigation practices that actually work. These results mirror findings from parallel studies in Central and Southern India (Karanth et al. 2012, 2013), suggesting widespread policy changes might be needed in the system.

The near absence of compensation reported by respondents' accords with other Asian countries, where compensation is low or nonexistent, and losses reported were high relative to annual household income. This potentially presents a challenge to maintaining tolerance for wildlife (Wang and Macdonald 2006; Spiteri and Nepal 2008; Karanth and Nepal 2012). However, there is an urgent need to corroborate losses reported by local residents to official records to identify gaps and improve compensation distribution to households reporting losses. India's current compensation policy is oriented more toward alleviating economic hardship (versus discouraging retaliation against endangered species) (Agarwala et al. 2010; Karanth et al. 2012, 2013). At one level, our study suggests that this policy is on target given that so few respondents reported a wish to kill wildlife raiding crops or hunting livestock. However, given that some respondents may have been reluctant to express this action and in light of the large area and great number of people reporting complaints, any initiative to expand compensation payments will have to be carefully designed and communicated. In developed and developing countries alike, offers of compensation have been associated with higher level of complaint (Naughton et al. 2003; Bulte and Rondeau 2005). Compensation schemes in place in these parks and those elsewhere in India (Karanth et al. 2012, 2013) might have to be revisited and restructured to be made more efficient and targeted toward species of focus.

Four percent of the surveyed households reported that an individual was injured or killed in confrontations with wildlife, and 99 % of households reported they did not kill problem animals. Tolerance for carnivores does not vary significantly across PAs, whereas it does for herbivores, namely KNP households appeared more likely to favor killing problem species of herbivores compared with RNP and NNP. Presented with hypothetical scenarios, households chose to take lethal action against herbivores destroying crops and carnivores harming people. For most scenarios, frightening and deterring animals was the most popular choice, although some respondents might have been reluctant to report their approval for killing animals in light of conservation rules. Our findings are in contrast to other studies that have found respondents more inclined to support lethal control against "problem" carnivores (Knight 2000; Ericsson and Heberlein 2003; Kleiven et al. 2004; Kaltenborn et al. 2006; Holmern et al. 2007).

Our results shed light on the similarities and differences across PAs in India with regard to human–wildlife conflict incidents, and experiences with compensation and mitigation. High densities of people live in close proximity to wildlife, and

historically have tolerated some species (Karanth et al. 2010). We find similar crop loss patterns across all PAs attributable to forage availability due to agricultural characteristics, and did not show clear evidence of wildlife coming only from the PAs. In contrast, livestock predation varied markedly by PA, individual household practices and management strategies. Our findings suggest that improving mitigation factors such as fencing and guarding might reduce crop loss, and simply grazing animals outside the PAs might reduce livestock loss. These results are common across all three PAs, although identifying specific mitigation strategies needs to be based on the individual species and practices in each PA. Improving crop protection is perhaps simpler in contrast to enforcing grazing regulations, where the availability of free forage will continue to attract grazing inside PAs despite the high risk of livestock loss.

Potential limitations of our study include possible bias in reported losses, and largely male respondents being accessible to us. Although we have no reason to think there was systematic bias across all three PAs. Our results are corroborated by studies in central India (Karanth et al. 2012) and southern India (Karanth et al. 2013). Our survey radius of 10 km and the number of households exceed those for most field-driven human–wildlife conflict surveys. Yet, our respondents' reports suggest that we could have sampled further from the PAs. These findings have to be corroborated by linking official compensation records to those reported by households, and followed by systematic monitoring of conflict and compensation processes as they occur. We find that despite significant losses and poor compensation, some tolerance for losses to wildlife in India still exists, although it varies by PA and species.

Acknowledgments The authors thank K. U. Karanth, P. M. Muthanna, S. K. Nepal, Centre for Wildlife Studies and Wildlife Conservation Society-India Program for support in the field. They also thank Jim Nichols and USGS Patuxent Wildlife Research Centre for hosting Karanth and Gopalaswamy in 2011. We are grateful to the forest departments of Rajasthan, Madhya Pradesh, and Karnataka, and S. Sharma, H.S. Pabla, H. S. Negi, R. Shukla, A.C. Poovaih, and U. Ramakrishnan for their support. Invaluable field assistance by A. Srivathsa and S. Gundecha enabled data collection. NSF Grant Number 1029219 supported this research. Columbia University IRB approved all the survey protocols.

Appendix

See Table 8.

Table 8 Covariates and a priori predictions about their influence on household risk to crop-raiding and livestock predation in Ranthambore, Kanha, and Nagarahole National Parks

Covariate category	Variable	Crop-raiding	Livestock predation
Park	a. Park (pa)	NNP > KNP > RNP	RNP > KNP > NNP
	b. Distance to park (dpa)	Decreases	Decreases
Park resources	a. Wood from park (wdpa)	Increases	Increases
	b. Water from park (wtrpa)	Increases	Increases
	c. Graze in park (grpa)	Increases	Increases
Agriculture	a. Number of crops (cp)	Increases	NA
	b. Number of harvests (harv)	Increases	NA
	c. Cropping months (crmn)	Increases	Increases
	d. Share crop (scrop)	Increases	Increases
Household	a. Land area (larea)	Increases	NA
	b. Years farming (yrfm)	Decreases	Decreases
	c. Drought experience (dht)	Increases	Increases
	d. Total people in house (hht)	Decreases	Decreases
Mitigation factors	a. Fencing (ma)	Decreases	Decreases
	b. Night watching (mb)	Decreases	Decreases
	c. Birthing structure (mc)	NA	Decreases
	d. Guard animals (md)	Decreases	Decreases
	e. Lighting (me)	Decreases	Decreases
	f. Removal of waste or dead animals (mf)	Decreases	Decreases
	g. Removal of brush (mg)	Decreases	Decreases
	h. Changing breeds (mh)	NA	Decreases
	i. Allowing hunting (mi)	Decreases	Decreases
	j. Closer eye on animals (mj)	Decreases	Decreases
	k. Stopping hunting (mk)	Decreases	Decreases
	l. Stopping feeding wildlife (ml)	Decreases	Decreases
	m. Stopping use of public land (mm)	NA	Decreases

References

Agarwala M, Kumar S, Treves A, Naughton-Treves L (2010) Paying for wolves in Solapur, India and Wisconsin, USA: comparing compensation rules and practice to understand goals and politics of wolf conservation. Biol Conserv 143:2945–2955

Athreya V, Odden M, Linnell JDC, Karanth KU (2010) Translocation as a tool for mitigating conflicts with leopards in human-dominated landscapes of India. Conserv Biol 25:133–141

Athreya V, Odden M, Linnell JDC, Krishnaswamy J, Karanth KU (2013) Big cats in our backyards: persistence of large carnivores in a human dominated landscape in India. PLoS One 8(3):e57872. doi:10.1371/journal.pone.0057872

Bagchi S, Mishra C, Bhatnagar YV (2004) Conflicts between traditional pastoralism and conservation of Himalayan Ibex (Capra sibirica) in the Trans-Himalayan mountains. Anim Conserv 7:121–128

Banerjee K, Jhala YV, Chauhan KS, Dave CV (2013) Living with Lions: the economics of coexistence in the Gir Forests India. Plos One 8(1):e49457

Barua M, Bhagwat SA, Jadhav S (2013) The hidden dimensions of human-wildlife conflict: health impacts, opportunity and transaction costs. Biol Conserv 157:309–316

Bulte EH, Rondeau D (2005) Research and management view point: why compensating wildlife damages may be bad for conservation. J. Wildlife Manag 69:14–19

Burnham KP, Anderson D (2002) Model Selection and Multimodal Inference. A practical information theoretic approach. Springer, New York

Chhangani AK, Robbins P, Mohnot SM (2010) Major threats to the faunal diversity of Kumbhalgarh Wildlife Sanctuary. Indian Forest 136:1410–1421

DeFries R, Karanth KK, Pareeth SJ (2010) Interactions between protected areas and their surroundings in human-dominated tropical landscapes. Biol Conserv 143:2870–2880

Dickman AJ (2010) Complexities of conflict: the importance of considering social factors for effectively resolving human-wildlife conflicts. Anim Conserv 13:458–466

Distefano, E., 2005. Human-wildlife conflict worldwide: collection of case studies, analysis of management practices and good practices. FAO Report. (www.fao.org)

Ericsson G, Heberlein TA (2003) Attitudes of hunters, locals, and the general public in Sweden now that the wolves are back. Biol Conserv 111:149–159

Fernando P, Wikramanayake E, Weerakoon D, Jayasinghe LKA, Gunawardene M, Janaka HK (2005) Perceptions and patterns of human-elephant conflict in old and new settlements in Sri Lanka. Int J Biodivers Conserv 14:2465–2481

Freckelton RP (2011) Dealing with collinearity in behavioural and ecological data: model averaging and the problems of measurement error. J Behav Ecol Soc 65:91–101

Gadd ME (2005) Conservation outside parks: attitudes of local people in Laikipia. Kenya Environ Conserv 30:50–63

Gelman A (2008) Scaling regression inputs by dividing by two standard deviations. Stat Med 27:2865–2873

Hill CM (2004) Farmers' perspectives of conflict at the wildlife–agriculture boundary: some lessons learned from African subsistence farmers. Hum Dim Wildl 9:279–286

Holmern T, Nyahongo J, Roskaft E (2007) Livestock loss caused by predators outside the Serengheti National Park. Tanzania Biol Conserv 135:518–526

Johnson A, Vongkhamheng C, Hedemark M, Saithongdam T (2006) Effects of human-carnivore conflict on tiger (*Panthera tigris*) and prey populations in Lao PDR. Anim Conserv 9:421–430

Kaltenborn BP, Bjerke T, Nyahongo J (2006) Living with problem animals: self-reported fear of potentially dangerous species in the Serengeti region. Tanzania Hum Dim Wildl 11:397–409

Karanth KK, DeFries R (2010) Conservation and management in human- dominated landscapes: case studies from India. Introduction Spec Issue Biol Conserv 143:2865–2869. doi:10.1016/j.biocon.2010.05.002

Karanth KK, DeFries R (2011) Nature-based tourism in Indian protected areas: new challenges for park management. Cons Letts 4:137–149

Karanth KK, Gopalaswamy AM, DeFries R, Ballal N (2012) Assessing Patterns of Human-Wildlife Conflicts and Compensation around a Central Indian Protected Area. PLoS One 7(12):e50433. doi:10.1371/journal.pone.0050433

Karanth KK, Gopalaswamy AM, Prasad P, Dasgupta S (2013) Patterns of human-wildlife conflicts and compensation: insights from Western Ghats Karnataka protected areas. Biol Cons 166:175–185

Karanth KU, Madhusudan MD (2002) Making parks work: identifying key factors to implementing parks in the tropics. In: Terborgh JW, van Schaik C, Davenport L, Rao M (eds) Mitigating human-wildlife conflicts in southern Asia. Island, Covelo, pp 250–264

Karanth KK, Nepal S (2012) Local perceptions of benefits and losses of living around protected areas in India and Nepal. Environ Manag 49:372–386

Karanth KK, Nichols JD, Karanth KU, Hines JE, Christensen NL (2010) The shrinking ark: patterns of large mammal extinctions in India. Proc Royal Soc Ldn-B 277:1971–1979

Kleiven J, Bjerke T, Kaltenborn BP (2004) Factors influencing the social acceptability of large carnivore behaviors. Int J Biodivers Conserv 13:1647–1658

Knight J (2000) Introduction. In: Knight J (ed) Natural enemies: people-wildlife conflicts in anthropological perspective. Routledge, London

Kolowski J, Holekamp K (2006) Spatial, temporal and physical characteristics of livestock depredations by large carnivores. Biol Conserv 128:529–541

Linkie M, Dinata Y, Nofrianto A, Leader-Williams N (2007) Patterns and perceptions of wildlife crop raiding in and around Kerinci Seblat National Park. Sumatra Anim Conserv 10:127–135

Madhusudan MD (2003) Living amidst large wildlife: livestock and crop depredation by large mammals in the interior villages of Bhadra Tiger Reserve, south India. Environ Manag 31:466–475

Manfredo MJ, Zinn HC, Sikorowski L, Jones J (1998) Public acceptance of mountain lion management: a case study of Denver, Colorado, and nearby foothills areas. Wildl Soc Bull 26:964–970

Mishra C, Allen P, Maccarthy T, Madhusudan MD, Bayarjargal A, Prins HHT (2003) The role of incentive programs in conserving the snow leopard. Conserv Biol 17:1512–1520

Naughton T, Grossberg R, Treves A (2003) Paying for tolerance: the impact of livestock predation and compensation payments on rural citizens attitudes towards wolves. Conserv Biol 17:1500–1511

Nyhus P, Tilson R (2004) Agroforestry, elephants and tigers: balancing conservation theory and practice in human-dominated landscapes of Southeast Asia. Agric Eco Environ 104:87–97

Ogra M, Badola R (2008) Compensating human-wildlife conflict in Protected Area Communities: ground level perspectives from Uttarkhand. India Hum Ecol 36:717–729

Rangarajan M (2001) India's wildlife history. Permanent Black, New Delhi

Sekhar N (1998) Crop and livestock depredation caused by wild animals in protected areas: the case of Sariska Tiger Reserve, Rajasthan. India. Environ Conserv 25:160–171

Sillero-Zuberi C, Sukumar R, Treves A (2007) Living with wildlife: the roots of conflict and the solutions. In: MacDonald DW, Service K (eds) Key topics in conservation biology. Oxford University, Oxford, pp 266–272

Sitati NW, Walpole MJ, Leader-Williams N (2005) Factors affecting susceptibility of farms to crop raiding by African elephants: using a predictive model to mitigate conflict. J Appl Ecol 42:1175–1182

Spiteri A, Nepal SK (2008) Evaluating local benefits from conservation in Nepal's Annapurna conservation area. Environ Manag 42:391–401

Suryavanshi KR, Bhatnagar YV, Redpath S, Mishra C (2013) People, predators and perceptions: patterns of livestock depredation by snow leopards and wolves. J Appl Ecol 50:550–560

Treves A (2006) The human dimensions of conflicts with wildlife around protected areas. In: Woodroof R, Thirgood S, Rabinowitz A (eds) People and wildlife, conflict or coexistence?. Cambridge University, Cambridge, pp 214–228

Treves A (2009) Hunting to conserve large carnivores. J Appl Ecol 46:1350–1356

Treves A, Karanth KU (2003) Human-carnivore conflict and perspectives on carnivore management worldwide. Conserv Biol 17:1491–1499

Treves A, Naughton-Treves L (2005) Evaluating lethal control in the management of human-wildlife conflict. In: Woodroof R, Thirgood S, Rabinowitz A (eds) People and wildlife, conflict or coexistence?. Cambridge University, Cambridge, pp 86–106

Velho N, Karanth KK, Laurance WF (2012) Hunting a serious and understudied threat to India, a globally significant conservation region. Biol Conserv 148:210–215

Wang SW, Macdonald DW (2006) Livestock predation by carnivores in Jigme Singye Wangchuck National Park. Bhutan Biol Conserv 129:558–565

Appendix O

Identifying species, sex and individual tigers and leopards in the Malenad-Mysore Tiger Landscape

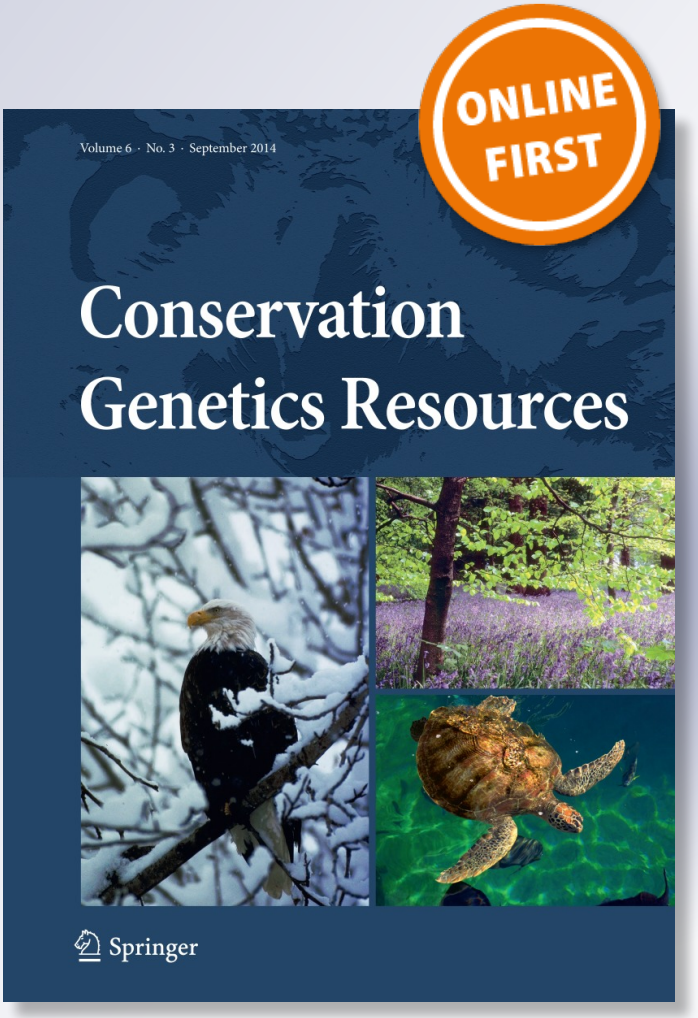
Identifying species, sex and individual tigers and leopards in the Malenad-Mysore Tiger Landscape, Western Ghats, India

Samrat Mondol, N. Samba Kumar, Arjun Gopalaswamy, Kartik Sunagar, K. Ullas Karanth & Uma Ramakrishnan

Conservation Genetics Resources

ISSN 1877-7252

Conservation Genet Resour
DOI 10.1007/s12686-014-0371-9



Springer

Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media Dordrecht. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com”.



Author's personal copy

Conservation Genet Resour
DOI 10.1007/s12686-014-0371-9

APPLICATION ESSAYS

Identifying species, sex and individual tigers and leopards in the Malenad-Mysore Tiger Landscape, Western Ghats, India

Samrat Mondol · N. Samba Kumar · Arjun Gopalaswamy · Kartik Sunagar · K. Ullas Karanth · Uma Ramakrishnan

Received: 14 June 2014 / Accepted: 24 October 2014
© Springer Science+Business Media Dordrecht 2014

Abstract Assessing the distribution, abundance and demographic ratios of endangered and elusive co-occurring carnivore species at a landscape level is important for their continued survival. Despite potential to determine distribution and dietary analyses, use of faecal samples has been relatively limited in the context of multiple sympatric species living at large landscapes. We developed and optimized a range of novel non-invasive molecular techniques for species, gender and individual identification of tiger (*Panthera tigris*) and leopard (*Panthera pardus*). We collected a large number of faecal samples as part of a pilot occupancy survey in the Malenad-Mysore Tiger Landscape, Western Ghats, India. We could genetically ascertain species and gender for 88 and 57 % of the field-collected samples respectively. Additionally, we also determined a panel of nine and eight polymorphic loci for tiger and leopard individual identification, resulting in 18 tigers and 39 leopards from varied quality field-collected samples. Our pilot study suggest such molecular approaches will help in future efforts to gather landscape level distribution, demographic and other ecological information on tigers and leopards across their distribution.

Keywords Species identification · Molecular sexing · *Panthera tigris* · *Panthera pardus* · Carnivore conservation

S. Mondol (✉)
Centre for Cellular and Molecular Platforms, GKVK Post, Bangalore 560 065, India
e-mail: samratmondol@gmail.com

S. Mondol · K. Sunagar · U. Ramakrishnan
National Centre for Biological Sciences, TIFR, GKVK Campus, Bellary Road, Bangalore 560 065, India

Present Address:
S. Mondol
INSPIRE Faculty, Wildlife Forensics and Conservation Genetics Department, Wildlife Institute of India, P.O. Box-18, Chandrabani, Dehradun 248 001, India

N. S. Kumar · A. Gopalaswamy · K. U. Karanth
Center for Wildlife Studies, Aga Abbas Ali Road, Bangalore 560 042, India

K. U. Karanth
Wildlife Conservation Society, Bronx, NY, USA

Introduction

Knowledge of species' occurrence and distribution patterns is critical in understanding ecology and prioritizing management decisions (Mittermeier et al. 1998). Acquiring this information can be particularly challenging for endangered mammals, as most are rare and elusive. Large carnivores are among the most endangered mammals in the world (Ceballos et al. 2005), and are conservation targets across their range (Weber and Rabinowitz 1996). Low densities, large home ranges and a secretive nature have always challenged studies to understand distribution patterns and other important biological questions for wide-ranging species (Schipper et al. 2008). In the current context of global climate change, changing land use practices, human persecution, habitat loss and fragmentation in biodiversity hotspots, accurate information about present distribution and abundance of large carnivores is critical for their future survival.

In recent time, direct and indirect noninvasive survey methods (for e.g. sign-track surveys, faecal encounters, remote cameras etc.) have been the primary choice to gather distributional information on elusive species (Palomares et al. 1991; Smallwood and Fitzhugh 1995; Foresman and Pearson 1998; Moruzzi et al. 2002; Karanth et al. 2009). While these approaches were successfully

Published online: 31 October 2014

Author's personal copy

employed in many carnivore studies, their utility has been particularly challenging in case of similar-body sized sympatric carnivores. Separating carnivores by sign-track surveys or faecal morphology (size, shape, scent and dietary contents) poses significant challenges in accurate species assignment and demographic studies (Zuercher et al. 2003; Prugh and Ritland 2005), and can particularly be error-prone in areas with co-occurring species (Davison et al. 2002; Fernandes et al. 2008). While remote cameras can easily distinguish species, they do not work for species in which individuals and sex are not uniquely identifiable. In this context, more accurate and reproducible DNA-based approaches have become a preferred alternative for accurate species identification (Beja-Pereira et al. 2009; Goossens and Bruford 2009), and are being increasingly used to study rare, threatened and cryptic populations. However, logistic and financial constraints have been an impediment in long-term studies of sympatric carnivores, resulting in limited studies over large spatial and temporal scales (Palomares et al. 2002; Bellemain et al. 2007; Bidlack et al. 2007, Wultsch et al. 2014).

In this paper, we describe a range of molecular tools to distinguish species, sex and individuals of tiger (*Panthera tigris*) and leopard (*Panthera pardus*). We further applied these approaches in a pilot genetic occupation/demographic distribution study for these species in the Malenad-Mysore Tiger Landscape (MMTL), Western Ghats, India. Both species exemplify wide-ranging, sympatric large carnivores, threatened due to recent massive habitat loss and global anthropogenic impacts (Uphyrkina et al. 2001; Walston et al. 2010). As most of the remaining tiger and leopard populations coexist in small, disjunct and fragmented habitats (Ranganathan et al. 2008) interspersed within human-dominated landscapes in the Indian subcontinent, reliable data on distribution and demography are vital in understanding their ecology and management. While considerable conservation investments are made to understand ecology and distribution for both species at local scales (Walston et al. 2010), but there is a paucity of data at metapopulation levels. We developed these genetic tools and tested in this pilot study with a long-term goal to study these large endangered carnivores over large spatial and temporal scale across their ranges.

Our objectives in this paper were, to (1) develop and validate a range of molecular tools for species, sex and individual identification from degraded biological samples of tigers and leopards and (2) test the described techniques on faecal samples collected under a pilot study conducted in MMTL landscape, Western Ghats. We believe that implementation of these newer molecular tools will be extremely valuable to assess distribution and demographic status of two of the largest endangered feline species throughout their range.

Materials and methods

Field test study area and sample collection

The Malenad-Mysore landscape (MMTL) in the state of Karnataka, India is the most productive habitat for large carnivores in the entire Western Ghats region of South-western India (Karanth et al. 2011). This landscape forms the south-central part of the Western Ghats, extending over 22,400 km^2 of moist-evergreen, moist-deciduous and the dry-deciduous forest types in Karnataka. The forest matrix includes 14 legally designated wildlife protected areas that cover over 5,500 km^2 of prime habitat. The rest of the landscape comprises of "multiple-use forests" that surround, abut or connect these protected areas providing a permeable matrix through which large carnivores can potentially disperse or perhaps even persist as non-breeding individuals (Fig. 1).

This study was part of a pilot survey to measure "habitat occupancy" of large carnivores over the entire MMTL landscape. Grid cells of about 188 km^2 (that corresponded to topographic maps of Survey of India in 1:25,000 scale) were superimposed on the entire MMTL study area. Only those grid cells that contained a minimum forest-patch size of 10 km^2 and more than 10 % of overall forest cover were sampled (for details see Karanth et al. 2011). Within each of this grid cell, survey teams walked along forest roads, trails, riverbanks, stream beds and other likely travel routes of tigers and leopards actively searching for their signs. The walk effort ranged from 4 to 40 km in each grid cell depending on the extent of forest cover available in it. Details of the field sampling protocols as well as the biological considerations and statistical survey design basis for these protocols are described elsewhere (Hines et al. 2010; Karanth et al. 2011). The survey was conducted in two phases (95 grid cells in February–June 2006 and 111 grid cells in November 2006–June 2007) and all the carnivore faecal samples collected in the second phase were used for this study (Fig. 1). A total of 2,293 km of walk effort during the second phase resulted in 353 faecal samples. Samples were collected in alcohol along with GPS coordinates of their locations and were stored in room temperature until further processing. Additionally, supplementary information such as scat diameter, secondary evidences like scrape marks, tracks etc., date, time and name of the collector were recorded for each sample.

Molecular markers

(a) *Species identification* We designed a novel multiplex PCR-Electrophoresis based approach in this study to unambiguously identify tigers and leopards from poor quality biological samples. To design primers,

Author's personal copy

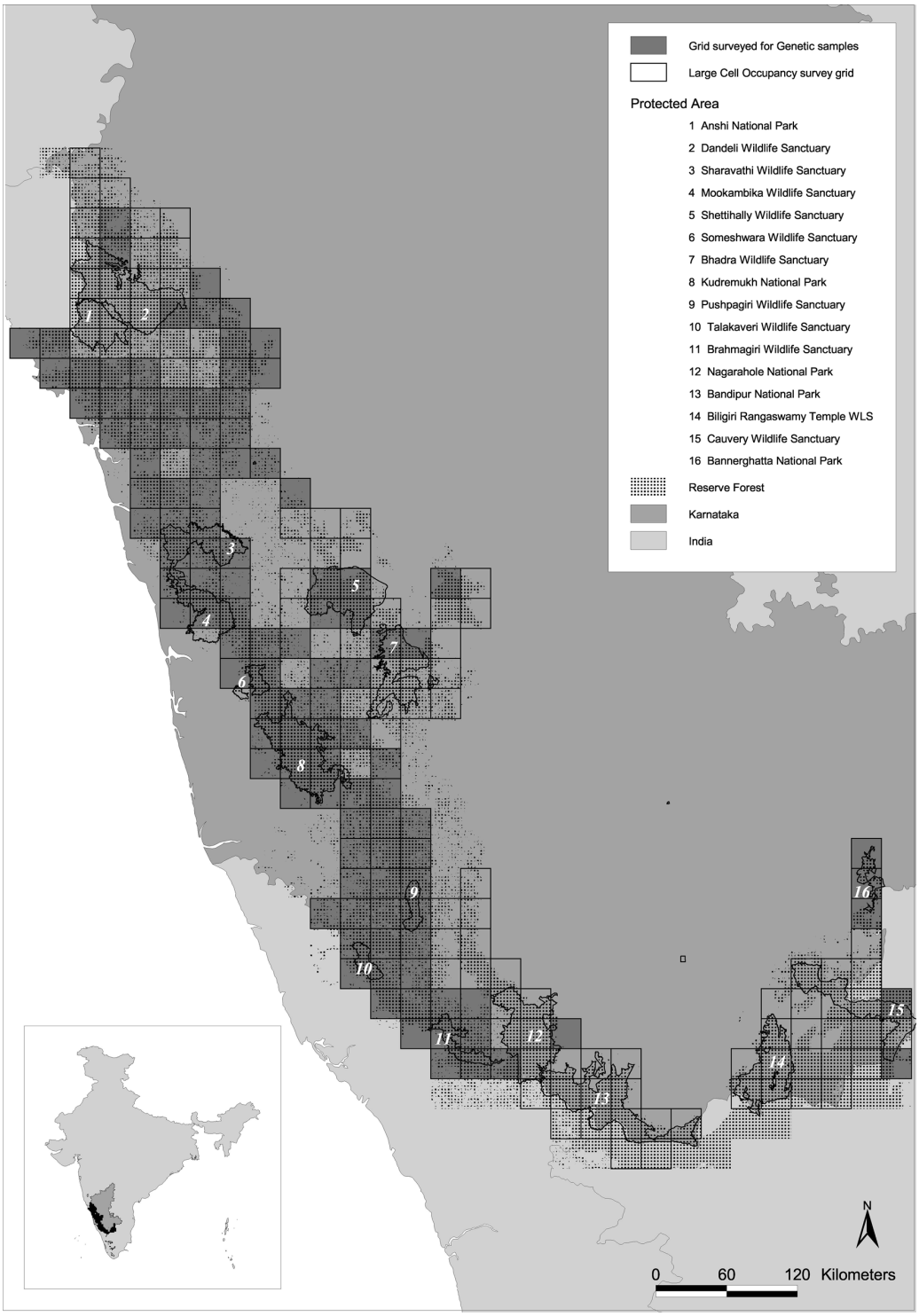


Fig. 1 Malenad-Mysore Landscape (MMTL) in Karnataka with protected areas and surrounding forest matrix with grid-cells imposed on it. The *inset map* shows the location of MMTL within India

Author's personal copy

we retrieved tiger and leopard mitochondrial DNA fragments (cytochrome-b, NADH2, NADH4, NADH5, 12s rRNA, 16s rRNA, ATPase8 and control region) from Genbank along with nuclear insertion (Numt) sequences of tigers (Kim et al. 2006). The sequences were aligned using MEGA 4 (Tamura et al. 2007) and screened visually for species-specific variations. From each of the mitochondrial fragments we designed one common forward primer and two highly species-specific reverse primers. A total of 24 primers were designed to amplify short PCR fragments (<250 bp). In addition, we also designed a pair of canid-specific primers to discriminate canids from other species using available genome sequences of sympatric canid species (wild dog, domestic dog, jackal).

(b) *Molecular sexing* We combined two previously described gender determination approaches (based on Amelogenin (Pilgrim et al. 2005) and ZFX-DBY (Sugimoto et al. 2006) markers) to design a multiplex molecular sexing system for tigers and leopards. This multiplex has two markers each from X and Y-chromosomes.

(c) *Individual identification* In this study, we assessed success rates of previously used microsatellite loci (Mondol et al. 2009a, b, c) for individual identification of tigers and leopards from field-collected faecal DNA samples. All these markers were tested using the same systematic approach explained in Mondol et al. (2009a) and the best loci were selected for further analyses.

DNA extraction and PCR standardization

DNA was extracted from all the samples following same protocols described in Mondol et al. (2009a). All initial PCR standardizations with the new primers were checked for species specificity with a set of reference samples including common sympatric carnivore and primary prey species (tiger, leopard, clouded leopard, wild dog, domestic dog, jackal, striped hyena, sloth bear, common palm civet, jungle cat, leopard cat, domestic cat, spotted deer, sambar, gaur and common langur- further called 'Reference samples'). After post-temperature standardizations, reactions were performed in Eppendorf thermocycler with 3.5 µl of multiplex buffer mix (QIAGEN Inc.), 4 µM BSA, 0.2 µM of primer mix and 1–3 µl of DNA extract with conditions including initial denaturation (95 °C for 15 min); 50 cycles of denaturation (94 °C for 30 s), annealing (50 °C for 30 s) and extension (72 °C for 35 s); followed by a final extension (72 °C for 10 min). Negative controls were included to monitor possible contamination. All PCR products were visualized in 2 % agarose gel, cleaned by Exo-SAP mixture (NEB) and sequenced from both ends to test the specificity of the primers for the targeted species. All sequences were visually inspected and matched against Genbank database for species confirmation. Further, selected tiger-leopard specific primers were multiplexed and tested with the same set of reference DNA samples and 40 faecal DNA samples of known felid species present in the lab (tiger, leopard, jungle cat, leopard cat and domestic cat- further called 'test samples') to check their specificity and evaluate their success with poor DNA sources. Similarly, the sexing primers were standardized as a multiplex with reference DNA samples of tiger, leopard, jungle cat and domestic cat following similar protocols described previously with annealing temperature of 55 °C.

Finally, all field-collected (further called 'field samples') samples were analyzed for species and sex identification. All tiger samples and selected leopard samples were individually identified using previously described microsatellites (Mondol et al. 2009a, b, c).

Analyses

All genotyping and individual identification for tigers and leopards was performed as described in Mondol et al. (2009a, c). Diversity statistics for all microsatellite markers were analyzed using ARLEQUIN 3.1 (Excoffier et al. 2005) and GIMLET (Valière 2002).

Results

Species identification

Among all newly designed mtDNA markers (n = 24) only NADH4 and NADH2 primers produced desired species-specific patterns from the reference samples. Primers from remaining regions showed non-specific amplifications from other species and were thus discarded. These six primers produced tiger-specific fragments of 85 and 110 bp and leopard-specific fragments of 190 and 130 bp, respectively. Amplifications with reference samples and test samples always produced species-specific band profiles for only tigers and leopards. No amplification was found with any other feline or other species. Sequencing of these fragments always resulted in accurate species identification when checked against Genbank database. Multiplex reaction with all six primers has always resulted in expected band patterns from tiger and leopard test samples and no cross-amplification was observed from any other carnivore or potential prey species. Therefore, any positive amplification observed with such band pattern was hence considered as originated from tigers or leopards. Similarly, the canid-

Author's personal copy

Table 1 Details of the species-specific mtDNA primers designed in this study

Primer ID	Target species	Primer sequence (5′–3′)	Amplicon size (bp)	Annealing temp.
TigParND2-F1	Common forward	TGTAGGYTGAATARCAGC		50 °C
TigND2-R1	Tiger	GGGCTGTATATGAGTAAC	110	
ParND2-R1	Leopard	GGGGACATTATTAGAACC	190	
TigParND4-F1	Common forward	TRATAGCTGCYTGATGAC		
TigND4-R1	Tiger	GCTACAAATAGCTCTGCA	85	
ParND4-R1	Leopard	GTTTGTGCCTATAAGGAC	130	
Canid-F1	Canids	CTCCACACATCTAAGCAACGC	200	59 °C
Canid-R1		GTTAAAGTGAGAAGATCGGCG		

specific primers resulted in a 200 bp fragment from reference canid samples (i.e. wild dog, domestic dog, jackal) and no amplification from any other species. Sequencing of these fragments confirmed respective canid species. Further, any sample with this amplification pattern was considered originating from canids. The primer details are presented in Table 1.

From all the field-collected samples (n = 353), 312 scats (88 %) were successfully assigned to species (Tiger/Leopard) or genera (Canid/Others) level. Overall, 229 samples were assigned to leopards, 39 to tigers and 44 to canids. The remaining 41 (12 %) samples did not show any amplification.

Sex identification

The sexing multiplex reaction PCR resulted in a four-band pattern (156, 194, 205 and 214 bp) for males and a two-band pattern (205 and 214 bp) for females. Same band patterns were always observed for felids in the reference and test samples and no amplification was found from any other species.

Success rate for sex discrimination for all the felid samples collected from field (n = 268) was 57 %. Sex was determined unambiguously for 134 leopards (n = 229, 78 females and 56 males) and 18 tigers (n = 39, 13 females and 5 males) (Fig. 2).

Individual identification

In this study, we tested previously used microsatellite loci (Mondol et al. 2009a, b, c) for tiger and leopard individual identification from extremely varied-quality faecal DNA samples. We used previously described 15 loci for tiger individual identification (Mondol et al. 2009a, b; Table 2) on all 39 tiger faecal samples. However, six of these 15 loci (FCA453, FCA391, F41, F115, FCA441 and FCA205) showed less than 20 % amplification success from the field-collected samples, compared to the remaining nine (>70 % success rate). We identified 18 unique individuals and five genetic 'recaptures' using these nice loci. These loci provided a cumulative probability of identity for siblings ($P_{ID(sibs)}$) of 1×10^{-4} and mean genotype error rate was <2 % (See Table 2).

Similarly we examined eight loci described at Mondol et al. (2009c) for leopard individual identification. Because this was a pilot study, we selected 50 geographically separated leopard faeces from the 229 field-collected samples for this. We identified 39 individual leopards from the selected faecal samples. The probability of identity for siblings ($P_{ID(sibs)}$) and mean genotype error rate was 1×10^{-4} and <3 %, respectively (See Table 2). Overall, 60 and 78 % of the selected tiger and leopard samples were individually identified with these microsatellite panels, respectively.

Discussion

We developed PCR-based species, sex and individual identification tools for tigers and leopards and used them in a pilot study to monitor tiger/leopard distribution and demography in parts of a globally important tiger conservation landscape. Earlier Karanth et al. (2011) have used sign-track surveys and faecal encounters to study these sympatric species in this landscape, but faeces similarity (both shape and size) has made species and sex identification a challenging task. Photographic surveys are also logistically challenging in such landscape studies as they require equipment, skilled personals, intensive effort in low animal density areas, particularly outside protected zones. The molecular tools developed in this study thus provide certain advantages over earlier approaches used in carnivore studies (Cossíos and Angers 2006; Sugimoto et al. 2006; Bidlack et al. 2007; Livia et al. 2007; Haag et al. 2009; Roques et al. 2010). The approaches proposed here facilitate rapid screening of large number of samples using fewer steps in sample processing (DNA extraction, single multiplex PCR, electrophoresis), reduced species/sex misidentifications, and are cheaper than PCR–RFLP (Nagata

Author's personal copy

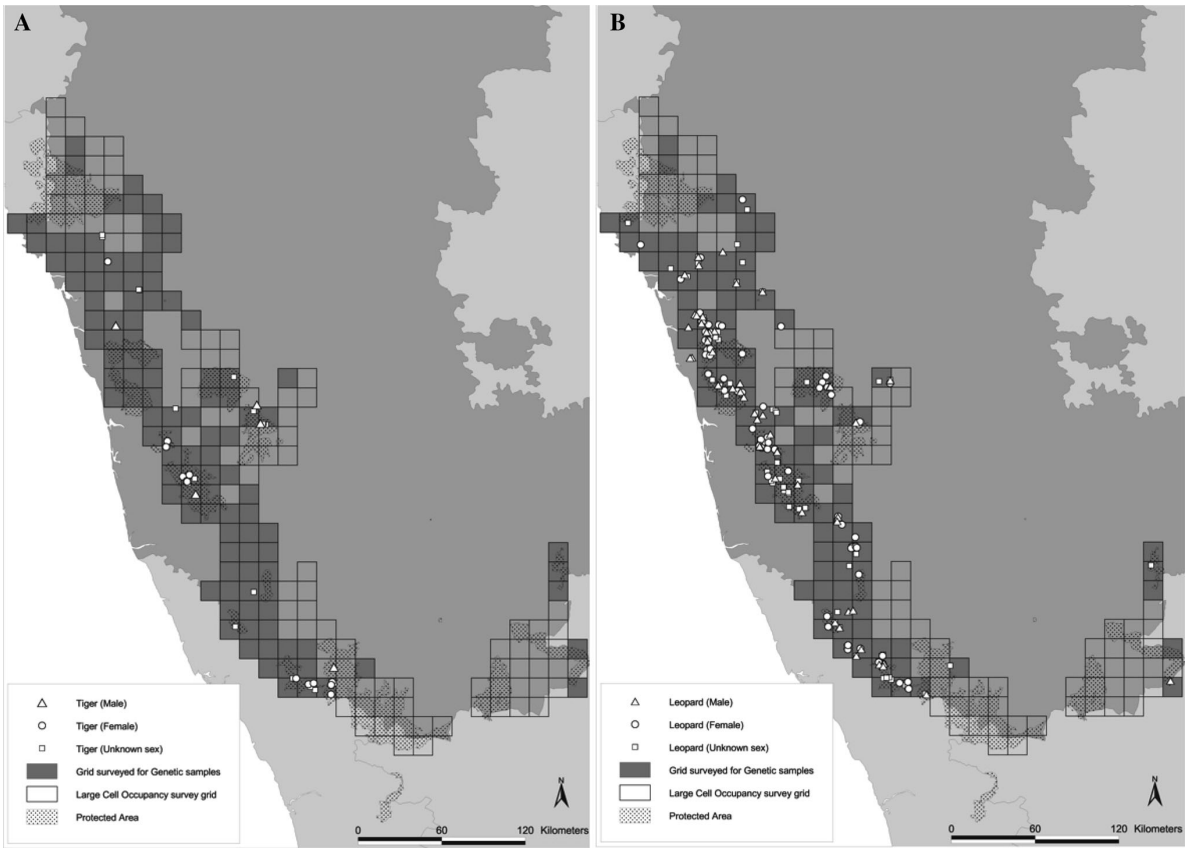


Fig. 2 Gender identification of tigers (**a**) and leopards (**b**) in Malenad-Mysore Landscape (MMTL) in Karnataka, India

et al. 2005; Cossíos and Angers 2006; Bidlack et al. 2007) or sequencing (Bhagavatula and Singh 2006; Perez et al. 2006; Pandey et al. 2007; Busby et al. 2009) approaches.

We were very careful while developing and standardizing these new molecular approaches as we aimed to study only two species in a carnivore hotspot. For example, we designed the species-specific markers from multiple genomic regions to ensure higher success rate and standardized them on large numbers of reference and field samples with varied qualities, assuring high success rates (88 %) from field-collected samples. Similarly, our four-marker-based sexing approach showed reduced rate of 'false negatives' and better success rate [57 % in this study compared to 47 % in Sugimoto et al. (2006)] in gender identification from a range of poor quality samples. Positive amplifications from other small felines (jungle cat, leopard cat) from the reference samples suggest this multiples sexing system would be useful in other small feline species across their distribution.

Microsatellite screening with extremely varied quality field-collected faecal samples in this study clearly showed very low amplification success from six of the 15 previously used loci (Mondol et al. 2009a, b) for tiger individual identification. Remaining nine loci showed high amplification success (<70 %), low genotyping error (>2 %) and high statistical confidence ($P_{ID(sibs)} = 1 \times 10^{-4}$), making this panel ideal for future tiger identification studies. On the other hand, previously described eight-loci panel for leopard individual identification showed higher statistical confidence ($P_{ID(sibs)} = 1 \times 10^{-4}$), lower genotyping error (>3 %) and higher level of genetic variations (Table 2) than earlier study by Mondol et al. (2009c). Adding recently developed polymorphic miniSTRs (Mondol et al. 2012) into these panels could be beneficial in future population genetics and forensic studies for these endangered species.

Finally, the pilot genetic study at the Malenad-Mysore region also provides useful insight on tiger and leopard ecology in this important conservation landscape. For example, results indicate about 40 % of the leopard samples (93 of 229) were found outside protected areas, compared to 28 % of tiger samples (11 of 39). This pattern is not surprising as tigers socially dominate leopards (Seidensticker 1976; McDougal 1988; Miquelle et al. 2005), and adapted to a wide-range of habitats outside

Author's personal copy

Table 2 Details of the tiger (n-23) and leopard (n = 39) microsatellite markers used in this study

Locus	Species	Success rate (%)	No. of alleles	H_E	H_O	Null allele	Allelic dropout (%)	False allele (%)	HWE	$P_{ID(sibs)}$	Reference
FCA126[a]	Tiger	92	10	0.78	0.62	No	0	0	Yes	$3.11e^{-1}$	Menotti-Raymond et al. (1999)
FCA069[a]	Tiger	85	11	0.82	0.50	No	0	0	Yes	$1.0e^{-1}$	Menotti-Raymond et al. (1999)
FCA232[a]	Tiger	87	10	0.84	0.50	No	1.5	0	Yes	$3.24e^{-2}$	Menotti-Raymond et al. (1999)
FCA090[a]	Tiger	92	12	0.87	0.54	No	0.5	0	Yes	$1.09e^{-2}$	Menotti-Raymond et al. (1999)
FCA672[a]	Tiger	88	12	0.89	0.65	No	1	0	Yes	$3.75e^{-3}$	Menotti-Raymond et al. (1999)
FCA304[a]	Tiger	78	10	0.87	0.45	No	0	0	Yes	$1.31e^{-3}$	Menotti-Raymond et al. (1999)
FCA230[a]	Tiger	83	11	0.85	0.40	No	0	0	Yes	$4.64e^{-4}$	Menotti-Raymond et al. (1999)
FCA279[a]	Tiger	87	8	0.70	0.33	Yes	0.5	0	Yes	$1.75e^{-4}$	Menotti-Raymond et al. (1999)
FCA628[a]	Tiger	71	8	0.83	0.68	No	0	0.5	Yes	$7.57e^{-5}$	Menotti-Raymond et al. (1999)
FCA453[b]	Tiger	40	6	–	–	–	–	–	–	–	Menotti-Raymond et al. (1999)
FCA391[b]	Tiger	17	3	–	–	–	–	–	–	–	Menotti-Raymond et al. (1999)
F41[b]	Tiger	32	3	–	–	–	–	–	–	–	Menotti-Raymond et al. (1999)
F115[b]	Tiger	42	4	–	–	–	–	–	–	–	Menotti-Raymond et al. (1999)
FCA441[b]	Tiger	50	5	–	–	–	–	–	–	–	Menotti-Raymond et al. (1999)
FCA205[b]	Tiger	38	5	–	–	–	–	–	–	–	Menotti-Raymond et al. (1999)
FCA126	Leopard	92	13	0.88	0.75	No	0	0	Yes	$3.07e^{-1}$	Menotti-Raymond et al. (1999)
FCA279	Leopard	87	8	0.74	0.49	No	2	0	Yes	$9.75e^{-2}$	Menotti-Raymond et al. (1999)
FCA304	Leopard	67	11	0.86	0.83	No	3	0	Yes	$3.14e^{-2}$	Menotti-Raymond et al. (1999)
FCA672	Leopard	97	13	0.85	0.75	No	0	0	Yes	$1.03e^{-2}$	Menotti-Raymond et al. (1999)
FCA090	Leopard	90	15	0.87	0.50	No	0.5	0	Yes	$3.45e^{-3}$	Menotti-Raymond et al. (1999)
FCA052	Leopard	97	12	0.85	0.76	No	1	0	Yes	$1.17e^{-3}$	Menotti-Raymond et al. (1999)
FCA309	Leopard	100	14	0.90	0.76	No	0	0	Yes	$3.99e^{-4}$	Menotti-Raymond et al. (1999)
FCA230	Leopard	97	13	0.89	0.78	No	0	0	Yes	$1.63e^{-4}$	Menotti-Raymond et al. (1999)

H_E expected heterozygosity, *H_O* observed heterozygosity, *HWE* Hardy–Weinberg equilibrium

[a] Loci finally used for tiger individual identification

[b] Loci discarded due to low amplification success. Due to low success rate some of the analysis were not conducted for these loci

protected areas (Seidensticker 1976). Results also showed that about 50 % of the samples outside protected areas (2 tigers and 47 leopards) were females, implicating both male and female driven dispersal in this landscape. However, given our knowledge on males-biased dispersal on tigers (Smith 1993; Patil et al. 2011), a longer-term study is

Author's personal copy

needed to conclusively assess such both-sex dispersal patterns at landscape scales.

Apart from the Malenad-Mysore Tiger Landscape, the Indian subcontinent currently hosts some of the most important tiger habitats (Sanderson et al. 2006). Survival of these endangered species depends on timely information on distribution, demography and ecology at landscape levels. Earlier studies on tigers and leopards have investigated the mechanisms facilitating their coexistence (Johnsingh 1983; Karanth and Sunquist 1995, 2000; Wang and Macdonald 2009; Wegge et al. 2009) or competition between them (Karanth and Nichols 1998; Odden et al. 2010) at local scales. The pilot study at MMTL demonstrates that faecal DNA-based species, sex and individual identification can help us to understand such patterns at a landscape level. We aim for a long-term genetic study to get regular information on landscape level population size estimates, sex-ratio, genetic relatedness and dispersal events throughout the subcontinent. Given the alarming rate of global population decline within the large carnivore families (Ceballos et al. 2005; Schipper et al. 2008), such synergetic approaches will perhaps be the key for their future survival.

Acknowledgments We are thankful to the forest department of Karnataka for providing permits to carry out the research in this landscape. We also thank the CWS field teams, local forest officials and volunteers for their assistance. We thank V. Kolipakam, R. Samant, P. Yadav and other Ramakrishnan lab members for their help during this study. SM was supported by Department of Science and Technology INSPIRE Faculty Award. This research was funded by Department of Science and Technology, Government of India grant awarded to KUK and UR, National Centre for Biological Sciences, and Wildlife Conservation Society. Center for Wildlife Studies provided logistical and material supports for the fieldwork. UR was supported by the Ramanujan fellowship.

References

Beja-Pereira A, Oliveira R, Alves PC, Schwartz MK, Luikart G (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. Mol Ecol Res 9:1279–1301

Bellemain E, Nawaz A, Valentini A, Swenson JE, Taberlet P (2007) Genetic tracking of the brown bear in northern Pakistan and implications for conservation. Biol Conserv 134:537–547

Bhagavatula J, Singh L (2006) Genotyping faecal samples of Bengal tiger (*Panthera tigris tigris*) for population estimation: a pilot study. BMC Genet 7:48

Bidlack AL, Reed SE, Palsbøll PJ, Getz WM (2007) Characterization of a western North American carnivore community using PCR–RFLP of cytochrome b obtained from fecal samples. Conserv Genet 8:1511–1513

Busby GBJ, Gottelli D, Wacher T et al (2009) Genetic analysis of scat reveals leopard *Panthera pardus* and cheetah *Acinonyx jubatus* in southern Algeria. Oryx 43:412–415

Ceballos G, Erlich PR, Soberon J, Salazar I, Fay JP (2005) Global mammal conservation: What must we manage? Science 309:603–607

Cossíos D, Angers B (2006) Identification of Andean felid species using PCR-RFLP. Mastozool Neotrop 13:239–244

Davison A, Birks JDS, Brookes RC, Braithwaite RC, Messenger JE (2002) On the origin of faeces: morphological versus molecular methods for surveying rare carnivores from their scats. J Zool 257:141–143

Excoffier L, Laval G, Schneider S (2005) ARLEQUIN ver 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50

Fernandes CA, Ginja C, Pereira I et al (2008) Species-specific mitochondrial DNA markers for identification of non-invasive samples from sympatric carnivores in the Iberian Peninsula. Conserv Genet 9:681–690

Foresman KR, Pearson DE (1998) Comparison of proposed survey techniques for detection of forest carnivores. J Wildl Manag 62:1217–1226

Goossens B, Bruford MW (2009) Non-invasive genetic analysis in conservation. In: Bertorelle G, Bruford MW, Hauffe HC, Rizzoli A, Vernesi C (eds) Population genetics for animal conservation. Cambridge University Press, United Kingdom, pp 167–201

Haag T, Santos AS, Angelo CD et al (2009) Development and testing of an optimized method for DNA-based identification of jaguar (*Panthera onca*) and puma (*Puma concolor*) faecal samples for use in ecological and genetic studies. Genetica 136:505–512

Hines JE, Nichols JD, Royle AJ et al (2010) Tigers on trails: occupancy modeling for luster sampling. Ecol Appl 20:1456–1466

Johnsingh AJT (1983) Large mammalian prey-predators in Bandipur. J Bombay Nat Hist Soc 80:1–57

Karanth KU, Nichols JD (1998) Estimation of tiger densities in India using photographic captures and recaptures. Ecology 79:2852–2862

Karanth KU, Sunquist ME (1995) Prey selection by tiger leopard and dhole in tropical forests. J Anim Ecol 64:439–450

Karanth KU, Sunquist ME (2000) Behavioural correlates of predation by tiger (Panthera tigris) leopard (*Panthera pardus*) and dhole (*Cuon alpinus*) in Nagarahole, India. J Zool 250:255–265

Karanth KK, Nichols JD, Hines JE, Karanth KU, Christensen NL (2009) Patterns and determinants of mammal species occurrence in India. J Appl Ecol 46:1189–1200

Karanth KU, Gopalaswamy A, Kumar NS et al (2011) Monitoring carnivore populations at the landscape scale: occupancy modelling of tigers from sign surveys. J Appl Ecol 48:1048–1056

Kim JH, Antunes A, Luo SJ et al (2006) Evolutionary analysis of a large mtDNA translocation (*numt*) into the nuclear genome of the *Panthera* genus species. Gene 366:292–302

Livia L, Francesca V, Antonella P, Fausto P, Bernardino R (2007) A PCR-RFLP method on faecal samples to distinguish *Martes martes*, *Martes foina*, *Mustela putorius* and *Vulpes vulpes*. Conserv Genet 8:757–759

McDougal C (1988) Leopard and tiger interactions at Royal Chitwan National Park, Nepal. J Bombay Nat Hist Soc 85:609–610

Menotti-Raymond et al (1999) A genetic linkage map of microsatellites in the domestic cat (*Felis catus*). Genomics 57:9–23

Miquelle DG, Stephens PA, Smirnov EN et al (2005) Tigers and wolves in the Russian Far East: competitive exclusion functional redundancy and conservation implications. In: Ray JC, Berger J, Redford KH, Steneck R (eds) Large carnivores and the conservation of biodiversity. Island Press, Washington, pp 179–207

Mittermeier RA, Myers N, Thomsen JB, Da Fonesca GAB, Oliveri S (1998) Biodiversity hotspots and major tropical wilderness areas: approaches to setting conservation priorities. Conserv Biol 12:516–520

Mondol S, Karanth KU, Kumar NS et al (2009a) Evaluation of non-invasive genetic sampling methods for estimating tiger population size. Biol Conserv 142:2350–2360

Mondol S, Karanth KU, Ramakrishnan U (2009b) Why the Indian subcontinent holds the key to global tiger recovery. PLoS Genet 5(8):e1000585

Author's personal copy

Mondol S, Navya R, Athreya V et al (2009c) A panel of microsatellites to individually identify leopards and its application to leopard monitoring in human dominated landscapes. BMC Genet 10:79

Mondol S, Thatte P, Yadav P, Ramakrishnan U (2012) A set of miniSTRs for population genetic analyses of tigers (*Panthera tigris*) with cross-species amplification for seven other Felidae. Conserv Genet Resour 4:63–66

Moruzzi TL, Fuller TK, DeGraaf RM, Brooks RT, Li W (2002) Assessing remotely triggered cameras for surveying carnivore distribution. Wildl Soc Bull 2:380–386

Nagata J, Aramilev VV, Belozor A, Sugimoto T, McCullough DR (2005) Fecal genetic analysis using PCR-RFLP of cytochrome b to identify sympatric carnivores the tiger *Panthera tigris* and the leopard *Panthera pardus* in far eastern Russia. Conserv Genet 6:863–866

Odden M, Wegge P, Fredriksen T (2010) Do tigers displace leopards? If so why? Ecol Res 25:875–881

Palomares F, Rodriguez R, Laffitte R, Delibes M (1991) The status and distribution of the Iberian lynx *Felis pardina* (Temminck) in Coto Donana area, SW Spain. Biol Conserv 57:59–69

Palomares F, Godoy JA, Piriz A, O'Brien SJ, Johnson WE (2002) Fecal genetic analysis to determine the presence and distribution of elusive carnivores: design and feasibility for the Iberian Lynx. Mol Ecol 11:2171–2182

Pandey PK, Dhotre DP, Dharne MS et al (2007) Conservation and Evolutionary Genetics of Indian Leopard *Panthera pardus fusca* as studied by mitochondrial 12S rRNA gene sequence analysis. Curr Sci 92:1129–1133

Patil N, Kumar NS, Gopalaswamy AM, Karanth KU (2011) Dispersing tiger makes a point. Oryx 45:472–475

Perez I, Geffen E, Mokady O (2006) Critically Endangered Arabian leopards *Panthera pardus nimr* in Israel: estimating population parameters using molecular scatology. Oryx 40:295–301

Pilgrim KL, Mckelvey KS, Riddle AE, Schwartz MK (2005) Felid sex-identification based on noninvasive genetic samples. Mol Ecol Notes 5:60–61

Prugh LR, Ritland CE (2005) Molecular testing of observer identification of carnivore feces in the field. Wildl Soc Bull 33:189–194

Ranganathan J, Chan KMA, Karanth KU, Smith JLD (2008) Where can tigers persist in the future? A landscape-scale density-based population model for the Indian subcontinent. Biol Conserv 141:67–77

Roques S, Adrados B, Chavez C et al (2010) Identification of neotropical felid faeces using RCP-PCR. Mol Ecol Res 11:171–175

Sanderson E, Forrest J, Loucks C et al (2006) Setting priorities for the conservation and recovery of wild tigers: 2005–2015. The Technical Assessment WCS WWF Smithsonian and NFWF-STF New York, Washington, DC

Schipper J, Chanson JS, Chiozza F et al (2008) The status of the world's land and marine mammals: diversity, threat and knowledge. Science 322:225–230

Seidensticker J (1976) On the ecological separation between tigers and leopards. Biotropica 8:225–234

Smallwood KS, Fitzhhugh EL (1995) A track count for estimating mountain lion *Felis concolor californica* population trend. Biol Conserv 71:251–259

Smith JLD (1993) The role of dispersal in structuring the Chitwan tiger population. Behaviour 124:165–195

Sugimoto T, Nagata J, Aramilev VV et al (2006) Species and sex identification from faecal samples of sympatric carnivores Amur leopard and Siberian tiger in the Russian Far East. Conserv Genet 7:799–802

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Uphyrkina O, Johnson WE, Quigley H et al (2001) Phylogenetics, genome diversity and origin of modern leopard *Panthera pardus*. Mol Ecol 10:2617–2633

Valière N (2002) GIMLET: a computer program for analyzing genetic identification data. Mol Ecol Notes 2:377–379

Walston J, Robinson JG, Bennett EL et al (2010) Bringing the tiger back from the brink—the six percent solution. PLoS Biol 8:e1000485

Wang S, Macdonald D (2009) Feeding habits and niche partitioning in a predator guild composed of tigers, leopards and dholes in a temperate ecosystem in central Bhutan. J Zool 277:275–283

Weber W, Rabinowitz A (1996) A global perspective on large carnivore conservation. Conserv Biol 10:1046–1054

Wegge P, Odden M, Pokharel CP, Storaas T (2009) Predator–prey relationships and responses of ungulates and their predators to the establishment of protected areas: a case study of tigers, leopards and their prey in Bardia National Park, Nepal. Biol Conserv 142:189–202

Wultsch C, Waits LP, Kelly MJ (2014) Noninvasive individual and species identification of jaguars (*Panthera onca*) pumas (*Puma concolor*) and ocelots (*Leopardus pardalis*) in Belize. Mol Ecol Res, Central America using cross-species microsatellites and faecal DNA. doi:10.1111/1755-0998.12266

Zuercher GL, Gipson PS, Stewart GC (2003) Identification of carnivore feces by local peoples and molecular analyses. Wildl Soc Bull 31:961–970

Bibliography

Abadi, F., Gimenez, O., Arlettaz, R. & Schaub, M. (2010a) An assessment of integrated population models: bias, accuracy, and violation of the assumption of independence. *Ecology*, **91**, 7–14.

Abadi, F., Gimenez, O., Ullrich, B., Arlettaz, R. & Schaub, M. (2010b) Estimation of immigration rate using integrated population models. *Journal of Applied Ecology*, **47**, 393–400.

Amstrup, S.C., McDonald, T.L. & Manly, B.F.J., eds. (2005) *Handbook of Capture-Recapture Analysis*. Princeton University Press, U.S.A.

Andheria, A.P. (2006) *Assessment of diet and abundance of large carnivores from field surveys of scats*. Master's thesis, Manipal University.

Archaux, F., Henry, P.Y. & Gimenez, O. (2012) When can we ignore the problem of imperfect detection in comparative studies? *Methods in Ecology and Evolution*, **3**, 188–194.

Axelrod, R. & Hamilton, W.D. (1981) The evolution of cooperation. *Science*, **211**, 1390–1396.

Baber, D.W. & Coblentz, B.E. (1986) Density, home range, habitat use, and reproduction in feral pigs on Santa Catalina Island. *Journal of Mammalogy*, **67**, 512–525.

Barea-Azcón, J., Virgós, E., Ballesteros-Duperon, E., Moleón, M. & Chirosa, M. (2007) Surveying carnivores at large spatial scales: a comparison of four broad-applied methods. *Vertebrate Conservation and Biodiversity*, pp. 387–404.

Barrette, C. (1977) Some aspects of the behaviour of muntjacs in Wilpattu National Park. *Mammalia*, **41**, 1–34.

Barrette, C. (1987) The comparative behavior and ecology of chevrotains, musk deer, and morphologically conservative deer. *Biology and management of the Cervidae* (ed. C.. Wemmer), pp. 200–213. Smithsonian Institution, Washington, USA.

Bengsen, A.J., Leung, L.K.P., Lapidge, S.J. & Gordon, I.J. (2011) Using a general index approach to analyze camera-trap abundance indices. *The Journal of Wildlife Management*, **75**, 1222–1227.

Bernardo, J.M. & Smith, A.F.M. (1994) *Bayesian Theory*. Chichester: Wiley.

Bolker, B.M. (2008) *Ecological models and data in R*. Princeton University Press, Princeton, New Jersey, USA.

Bonesi, L. & Macdonald, D. (2004) Evaluation of sign surveys as a way to estimate the relative abundance of American mink (Mustela vison). *Journal of Zoology*, **262**, 65–72.

Borchers, D.L. & Efford, M.G. (2008) Spatially explicit maximum likelihood methods for capture–recapture studies. *Biometrics*, **64**, 377–385.

Brooks, S.P. & Gelman, A. (1998) General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, **7**, 434–455.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. & Thomas, L. (2001) *Introduction to Distance Sampling*. Oxford University Press, Oxford, UK.

Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Verlag.

Caley, P. (1997) Movements, activity patterns and habitat use of feral pigs (Sus scrofa) in a tropical habitat. *Wildlife Research*, **24**, 77–87.

Carbone, C. & Gittleman, J.L. (2002) A common rule for the scaling of carnivore density. *Science*, **295**, 2273–2276.

Casella, G. & Berger, R.L. (1990) *Statistical Inference*. Duxbury Press, Belmont, CA, USA.

Ceballos, G., Ehrlich, P.R., Soberon, J., Salazar, I. & Fay, J.P. (2005) Global Mammal Conservation: What Must We Manage? *Science*, **309**, 603–607.

Chandler, R.B. & Royle, J.A. (2013) Spatially explicit models for inference about density in unmarked or partially marked populations. *The Annals of Applied Statistics*, **7**, 936–954.

Chevin, L.M., Lande, R. & Mace, G.M. (2010) Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *PLoS biology*, **8**, e1000357.

Clark, J.S., Carpenter, S.R., Barber, M., Collins, S., Dobson, A., Foley, J.A., Lodge, D.M., Pascual, M., Pielke Jr, R., Pizer, W. *et al.* (2001) Ecological forecasts: an emerging imperative. *Science*, **293**, 657–660.

Clutton-Brock, T. & Sheldon, B.C. (2010) Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology & Evolution*, **25**, 562–573.

Cochran, W. (1937) Problems arising in the analysis of a series of similar experiments. *Supplement to the Journal of the Royal Statistical Society*, **4**, 102–118.

Conroy, M.J., Runge, J.P., Barker, R.J., Schofield, M.R. & Fonnesbeck, C.J. (2008) Efficient estimation of abundance for patchily distributed populations via two-phase, adaptive sampling. *Ecology*, **89**, 3362–3370.

Conry, J.P. (1989) Gaur *Bos gaurus* and development in Malaysia. *Biological Conservation*, **49**, 47–65.

Coulson, T., Benton, T., Lundberg, P., Dall, S. & Kendall, B. (2006) Putting evolutionary biology back in the ecological theatre: a demographic framework mapping genes to communities. *Evolutionary Ecology Research*, **8**, 1155–1171.

Das, A., Krishnaswamy, J., Bawa, K.S., Kiran, M.C., Srinivas, V., Kumar, N.S. & Karanth, K.U. (2006) Prioritization of conservation areas in the Western Ghats, India. *Biological Conservation*, **133**, 16–31.

Dawson, D.K. & Efford, M.G. (2009) Bird population density estimated from acoustic signals. *Journal of Applied Ecology*, **46**, 1201–1209.

De Valpine, P. (2003) Better inferences from population-dynamics experiments using Monte Carlo state-space likelihood methods. *Ecology*, **84**, 3064–3077.

Dice, L.R. (1941) Methods for estimating populations of mammals. *The Journal of Wildlife Management*, **5**, 398–407.

Diefenbach, D.R., Conroy, M.J., Warren, R.J., James, W.E., Baker, L.A. & Hon, T. (1994) A test of the scent-station survey technique for bobcats. *Journal of Wildlife Management*, **58**, 10–17.

Dinerstein, E. (1979) An ecological survey of the Royal Karnali-Bardia Wildlife Reserve, Nepal, Part II; habitat/animal interaction. *Biological Conservation*, **16**, 265–300.

Dingemanse, N.J., Kazem, A.J., Réale, D. & Wright, J. (2010) Behavioural reaction norms: animal personality meets individual plasticity. *Trends in Ecology & Evolution*, **25**, 81–89.

Dorazio, R.M. & Royle, J.A. (2003) Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics*, **59**, 351–364.

Dunn, J.E. & Gipson, P.S. (1977) Analysis of radio telemetry data in studies of home range. *Biometrics*, pp. 85–101.

Eberhardt, L.L. & Simmons, M.A. (1987) Calibrating population indices by double sampling. *The Journal of Wildlife Management*, **51**, 665–675.

Efford, M. (2004) Density estimation in live-trapping studies. *Oikos*, **106**, 598–610.

Efford, M.G., Dawson, D.K. & Borchers, D.L. (2009) Population density estimated from locations of individuals on a passive detector array. *Ecology*, **90**, 2676–2682.

Efford, M.G., Dawson, D.K. & Robbins, C.S. (2004) DENSITY: software for analysing capture-recapture data from passive detector arrays. *Animal Biodiversity and Conservation*, **27**, 217–228.

Efford, M.G. & Fewster, R.M. (2013) Estimating population size by spatially explicit capture–recapture. *Oikos*, **122**, 918–928.

Efron, B. & Tibshirani, R. (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical science*, **1**, 54–75.

Eggert, L.S., Eggert, J.A. & Woodruff, D.S. (2003) Estimating population sizes for elusive animals: the forest elephants of Kakum National Park, Ghana. *Molecular Ecology*, **12**, 1389–1402.

Ellison, A.M. (2004) Bayesian inference in ecology. *Ecology Letters*, **7**, 509–520.

Ergon, T. & Gardner, B. (2013) Separating mortality and emigration: modelling space use, dispersal and survival with robust-design spatial capture–recapture data. *Methods in Ecology and Evolution*.

Estes, J.A., Terborgh, J., Brashares, J.S., Power, M.E., Berger, J., Bond, W.J., Carpenter, S.R., Essington, T.E., Holt, R.D., Jackson, J.B.C. *et al.* (2011) Trophic Downgrading of Planet Earth. *Science*, **333**, 301.

Freeman, M.F. & Tukey, J.W. (1950) Transformations related to the angular and the square root. *The Annals of Mathematical Statistics*, **21**, 607–611.

Fretwell, S.D. (1972) *Populations in a seasonal environment*. Princeton University Press.

Funston, P., Frank, L., Stephens, T., Davidson, Z., Loveridge, A., Macdonald, D., Durant, S., Packer, C., Mosser, A. & Ferreira, S. (2010) Substrate and species constraints on the use of track incidences to estimate African large carnivore abundance. *Journal of Zoology*, **281**, 56–65.

Gardner, B., Reppucci, J., Lucherini, M. & Royle, J.A. (2010a) Spatially explicit inference for open populations: estimating demographic parameters from camera-trap studies. *Ecology*, **91**, 3376–3383.

Gardner, B., Royle, J.A. & Wegan, M.T. (2009) Hierarchical models for estimating density from DNA mark-recapture studies. *Ecology*, **90**, 1106–1115.

Gardner, B., Royle, J.A., Wegan, M.T., Rainbolt, R.E. & Curtis, P.D. (2010b) Estimating black bear density using DNA data from hair snares. *The Journal of Wildlife Management*, **74**, 318–325.

Gelman, A., Roberts, G. & Gilks, W. (1996) Efficient metropolis jumping rules. *Bayesian Statistics*, **5**, 599–608.

Gelman, A. & Rubin, D.B. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–472.

Gelman, A. & Shalizi, C.R. (2010) Philosophy and the practice of Bayesian statistics. *Arxiv preprint arXiv:1006.3868*.

Geweke, J. (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics* (eds. J.M. Bernardo, J.O. Berger, A.P. Dawid & A.F.M. Smith), vol. 4. Oxford University Press, Oxford, UK.

Goldstein, J., Haran, M., Simeonov, I., Fricks, J. & Chiaromonte, F. (2014) An attraction-repulsion point process model for respiratory syncytial virus infections. *arXiv preprint arXiv:1401.5506*.

Goodrich, J.M., Miquelle, D.G., Smirnov, E.N., Kerley, L.L., Quigley, H.B. & Hornocker, M.G. (2010) Spatial structure of Amur (Siberian) tigers (Panthera tigris altaica) on Sikhote-Alin Biosphere Zapovednik, Russia. *Journal of Mammalogy*, **91**, 737–748.

Gopalaswamy, A.M., Royle, A.J., Hines, J.E., Singh, P., Jathanna, D., Kumar, N.S. & Karanth, K.U. (2012a) *SPACECAP: A Program to Estimate Animal Abundance and Density using Bayesian Spatially-Explicit Capture-Recapture Models*. Wildlife Conservation Society, India Program, Centre for Wildlife Studies, Bangalore, India. Version 1.0.6.

Gopalaswamy, A., Karanth, K., Kumar, N. & Macdonald, D. (2012b) Estimating tropical forest ungulate densities from sign surveys using abundance models of occupancy. *Animal Conservation*, **15**, 669–679.

Gopalaswamy, A.M. (2013) Capture–Recapture Models, Spatially Explicit. *Wiley StatsRef: Statistics Reference Online*.

Gopalaswamy, A.M., Delampady, M., Karanth, K.U., Kumar, N.S. & Macdonald, D.W. (In review) An examination of index-calibration experiments: counting tigers at macroecological scales. *Methods in Ecology and Evolution*.

Gopalaswamy, A.M., Royle, J.A., Delampady, M., Nichols, J.D., Karanth, K.U. & Macdonald, D.W. (2012c) Density estimation in tiger populations: combining information for strong inference. *Ecology*, **93**, 1741–1751.

Gopalaswamy, A.M., Royle, J.A., Hines, J.E., Singh, P., Jathanna, D., Kumar, N. & Karanth, K.U. (2012d) Program SPACECAP: software for estimating animal density using spatially explicit capture–recapture models. *Methods in Ecology and Evolution*, **3**, 1067–1072.

Goswami, V., Madhusudan, M.D. & Karanth, K.U. (2007) Application of photographic capture-recapture modeling to the estimation of demographic parameters in male Asian elephants. *Animal Conservation*, **10**, 391–399.

Gruber, B., Reineking, B., Calabrese, J.M., Kranz, A., Poledníková, K., Poledník, L., Klenke, R., Valentin, A. & Henle, K. (2008) A new method for estimating visitation rates of cryptic animals via repeated surveys of indirect signs. *Journal of Applied Ecology*, **45**, 728–735.

Gusset, M. & Burgener, N. (2005) Estimating larger carnivore numbers from track counts and measurements. *African Journal of Ecology*, **43**, 320–324.

Hanski, I. & Hanski, I.A. (1999) *Metapopulation ecology*, vol. 232. Oxford University Press Oxford.

Hanski, I., Henttonen, H. & Hansson, L. (1994) Temporal variability and geographical patterns in the population density of microtine rodents: a reply to Xia and Boonstra. *American Naturalist*, **144**, 329–342.

Hassell, M.P., Comins, H. & May, R.M. (1984) Species coexistence and self-organizing spatial dynamics. *Nature*, **370**.

Hastings, A., Arzberger, P., Bolker, B., Collins, S., Ives, A., Johnson, N. & Palmer, M. (2005) Quantitative bioscience for the 21st century. *BioScience*, **55**, 511–517.

Hatler, D.F., Nagorsen, D.W. & Beal, A.M. (2009) Non-invasive survey methods for carnivores. *Mammalia*, **73**, 273–280.

Hiby, L., Lovell, P., Patil, N., Kumar, N.S., Gopalaswamy, A.M. & Karanth, K.U. (2009) A tiger cannot change its stripes: using a three-dimensional model to match images of living tigers and tiger skins. *Biology Letters*, **5**, 383.

Hines, J.E. (2006) *PRESENCE: Software to estimate patch occupancy rates and related parameters*. Patuxent Wildlife Research Center, Laurel, Maryland.

Hines, J.E., Nichols, J.D., Royle, J.A., MacKenzie, D.I., Gopalaswamy, A.M., Kumar, N.S. & Karanth, K.U. (2010) Tigers on trails: occupancy modeling for cluster sampling. *Ecological Applications*, **20**, 1456–1466.

Hlynka, M. & Loach, D. (2005) Generating Uniform Random Points in a Regular n Sided Polygon.
URL `http://www2.uwindsor.ca/hlynka/loachpolygon.pdf`

Holt, R.D. (1993) Ecology at the mesoscale: the influence of regional processes on local communities. *Species diversity in ecological communities. University of Chicago Press, Chicago*, **101**, 77–88.

Horne, J.S. & Garton, E.O. (2006) Selecting the best home range model: an information-theoretic approach. *Ecology*, **87**, 1146–1152.

Hutchinson, G.E. (1965) *The Ecological Theater and the Evolutionary Play*. Yale University Press, New Haven, Connecticut.

Isaac, N.J.B., Cruickshanks, K.L., Weddle, A.M., Rowcliffe, J.M., Brereton, T.M., Dennis, R.L.H., Shuker, D.M. & Thomas, C.D. (2011) Distance sampling and the challenge of monitoring butterfly populations. *Methods in Ecology and Evolution*, **2**, 585–594.

Jathanna, D., Karanth, K.U. & Johnsingh, A.J.T. (2003) Estimation of large herbivore densities in the tropical forests of southern India using distance sampling. *Journal of Zoology*, **261**, 285–290.

Jhala, Y.V., Qamar, Q., Gopal, R. & Sinha, P.R. (2011a) Status of tigers, co-predators and prey in India, 2010. , National Tiger Conservation Authority, Govt. of India, New Delhi, and Wildlife Institute of India, Dehradun.

Jhala, Y.V., Qureshi, Q. & Gopal, R. (2011b) Can the abundance of tigers be assessed from their signs? *Journal of Applied Ecology*, **48**, 14–24.

Johnson, D.H. (2008) In defense of indices: the case of bird surveys. *The Journal of Wildlife Management*, **72**, 857–868.

Jolly, G.M. (1965) Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika*, **52**, 225–247.

Juliano, S.A. & Williams, F. (1987) A comparison of methods for estimating the functional response parameters of the random predator equation. *The Journal of Animal Ecology*, **56**, 641–653.

Karanth, K.K. (2007) Making resettlement work: The case of India's Bhadra Wildlife Sanctuary. *Biological Conservation*, **139**, 315–324.

Karanth, K.U. (1982) Bhadra wildlife sanctuary and its endangered ecosystem. *Journal of Bombay Natural History*, **79**, 79–86.

Karanth, K.U., Gopalaswamy, A.M., Kumar, N.S., Delampady, M., Nichols, J.D., Seidensticker, J., Noon, B.R. & Pimm, S.L. (2011a) Counting India's Wild Tigers Reliably. *Science*, **332**, 791.

Karanth, K.U., Gopalaswamy, A.M., Kumar, N.S., Vaidyanathan, S., Nichols, J.D. & MacKenzie, D.I. (2011b) Monitoring carnivore populations at the landscape scale: occupancy modelling of tigers from sign surveys. *Journal of Applied Ecology*, **48**, 1048–1056.

Karanth, K.U. & Kumar, N.S. (2005) Distribution and Dynamics of Tiger and Prey Populations in Karnataka, Final Technical Report. , Centre for Wildlife Studies, Bangalore.

Karanth, K.U., Kumar, N.S., Gopalaswamy, A.M. & Srinivas, V. (2008) Distribution and Dynamics of Tiger and Prey Populations in Karnataka, Final Technical Report.

Karanth, K.U. & Nichols, J.D. (1998) Estimation of tiger densities in India using photographic captures and recaptures. *Ecology*, **79**, 2852–2862.

Karanth, K.U. & Nichols, J.D. (2000) Ecological status and conservation of tigers in India. , Centre for Wildlife Studies, Bangalore, India.

Karanth, K.U. & Nichols, J.D. (2002) *Monitoring Tigers and their Prey: A manual for researchers, managers and conservationists in tropical Asia*. Centre for Wildlife Studies, Bangalore.

Karanth, K.U., Nichols, J.D., Kumar, N.S. & Hines, J.E. (2006) Assessing tiger population dynamics using photographic capture-recapture sampling. *Ecology*, **87**, 2925–2937.

Karanth, K.U., Nichols, J.D., Kumar, N.S., Link, W.A. & Hines, J.E. (2004) Tigers and their prey: predicting carnivore densities from prey abundance. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 4854.

Karanth, K.U. & Sunquist, M.E. (1992) Population structure, density and biomass of large herbivores in the tropical forests of Nagarahole, India. *Journal of Tropical Ecology*, **8**, 21–35.

Karanth, K.U., Thomas, L. & Kumar, N.S. (2002) Field surveys: Estimating absolute densities of prey species using line transect sampling. *Monitoring Tigers and their Prey: A manual for researchers, managers and conservationists in tropical Asia* (eds. K.U. Karanth & J.D. Nichols), pp. 111–120. Centre for Wildlife Studies, Bangalore, India.

Karanth, K.U., Gopalaswamy, A.M., Karanth, K.K., Goodrich, J., Seidensticker, J. & Robinson, J.G. (2013a) Sinks as saviors: Why flawed inference cannot assist tiger recovery. *Proceedings of the National Academy of Sciences*, **110**, E110–E110.

Karanth, K.K., Gopalaswamy, A.M., DeFries, R. & Ballal, N. (2012) Assessing patterns of human-wildlife conflicts and compensation around a central Indian protected area. *PloS one*, **7**, e50433.

Karanth, K.K., Gopalaswamy, A.M., Prasad, P.K. & Dasgupta, S. (2013b) Patterns of human–wildlife conflicts and compensation: Insights from Western Ghats protected areas. *Biological Conservation*, **166**, 175–185.

Karanth, K.K., Naughton-Treves, L., DeFries, R. & Gopalaswamy, A.M. (2013c) Living with Wildlife and Mitigating Conflicts Around Three Indian Protected Areas. *Environmental management*, **52**, 1320–1332.

Katzner, T.E., Ivy, J.A.R., Bragin, E.A., Milner-Gulland, E.J. & DeWoody, J.A. (2011) Conservation implications of inaccurate estimation of cryptic population size. *Animal Conservation*, **14**, 328–332.

Keane, A., Jones, J.P.G. & Milner-Gulland, E.J. (2011) Encounter data in resource management and ecology: pitfalls and possibilities. *Journal of Applied Ecology*, **48**, 1164–1173.

Kendall, W.L. & Pollock, K.H. (2001) The robust design in capture-recapture studies: a review and evaluation by Monte Carlo simulation. *Wildlife*, pp. 31–43.

Kendall, W.L., Pollock, K.H. & Brownie, C. (1995) A likelihood-based approach to capture-recapture estimation of demographic parameters under the robust design. *Biometrics*, pp. 293–308.

Kendall, W., Nichols, J. & Hines, J. (1997) Estimating temporary emigration using capture-recapture data with Pollock's robust design. *Ecology*, **78**, 563–578.

Kéry, M., Gardner, B., Stoeckle, T., Weber, D. & Royle, J.A. (2010) Use of Spatial Capture-Recapture Modeling and DNA Data to Estimate Densities of Elusive Animals. *Conservation Biology*, **25**, 356–364.

Kindberg, J., Ericsson, G. & Swenson, J.E. (2009) Monitoring rare or elusive large mammals using effort-corrected voluntary observers. *Biological conservation*, **142**, 159–165.

Knape, J. & de Valpine, P. (2012) Are patterns of density dependence in the Global Population Dynamics Database driven by uncertainty about population abundance? *Ecology letters*, **15**, 17–23.

Kruuk, H. (1978) Spatial organization and territorial behaviour of the European badger Meles meles. *Journal of Zoology*, **184**, 1–19.

Kumar, N.S. (2011) *Assessment of distribution and abundance of ungulate prey using spatial models in Nagarahole and Bandipur Tiger Reserves of India*. Ph.D. thesis, Manipal University.

Kuussaari, M., Bommarco, R., Heikkinen, R.K., Helm, A., Krauss, J., Lindborg, R., Öckinger, E., Pärtel, M., Pino, J., Roda, F. *et al.* (2009) Extinction debt: a challenge for biodiversity conservation. *Trends in ecology & evolution*, **24**, 564–571.

Kvålseth, T.O. (1985) Cautionary note about R 2. *The American Statistician*, **39**, 279–285.

Lachish, S., Gopalaswamy, A.M., Knowles, S.C.L. & Sheldon, B.C. (2012) Site-occupancy modelling as a novel framework for assessing test sensitivity and estimating wildlife disease prevalence from imperfect diagnostic tests. *Methods in Ecology and Evolution*, **3**, 339–348.

Laing, S.E., Buckland, S.T., Burn, R.W., Lambie, D. & Amphlett, A. (2003) Dung and nest surveys: estimating decay rates. *Journal of Applied Ecology*, **40**, 1102–1111.

Lambin, X., Petty, S.J. & Mackinnon, J.L. (2000) Cyclic dynamics in field vole populations and generalist predation. *Journal of animal ecology*, **69**, 106–119.

Lebreton, J.D., Burnham, K.P., Clobert, J. & Anderson, D.R. (1992) Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological monographs*, **62**, 67–118.

Levin, S.A. (1976) Population dynamic models in heterogeneous environments. *Annual review of ecology and systematics*, **7**, 287–310.

Levin, S.A. (1995) *The problem of pattern and scale in ecology*. Springer.

Lindley, D.V. & Smith, A.F.M. (1972) Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**, 1–41.

Lindström, E. (1986) Territory inheritance and the evolution of group-living in carnivores. *Animal Behaviour*, **34**, 1825–1835.

Link, W.A. & Barker, R.J. (2010) *Bayesian inference: with ecological applications*. Academic Press.

Link, W.A. & Nichols, J.D. (1994) On the importance of sampling variance to investigations of temporal variation in animal population size. *Oikos*, **69**, 539–544.

Link, W.A., Yoshizaki, J., Bailey, L.L. & Pollock, K.H. (2010) Uncovering a latent multinomial: analysis of mark–recapture data with misidentification. *Biometrics*, **66**, 178–185.

Long, R.A. & MacKay, P. (2008) *Noninvasive survey methods for carnivores*. Island Pr.

Lukacs, P.M. (2002) WebSim: simulation software to assist in trapping web design. *Wildlife Society Bulletin*, pp. 1259–1261.

Luniak, M. (2004) Synurbization–adaptation of animal wildlife to urban development. *Proc. 4th Int. Symposium Urban Wildl. Conserv. Tucson*, pp. 50–55.

Lunn, D.J., Thomas, A., Best, N. & Spiegelhalter, D. (2000) WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, **10**, 325–337.

MacArthur, R.H. & Pianka, E.R. (1966) On optimal use of a patchy environment. *American Naturalist*, **100**, 603–609.

Macdonald, D.W. (1983) The ecology of carnivore social behaviour. *Nature*, **301**, 379–384.

Macdonald, D.W., ed. (2001) *The new encyclopedia of mammals*. Oxford University Press, Oxford, UK.

Macdonald, D.W. & Johnson, D.D.P. (In Press) Patchwork planet: resource dispersion and the ecology of life. *Journal of Zoology*.

Macdonald, D.W. & Loveridge, A.J. (2010) *The biology and conservation of wild felids*. Oxford Univ Press.

Macdonald, D. & Sillero-Zubiri, C. (2004) *The biology and conservation of wild canids*. Oxford University Press, USA.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. & Langtimm, C.A. (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**, 2248–2255.

MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E. (2006) *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Academic Press.

Madhusudan, M.D. & Mishra, C. (2003) Why big, fierce animals are threatened: conserving large mammals in densely populated landscapes. *Battles over nature: science and the politics of conservation* (eds. V. Saberwal & M. Rangarajan), pp. 31–55. Permanent Black, New Delhi, India.

Maher, C. & Lott, D. (1995) Definitions of territoriality used in the study of variation in vertebrate spacing systems. *Animal Behaviour*, **49**, 1581–1597.

Mandelbrot, B.B. (1983) *The fractal geometry of nature*, vol. 173. Macmillan.

Marques, F.F.C., Buckland, S.T., Goffin, D., Dixon, C.E., Borchers, D.L., Mayle, B.A. & Peace, A.J. (2001) Estimating deer abundance from line transect surveys of dung: sika deer in southern Scotland. *Journal of Applied Ecology*, **38**, 349–363.

Martin, J., Runge, M.C., Nichols, J.D., Lubow, B.C. & Kendall, W.L. (2009) Structured decision making as a conceptual framework to identify thresholds for conservation and management. *Ecological Applications*, **19**, 1079–1090.

May, R. (1999) Unanswered questions in ecology. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **354**, 1951–1959.

May, R.M. (1978) Host-parasitoid systems in patchy environments: a phenomenological model. *The Journal of Animal Ecology*, **47**, 833–844.

McClintock, B.T., Nichols, J.D., Bailey, L.L., MacKenzie, D.I., Kendall, W.L. & Franklin, A.B. (2010) Seeking a second opinion: uncertainty in disease ecology. *Ecology Letters*, **13**, 659–674.

Millar, R.B. (2009) Comparison of hierarchical Bayesian models for overdispersed count data using DIC and Bayes' Factors. *Biometrics*, **65**, 962–969.

Mishra, H.R. (1982) *The ecology and behaviour of chital (Axis axis) in the Royal Chitwan National Park, Nepal.* Ph.D. thesis, University of Edinburg, Edinburgh, UK.

Mishra, H.R. & Wemmer, C.M. (1987) The comparative breeding ecology of four cervids Royal Chitwan National Park. *In Biology and management of the Cervida* (ed. C.M. Wemmer), pp. 259–271. Smithsonian Institution Press, Washington, D.C., USA. Research Symposia of the National Zoological Park.

Moe, S.R. & Wegge, P. (1997) Spacing behaviour and habitat use of axis deer (Axis axis) in lowland Nepal. *Canadian Journal of Zoology*, **72**, 1735–1744.

Møller, J. & Waagepetersen, R.P. (2004) *Statistical inference and simulation for spatial point processes*, vol. 100. CRC Press.

Mondol, S., Karanth, K.U., Kumar, N.S., Gopalaswamy, A.M., Andheria, A. & Ramakrishnan, U. (2009) Evaluation of non-invasive genetic sampling methods for estimating tiger population size. *Biological Conservation*, **142**, 2350–2360.

Mondol, S., Kumar, N.S., Gopalaswamy, A.M., Sunagar, K., Karanth, K.U. & Ramakrishnan, U. (In press) Identifying species, sex and individual tigers and leopards in the Malenad-Mysore Tiger Landscape, Western Ghats, India. *Conservation Genetics Resources*, pp. 1–9.

Nag, K. (2008) Assessing animal abundance using photographic capture data using an occupancy approach.

Nagelkerke, N.J. (1991) A note on a general definition of the coefficient of determination. *Biometrika*, **78**, 691–692.

Nchanji, A.C. & Plumptre, A.J. (2001) Seasonality in elephant dung decay and implications for censusing and population monitoring in south-western Cameroon. *African Journal of Ecology*, **39**, 24–32.

Nichols, J.D., Hines, J.E., Sauer, J.R., Fallon, F.W., Fallon, J.E. & Heglund, P. (2000) A double-observer approach for estimating detection probability and abundance from point counts. *Auk*, **117**, 393–408.

Novaro, A.J., Funes, M.C. & Walker, R.S. (2005) An empirical test of source–sink dynamics induced by hunting. *Journal of Applied Ecology*, **42**, 910–920.

Nowak, M.A. (2006) Five rules for the evolution of cooperation. *science*, **314**, 1560–1563.

O'Connell, A.F., Nichols, J.D. & Karanth, K.U., eds. (2010a) *Camera traps in Ecology*. Springer, Tokyo, Japan.

O'Connell, A.F., Nichols, J.D. & Karanth, U.K. (2010b) *Camera traps in animal ecology: Methods and analyses*. Springer.

Odden, M. & Wegge, P. (2007) Predicting spacing behavior and mating systems of solitary cervids: a study of hog deer and Indian muntjac. *Zoology*, **110**, 261–270.

O'Hara, R.B. & Kotze, D.J. (2010) Do not log-transform count data. *Methods in Ecology and Evolution*, **1**, 118–122.

Oksanen, L. & Oksanen, T. (1992) Long-term microtine dynamics in north Fennoscandian tundra: the vole cycle and the lemming chaos. *Ecography*, **15**, 226–236.

Otis, D.L., Burnham, K.P., White, G.C. & Anderson, D.R. (1978) Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, **62**, 3–135.

Parmenter, R.R., Yates, T.L., Anderson, D.R., Burnham, K.P., Dunnum, J.L., Franklin, A.B., Friggens, M.T., Lubow, B.C., Miller, M., Olson, G.S. *et al.* (2003) Small-mammal density estimation: a field comparison of grid-based vs. web-based density estimators. *Ecological monographs*, **73**, 1–26.

Patil, N., Kumar, N., Gopalaswamy, A.M. & Karanth, K.U. (2011) Dispersing tiger makes a point... *Oryx*, **45**, 472–472.

Pielou, E.C. (1969) *An introduction to mathematical ecology*. Wiley-Interscience.

Piersma, T. & Drent, J. (2003) Phenotypic flexibility and the evolution of organismal design. *Trends in Ecology & Evolution*, **18**, 228–233.

Pledger, S. (2000) Unified maximum likelihood estimates for closed capture–recapture models using mixtures. *Biometrics*, **56**, 434–442.

Plumptre, A.J. & Harris, S. (1995) Estimating the biomass of large mammalian herbivores in a tropical montane forest - a method of faecal counting that avoids assuming a steadystate system. *Journal of Applied Ecology*, **32**, 111–120.

Pollock, K.H., Nichols, J.D., Brownie, C. & Hines, J.E. (1990) Statistical inference for capture-recapture experiments. *Wildlife Monographs*, pp. 3–97.

Pollock, K.H., Nichols, J.D., Simons, T.R., Farnsworth, G.L., Bailey, L.L. & Sauer, J.R. (2002) Large scale wildlife monitoring studies: statistical methods for design and analysis. *Environmetrics*, **13**, 105–119.

Pollock, K. (1982) A capture-recapture design robust to unequal probability of capture. *The Journal of Wildlife Management*, **46**, 752–757.

Pradel, R., Hines, J.E., Lebreton, J.D. & Nichols, J.D. (1997) Capture-recapture survival models taking account of transients. *Biometrics*, pp. 60–72.

Price, A.J. & Rachlow, J.L. (2011) Development of an index of abundance for pygmy rabbit populations. *The Journal of Wildlife Management*, **75**, 929–937.

Pulliam, H.R. (1988) Sources, sinks, and population regulation. *American naturalist*, **132**, 652–661.

R Development Core Team (2009) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [WWW document]. URL http://www.r-project.org/.

Rainey, P.B. & Rainey, K. (2003) Evolution of cooperation and conflict in experimental bacterial populations. *Nature*, **425**, 72–74.

Rayar, K. (2010) *Assessing potential tiger habitats in Cauvery Wildlife Sanctuary, Karnataka, using occupancy modeling approaches*. Master's thesis, National Centre for Biological Sciences, Bangalore, India.

Redford, K.H. (1992) The empty forest. *BioScience*, **42**, 412–422.

Reich, B.J. & Gardner, B. (2014) A spatial capture-recapture model for territorial species. *arXiv preprint arXiv:1405.1976*.

Revilla, E. (2003) What does the Resource Dispersion Hypothesis explain, if anything? *Oikos*, **101**, 428–432.

Roughgarden, J. (1975) A simple model for population dynamics in stochastic environments. *American Naturalist*, pp. 713–736.

Rovero, F. & Marshall, A.R. (2009) Camera trapping photographic rate as an index of density in forest ungulates. *Journal of Applied Ecology*, **46**, 1011–1017.

Rowcliffe, J.M., Carbone, C., Jansen, P.A., Kays, R. & Kranstauber, B. (2011) Quantifying the sensitivity of camera traps: an adapted distance sampling approach. *Methods in Ecology and Evolution*, **2**, 464–476.

Rowcliffe, J.M., Field, J., Turvey, S.T. & Carbone, C. (2008) Estimating animal density using camera traps without the need for individual recognition. *Journal of Applied Ecology*, **45**, 1228–1236.

Royle, J.A. (2006) Site occupancy models with heterogeneous detection probabilities. *Biometrics*, **62**, 97–102.

Royle, J.A. & Dorazio, R.M. (2008) *Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities*. Academic Press, San Diego, CA, USA.

Royle, J.A., Dorazio, R.M. & Link, W.A. (2007) Analysis of multinomial models with unknown index using data augmentation. *Journal of Computational and Graphical Statistics*, **16**, 67–85.

Royle, J.A., Gardner, B., O'Connell, A.F., Nichols, J.D. & Karanth, K.U. (2011a) Hierarchical spatial capture-recapture models for estimating density from trapping arrays. *Camera traps in animal ecology: methods and analyses.*, pp. 163–190.

Royle, J.A., Karanth, K.U., Gopalaswamy, A.M. & Kumar, N.S. (2009a) Bayesian inference in camera trapping studies for a class of spatial capture-recapture models. *Ecology*, **90**, 3233–3244.

Royle, J.A., Kéry, M. & Guélat, J. (2011b) Spatial capture-recapture models for search-encounter data. *Methods in Ecology and Evolution*, **2**, 602–611.

Royle, J.A. & Nichols, J.D. (2003) Estimating abundance from repeated presence-absence data or point counts. *Ecology*, **84**, 777–790.

Royle, J.A., Nichols, J.D., Karanth, K.U. & Gopalaswamy, A.M. (2009b) A hierarchical model for estimating density in camera-trap studies. *Journal of Applied Ecology*, **46**, 118–127.

Royle, J.A. & Young, K.V. (2008) A hierarchical model for spatial capture-recapture data. *Ecology*, **89**, 2281–2289.

Royle, J.A., Chandler, R.B., Gazenski, K.D. & Graves, T.A. (2013a) Spatial capture-recapture models for jointly estimating population density and landscape connectivity. *Ecology*, **94**, 287–294.

Royle, J.A., Chandler, R.B., Sollmann, R. & Gardner, B. (2013b) *Spatial capture-recapture*. Academic Press.

Royle, J.A., Chandler, R.B., Sun, C.C. & Fuller, A.K. (2013c) Integrating resource selection information with spatial capture–recapture. *Methods in Ecology and Evolution*, **4**, 520–530.

Sandell, M. (1989) The mating tactics and spacing patterns of solitary carnivores. *Carnivore behavior, ecology, and evolution*, pp. 164–182. Springer.

Sankar, K. & Acharya, B. (2004) Sambar (Cervus unicolor). *Ungulates of Peninsular India* (eds. K. Sankar & S. Goyal), vol. 7, pp. 163–170. Wildlife Institute of India, Dehra Dun, India.

Sankar, K., Qureshi, Q., Pasha, M.K.S. & Areendran, G. (2000) Ecology of gaur *Bos gaurus* in Pench Tiger Reserve, Madhya Pradesh.

Schaller, G.B. (1967) *The deer and the tiger*. University of Chicago Press, Chicago, USA.

Schaub, M., Gimenez, O., Sierro, A. & Arlettaz, R.E.L. (2007) Use of integrated modeling to enhance estimates of population dynamics obtained from limited data. *Conservation Biology*, **21**, 945–955.

Scheiner, S.M. (1993) Genetics and evolution of phenotypic plasticity. *Annual review of ecology and systematics*, pp. 35–68.

Schipper, J., Chanson, J.S., Chiozza, F., Cox, N.A., Hoffmann, M., Katariya, V., Lamoreux, J., Rodrigues, A.S.L., Stuart, S.N., Temple, H.J. *et al.* (2008) The status of the world's land and marine mammals: diversity, threat, and knowledge. *Science*, **322**, 225.

Seber, G.A.F. (1965) A note on the multiple-recapture census. *Biometrika*, **52**, 249.

Sharma, K., Wright, B., Joseph, T. & Desai, N. (2014) Tiger poaching and trafficking in India: Estimating rates of occurrence and detection over four decades. *Biological Conservation*, **179**, 33–39.

Sharma, R.K., Jhala, Y.V., Qureshi, Q., Vattakaven, J., Gopal, R. & Nayak, K. (2010) Evaluating capture–recapture population and density estimation of tigers in a population with known parameters. *Animal Conservation*, **13**, 94–103.

Signor, P.W. & Lipps, J.H. (1982) Sampling bias, gradual extinction patterns and catastrophes in the fossil record. *Geological Society of America Special Paper*, **190**, 291–296.

Sillero-Zubiri, C. & Macdonald, D.W. (1998) Scent-marking and territorial behaviour of Ethiopian wolves Canis simensis. *Journal of Zoology*, **245**, 351–361.

Singer, F.J. (1981) Wild pig populations in the national parks. *Environmental Management*, **5**, 263–270.

Singh, P., Gopalaswamy, A.M. & Karanth, K.U. (2010) Factors influencing densities of striped hyenas (Hyaena hyaena) in arid regions of India. *Journal of Mammalogy*, **91**, 1152–1159.

Smith, J.L.D., McDougal, C. & Miquelle, D. (1989) Scent marking in free-ranging tigers, *Panthera tigris*. *Animal Behaviour*, **37**, 1–10.

Soisalo, M.K. & Cavalcanti, S. (2006) Estimating the density of a jaguar population in the Brazilian Pantanal using camera-traps and capture-recapture sampling in combination with GPS radio-telemetry. *Biological Conservation*, **129**, 487–496.

Sollmann, R., Furtado, M.M., Gardner, B., Hofer, H., Jácomo, A.T., Tôrres, N.M. & Silveira, L. (2011) Improving density estimates for elusive carnivores: accounting for sex-specific detection and movements using spatial capture–recapture models for jaguars in central Brazil. *Biological Conservation*, **144**, 1017–1024.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & Van der Linde, A. (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **64**, 583–639.

Spiess, A.N. & Neumeyer, N. (2010) An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. *BMC pharmacology*, **10**, 6.

Stephens, P.A., Zaumyslova, O.Y., Miquelle, D.G., Myslenkov, A.I. & Hayward, G.D. (2006) Estimating population density from indirect sign: track counts and the Formozov–Malyshev–Pereleshin formula. *Animal Conservation*, **9**, 339–348.

Stewart, G. (2010) Meta-analysis in applied ecology. *Biology letters*, **6**, 78.

Sunquist, M.E. & Sunquist, F. (2002) *Wild cats of the world*. University of Chicago Press.

Tarver, J.E., Braddy, S.J. & Benton, M.J. (2007) The effects of sampling bias on Palaeozoic faunas and implications for macroevolutionary studies. *Palaeontology*, **50**, 177–184.

Thomas, L., Buckland, S.T., Rexstad, E.A., Laake, J.L., Strindberg, S., Hedley, S.L., Bishop, J.R.B. & Marques, T.A. (2010) Distance software: design and analysis of distance sampling surveys for estimating population size. *Journal of Applied Ecology*, **47**, 5–14.

Thomas, L. & Karanth, K.U. (2002) Statistical concepts: Estimating absolute densities of prey species using line transect sampling. *Monitoring Tigers and their Prey: A manual for researchers, managers and conservationists in tropical Asia* (eds. K.U. Karanth & J.D. Nichols), pp. 87–110. Centre for Wildlife Studies, Bangalore, India.

Thompson, S.K. (2002) *Sampling*. Wiley, New York, USA, 2nd edn.

Thompson, W.L. (2004) *Sampling rare or elusive species: concepts, designs, and techniques for estimating population parameters*. Island Press.

Tierney, L. (1994) Markov chains for exploring posterior distributions. *The Annals of Statistics*, **22**, 1701–1728.

Tilman, D., May, R.M., Lehman, C.L. & Nowak, M.A. (1994) Habitat destruction and the extinction debt. *Nature*, **371**, 65–66.

de Valpine, P. & Hastings, A. (2002) Fitting population models incorporating process noise and observation error. *Ecological Monographs*, **72**, 57–76.

Van Horne, B. (1983) Density as a misleading indicator of habitat quality. *The Journal of Wildlife Management*, **47**, 893–901.

Venables, W.N. & Ripley, B.D. (2002) *Modern applied statistics with S*. Springer verlag.

Via, S., Gomulkiewicz, R., De Jong, G., Scheiner, S.M., Schlichting, C.D. & Van Tienderen, P.H. (1995) Adaptive phenotypic plasticity: consensus and controversy. *Trends in Ecology & Evolution*, **10**, 212–217.

Walston, J., Robinson, J., Bennett, E., Breitenmoser, U., da Fonseca, G., Goodrich, J., Gumal, M., Hunter, L., Johnson, A., Karanth, K. *et al.* (2010) Bringing the tiger back from the brink – The six-percent solution. *PLoS Biology*, **8**, e1000485.

Watkinson, A.R. & Sutherland, W.J. (1995) Sources, sinks and pseudo-sinks. *Journal of Animal Ecology*, **64**, 126–130.

White, G.C. & Burnham, K.P. (1999) Program MARK: survival estimation from populations of marked animals. *Bird Study*, **46**, 120–139.

White, G.C., Burnham, K.P., Otis, D.L. & Anderson, D.R. (1978) *Users Manual for Program CAPTURE*. Utah State University Press, Logan, Utah.

Wiens, J.A. (1989) Spatial scaling in ecology. *Functional ecology*, **3**, 385–397.

Willett, J.B. & Singer, J.D. (1988) Another cautionary note about R 2: Its use in weighted least-squares regression analysis. *The American Statistician*, **42**, 236–238.

Williams, B.K., Nichols, J.D. & Conroy, M.J. (2002) *Analysis and management of animal populations: modeling, estimation, and decision making*. Academic Press, San Diego, CA, USA.

Wilson, K.R. & Anderson, D.R. (1985) Evaluation of two density estimators of small mammal population size. *Journal of Mammalogy*, **66**, 13–21.

Wolff, J.O. (1980) The role of habitat patchiness in the population dynamics of snowshoe hares. *Ecological Monographs*, pp. 111–130.

Wolff, J.O. (1997) Population regulation in mammals: an evolutionary perspective. *Journal of Animal Ecology*, pp. 1–13.

Woodroffe, R. & Ginsberg, J.R. (1998) Edge effects and the extinction of populations inside protected areas. *Science*, **280**, 2126–2128.