Using the *Penguin Watch* network to examine phenology and breeding success in *Pygoscelis* penguins



By

Fiona M. Jones

Department of Zoology, University of Oxford

Green Templeton College

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2019



Gentoo penguins (*Pygoscelis papua*) with chicks, Booth Island, Antarctic Peninsula (65.05°S, 63.96°W), on 16th January 2017. Photo: F. M. Jones.

Using the *Penguin Watch* network to examine phenology and breeding success in *Pygoscelis* penguins

Fiona M. Jones

Green Templeton College

A thesis submitted for the degree of Doctor of Philosophy

Trinity, 2019

Abstract

Remote time-lapse cameras can facilitate high-frequency monitoring of wild animal populations over large spatial and temporal scales, yet the volume of imagery they produce can be overwhelming for small research teams. As ecological studies transition towards a 'big data' approach, we must develop ways to deal with the 'data deluge', in the form of integrated processing frameworks. In this thesis, I present and examine data-processing pipelines for the *Penguin Watch* project – a network of approximately 80 time-lapse cameras positioned in penguin colonies across the Scotia Arc region of the Southern Ocean. I show that both the *Penguin Watch* online citizen science project and the *Pengbot* computer vision algorithm are efficient and reliable image annotation tools, which can be used in combination to maximise output and minimise error. I present new methodologies to extract biologically meaningful metrics from citizen science annotations, and show that spatial data – for example 'nearest neighbour distances' – can be used to examine behaviour, such as chick huddling.

Seabirds, including penguin species, are often considered indicators of marine health. As such, effective monitoring of their populations offers an invaluable opportunity to assess the health of the wider marine ecosystem. An essential part of this process is obtaining baseline data, and here I use the *Penguin Watch* project to extract phenological dates for a number of new Chinstrap penguin (*Pygoscelis antarcticus*) colonies and seasons – a species which has historically been understudied. Through a 'virtual' mark-recapture study I examine breeding success in Gentoo penguin (*P. papua*) colonies across a substantial part of their range in the Atlantic sector of the Southern Ocean. Investigation of the impact of three key anthropogenic stressors (precipitation, the krill fishery, and tourism) on breeding success revealed no significant effects over the study period. However, there is evidence that the relationship between chick survival and precipitation could be highly non-linear, with severe precipitation events associated with high chick mortality. This has implications for a number of Southern Ocean bird species, and with precipitation levels expected to increase under climate-change scenarios, continued monitoring is essential. Repetition of this study on Adélie (*P. adeliae*) and Chinstrap penguins – species which have a higher dependency on krill than Gentoos – is a key future research priority.

Contents

Acknowledgements……1

Chapter 1: Introduction……5

Chapter 2: Time-lapse imagery and volunteer classifications from the Zooniverse *Penguin Watch* project……63

Chapter 3: Processing citizen science- and machine-annotated time-lapse imagery for biologically meaningful metrics……94

Chapter 4: Using citizen science to examine phenology and behaviour: new insights from the *Penguin Watch* project……137

Chapter 5: Large-scale assessment of intra- and inter-annual nesting success using a remote camera network……181

Chapter 6: Discussion and Conclusions……217

Chapter 7: References……237

Appendix 1: Time-lapse imagery as a low-cost monitoring option for colonial species..271

Appendix 2: *Penguin Watch* Cameras……300

Appendix 3: Fieldwork Permit Details……304

Appendix 4: Chapter 2 – Metadata Tables……306

Appendix 5: Chapter 3 – Metadata Tables……313

Appendix 6a: Chapter 4 – Huddle Scores……320

Appendix 6b: Chapter 4 – Assessing Temperature as a Driver of Huddling Behaviour – Temperature Data and Analysis……321

Appendix 6c: Chapter 4 – Investigating the Crèche Phase using Citizen Science Data....325

Appendix 7a: Chapter 5 – Supplementary Materials – Part 1 (S1)……328

Appendix 7b: Chapter 5 – Supplementary Materials – Part 2 (S2)……355

Appendix 8: *Penguin Watch* as a Tool for Public Engagement……359

Acknowledgements

I wish to express my sincere thanks to my primary supervisor, Dr Tom Hart, for his support, advice and kindness throughout my DPhil, and of course for initiating the *Penguin Watch* project. Not only has Tom helped me to develop my academic research skills, but he has turned someone who had never even been on a Scout camp into a competent polar fieldworker, and I cannot thank him enough for the unforgettable and awe-inspiring experience of working in Antarctica. I also wish to thank my secondary supervisor, Professor Chris Lintott, for his continued support and enthusiasm, and for a fascinating insight into the world of citizen science. I thank him also for his advice and encouragement regarding my interest in science communication, and for providing me with fantastic opportunities to hone my outreach skills.

I would also like to thank members of the Zooniverse team – both in Oxford and further afield – for their friendly advice and assistance throughout my DPhil. Before working on *Penguin Watch*, I had barely written a line of code, and their willingness to offer support (and their patience!) never ceases to amaze me. Special thanks go to Campbell Allen, Dr Grant Miller, Dr Coleman Krawczyk, Shaun Noordin, Dr Brooke Simmons, Dr Greg Hines, Dr Darryl Wright, Amy Boyer, and Adam McMaster, although many others have helped along the way.

I also wish to thank my fellow members of the *Penguin Watch* team – Ignacio (Nacho) Martinez Juarez for his enthusiasm, kind words of encouragement, and research chats (not to mention his patient Spanish language tuition prior to fieldwork!), and Laure Cugnière, our project coordinator, for her incredible efficiency, hard-work and dedication to the

project. I also wish to thank Dr Caitlin Black for her help and advice, and Dr Gemma Clucas and Hila Levy Williams for their fieldwork tips.

I give my heartfelt thanks to all of my research collaborators, without whom this DPhil would not have been possible. In addition to those already mentioned, these are: Dr Casey Youngflesh, Dr Heather Lynch, Dr Carlos Arteta, Professor Andrew Zisserman, Dr Victor Lempitsky, Victoria Hope, Dr Louise Emmerson, Dr Colin Southwell, Dr Robin Freeman, Dr Robert Simpson, Alasdair Davies, Dr Eamonn Maguire, Professor Peter Barham, Dr Mark Jessopp, Joan Arthur, Holly Torsey and Zuzi Ročkaiová (nee Macháčková). I have thoroughly enjoyed working with so many interesting and inspiring scientists, across such an amazing diversity of disciplines. Further thanks go to the *Penguin Watch* moderators (Joan, Holly and Zuzi), for their infectious enthusiasm, relentless dedication, and willingness to go above-and-beyond to help both me and the project.

Special thanks go to Quark Expeditions, Dr Andrea Raya Rey and Ana Craven, who have provided invaluable logistical support for each of my fieldwork trips, and everybody I have met and worked with in Antarctica. I am extremely grateful to have worked with such a knowledgeable and experienced team, with such an obvious passion for Antarctica and its wildlife. I thank them for their warm welcome and friendliness, in addition to their practical advice, help and support, both to me personally and to *Penguin Watch*. There are simply too many people to name, so I have included a team photograph from my first expedition aboard the MS Ocean Endeavour, which shows just some of the incredible people I have worked with over my three years' fieldwork. I also wish to thank all of those who have provided training for fieldwork, including the staff of Warsash Maritime Academy and Marine Matters, and Dr Harvey Pynn and Dr Craig Miller from Wilderness

Medical Training. Additionally I would like to thank NERC for funding my DPhil, including essential training opportunities.

It almost goes without saying that I am extremely grateful to the 50,000+ *Penguin Watch* citizen scientists who make our research possible. I am particularly thankful to all of the school teachers who have chosen to use *Penguin Watch* as an educational tool, and hope it continues to be used for this purpose into the future. I am also thankful to Kirsty Heber-Smith, Dr Michaela Livingstone-Banks, Carly Smith-Huggins and all others who have organised outreach initiatives that allow us to share our research with the general public, such as via Oxford Sparks, Curiosity Carnival, Super Science Saturdays, Pint of Science and other public lectures and school visits.

I would like to thank all of my friends for their kind interest in my research, and their encouragement and support along the way. In particular I would like to thank my friends from Surrey, Oxford, Kinecroft Academy of Dance and the Oxford Lindyhoppers. I would also like to thank those at St Mary's Church for their friendliness and warmth, and words cannot express how grateful I am to God for the opportunities given to me, and the people who I get to share life with.

Last, but by no means least, I would like to thank my family for their endless love, support and encouragement. In addition to their enthusiasm and interest in my work, my parents have been there for me every step of the way, reading my papers, helping me prepare for fieldwork, taking me to the airport, and even letting me store large amounts of polar fieldwork kit in their house…I thank my brothers, Christopher and Timothy, and my sisters-in-law, Sarah and Kay, for all their interest and encouragement, and would like to thank Christopher for sharing his knowledge of photography, and for the camera which

accompanied me on each of my fieldwork trips. Finally, I would like to thank my boyfriend Martin, for his never-ending enthusiasm in my endeavours, helpful PhD advice, support when things haven't gone to plan, and for listening to me talk about penguins for hours on end.



The Expedition Team (Quark Expeditions) aboard the MS Ocean Endeavour in December 2016.

Chapter 1: Introduction

i) Penguins – sentinels of change

Globally, many ecosystems are under threat from anthropogenic stressors (IUCN-CEM, 2016). In some cases, environmental degradation can be attributed to direct human intervention, a key example being the aridification of the Aral Sea (classified by the International Union for Conservation of Nature (IUCN) as a 'collapsed' ecosystem) through increasingly intensive irrigation practices (Keith, 2013). In other cases, the threats are indirect – perhaps the most obvious being consequences of anthropogenic climate change (e.g. ocean warming and acidification leading to coral reef loss (Bland et al., 2017) or decreased precipitation levels threatening cloud forest habitat (Auld & Leishman, 2015)). Often, an ecosystem is subject to multiple threats simultaneously; when these act synergistically, it can be difficult to discern which is the primary stressor, and thus where mitigation priorities lie. Gauging the health of an ecosystem, particularly in the absence of robust baseline data, and disentangling the effects of the myriad threats facing it, remains a significant challenge in conservation biology.

Certain taxa can be considered indicators of ecosystem health, because of their close relationship with, or dependency on, different facets of their biotic and abiotic environment. Owing to their longevity, wide geographical distribution, and complex suite of interactions with the marine environment, seabirds are often considered indicators of marine ecosystem health (Croxall et al., 2012; Furness & Camphuysen, 1997; Paleczny et al., 2015; Parsons et al., 2008). They are capable of integrating long-term and large-scale signals, and interact with different aspects of the coastal or marine environment during

their yearly cycle and lifetime (Paleczny et al., 2015). Hence, improved monitoring of seabird populations, leading to increased understanding of the mechanisms underlying any observed changes, should in turn lead to an enhanced understanding of the health of the marine environment.

Paleczny et al. (2015) estimate a 69.7% decline in the global monitored seabird population between 1950 and 2010. Moreover, the pace of change seems to be increasing, with more rapid declines likely coinciding with new and intensifying threats (Croxall et al., 2012). During the breeding period, seabirds are exposed to land-based threats, such as invasive species or human disturbance – through tourism (Anderson & Keith, 1980), infrastructure development and habitat removal (Schippers et al., 2009), or exploitation (e.g. hunting or egg harvesting) (Boersma et al., 2002; Croxall et al., 2012; Paleczny et al., 2015). While at sea, the primary threats include conflict with fisheries (through competition for fish stocks, and entanglement in fishing equipment – termed 'bycatch' (Anderson et al., 2011; Gianuca et al., 2017; Montevecchi, 2002)), pollution (including oil spills, plastics and bioaccumulating persistent organic pollutants (POPs) (Millow et al., 2015; Votier et al., 2005; Walker, 1990; Wilcox et al., 2015)), and an increasing frequency of severe weather events (Descamps et al., 2015; Frederiksen et al., 2008).

Amongst the most threatened seabirds are the penguins (Family: Spheniscidae), with ten of 18 species classified as either 'vulnerable' or 'endangered' by the IUCN (IUCN, 2018; Trathan et al., 2015). Penguins spend much of their life in water, and are highly adapted to this environment as specialised swimmers and divers (Boersma, 2008; Green et al., 2003; Li et al., 2014; Wienecke et al., 2007; Wilson & Zimmer, 2004). However, they are also fully dependent on a terrestrial or sea ice environment to breed (Davis & Darby, 1990).

As such, they are sensitive to changes across a wide range of habitat types, and sample the oceanic environment to a greater extent than non-diving seabirds (Boersma, 2008). Furthermore, behavioural plasticity allows them to react quickly to changes in resource availability, as discussed by Pichegru et al. (2010), who examine the effect of a new fishing 'no-take zone' on a population of African penguins. Penguins have thus been described as "marine sentinels" (Boersma, 2008), and understanding the way in which their populations are changing – in terms of distribution, demographics or phenology – is likely to shed light on changes occurring at the ecosystem or even oceanic scale.

In the following introductory sections I provide a brief overview of the Spheniscidae family (section *ii*), before introducing the *Pygoscelis* genus (section *iii*) – which comprises the focal species of this thesis – and outlining the key threats to Southern Ocean penguins (section *iv*). I continue by discussing the various methods and technologies which have been used to monitor these penguins in the past (section *v*), and present remote sensing as an ecological monitoring tool with great potential for the future (*vi*). Remote sensing devices often produce large quantities of image data, and in section *vii* I discuss the methods available to process them, with an emphasis on citizen science. I explore the history of the Antarctic Treaty System in section *viii*, and conclude by laying out the aims of this thesis in section *ix*. An ethics statement, which includes details of fieldwork permits, is given in section *x*.

ii) *An introduction to the family Spheniscidae*

> *"Behind the feathered façade of the archetypal tuxedoed penguin lies a remarkable biological entity. Stripped bare of its romantic and anthropomorphic suit, the penguin begs scientific scrutiny; it is like no other bird"*
>
> *(Lloyd S. Davis & John T. Darby)*

In the mind's eye, penguins (birds of the family Spheniscidae) are often given a backdrop of snow and ice, perhaps marching great distances to breed, or preciously guarding a single egg. In reality, penguins exhibit a wide diversity of life histories, and only four of the 18 recognised species reside on, or close to, the Antarctic continent. While confined to the Southern Hemisphere, penguins are distributed as far north as the Equator (the Galápagos penguin (*Spheniscus mendiculus*)), and certain species – in contrast to their stereotype – can be found nesting in coastal burrows or temperate rainforest (e.g. the Little penguin (also known as the Blue or Fairy penguin (*Eudyptula minor*)) and the Fiordland Crested penguin (*Eudyptes pachyrhynchus*), respectively (Miyazaki & Waas, 2003; Otley et al., 2017)).

Although penguins vary markedly in size (from the Little penguin, which measures ~33cm in height and weighs only 1.1-1.2Kg, to the Emperor (*Aptenodytes forsteri*; Fig. 1), which can reach 1m in height and weigh up to 40Kg (Le Maho, 1977; Overeem et al., 2008)), their body plan varies little between species (Cruz, 2007). Indeed, penguins spend much of their lives at sea, and their morphology is extremely well adapted to the marine environment (Cruz, 2007; Davis & Darby, 1990). Penguins are flightless and are specialised divers – traits which are reflected in their skeletal composition and structure (Cruz, 2007; Louw, 1992).



Figure 1: Emperor penguin (*Aptenodytes forsteri*) with chick; Snow Hill Island, Antarctic Peninsula (64°28'S, 57°12'W). Emperors are the largest and heaviest penguin species, and are the only species to breed during the austral winter, on sea ice. Photo: F. M. Jones.

The majority of birds have hollow bones, which they have evolved coincident with flight (O'Connor & Claessens, 2005; Sullivan et al., 2017). In contrast, penguins possess solid bones which are absent of pneumatic features, alongside an evolutionary loss of function (Cruz, 2007). Furthermore, the tarsometatarsus (a bone of the hind limb) is shorter than that seen in flying taxa, and the 'wing' is a specialised flipper (Cruz, 2007). The flipper is dorso-ventrally flattened and paddle-like, and is rotated along its long axis – powered by the pectoral muscles – to propel the bird through the water (Cruz, 2007; Louw, 1992). The flipper also has thermoregulatory functionality: its blood vessels are comparatively small, and the limb is maintained at a lower temperature than the core body, restricting loss of heat (Louw, 1992). A larger surface area allows greater heat dissipation, hence species

living in temperate environments have broader flippers than their Antarctic relatives (Louw, 1992). These anatomical features, alongside physiological and behavioural adaptations, allow penguins to move efficiently through water, making them excellent marine predators. Diet varies both within and between species, but broadly comprises crustaceans, cephalopods and fish (Cherel et al., 1993; Thomas et al., 2013). For instance, Adélie penguins (*Pygoscelis adeliae*) consume large amounts of krill (Euphausiaceae), whereas the Little, Fiordland Crested and Humboldt (*Spheniscus humboldti*) penguins mainly feed on teleost fish or cephalopods (Flemming et al., 2013; Lynnes et al., 2004; Thomas et al., 2013). An example of temporal (seasonal) variability in diet comes from the King penguin (*Aptenodytes patagonicus*) colony at the Crozet Islands in the Southern Indian Ocean. Cherel et al. (1993) found that three species of lanternfish (family: Myctophidae) form the main component of diet during the chick-rearing period in the spring and autumn, whereas the squid species *Onykia ingens* (family: Onychotheuthidae) is a key part of chick diet in winter. Furthermore, an alteration in parents' foraging area was also detected, with neritic species (including the benthopelagic *O. ingens*) being the primary prey during winter, at least to provision chicks, and more distant waters fished during spring and autumn (Cherel et al., 1993). Studies suggest that there is also spatial variability in diet, with Hindell (1988) finding cephalopods to be a minor component of King penguin diet at Macquarie Island, in the Southern Pacific Ocean. A more extreme example of this spatial variability has been reported by Lescroël et al. (2004), who found differences in the diet of Gentoo penguins (*Pygoscelis papua*) within the Kerguelen Islands archipelago.

The key stages in the penguin annual cycle are shown in Fig. 2. While these are relatively conserved between species, there are marked differences in phenological timings, characteristic breeding behaviours and migration distances, even amongst penguins of the same genus. Some of these key inter-species differences are briefly discussed here.

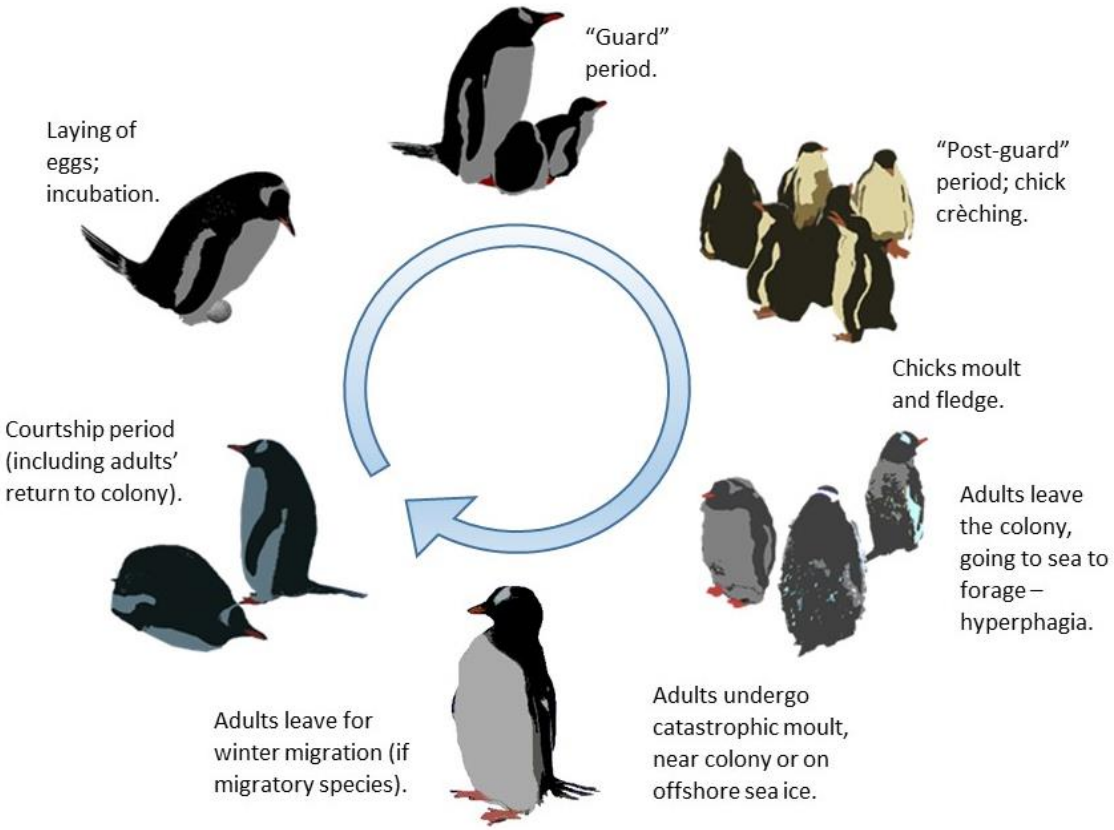


Figure 2: Stages in the penguin annual cycle. Traits such as clutch size and winter migration distance differ between species (and in Kings (*Aptenodytes patagonicus*), the cycle is not completed every year – see text).

A prime example comes from the *Aptenodytes*, where the breeding biology of the Emperor penguin differs greatly from that of the King penguin (Le Maho, 1977). The Emperor penguin is the only species to breed in the Antarctic winter, on sea ice, with females laying

a single egg in early May and chicks fledging in December upon break-up of the sea ice (Le Maho, 1977). King penguins also raise a single chick, but begin their breeding cycle in early spring, with eggs laid in November or December (Olsson, 1996). Chicks over-winter at the colony, leading to a long (approximately 14 month) breeding cycle. This in turn leads to bimodality in breeding activities at the population scale, since parents who successfully raise a chick one year are delayed in their (likely unsuccessful) breeding attempt the following year (Fig. 3; Olsson, 1996).



Figure 3: King penguin (*Aptenodytes patagonicus*) colony at Salisbury Plain, South Georgia (54°3'S, 37°21'W), 6th January 2017. King penguins have a seasonally bimodal breeding strategy, with each full cycle taking approximately 14 months. In this photograph, some penguins are on eggs (bottom right), while crèches of the previous year's chicks can also be seen. Photo: F. M. Jones.

As in other bird species, penguins incubate their eggs against a specialised 'brood patch' – a featherless area of abdominal skin which becomes dilated at the beginning of the incubation period (Le Maho, 1977; Massaro et al., 2006). The *Aptenodytes* species rest their one egg between their feet and body, where a fold of skin forms a pouch (Le Maho, 1977). In contrast, all other species lay and incubate their eggs in nests, whether these be inside burrows (e.g. the Magellanic (*Spheniscus magellanicus*) and Little penguins), within holes in rocks or tree roots (e.g. the Fiordland crested penguin), or constructed above ground out of available materials (e.g. the Gentoo penguin, which uses mud and vegetation to build nests on sub-Antarctic Islands, and small stones on the Antarctic Peninsula (Bost & Jouventin, 1990)). With the exception of the *Aptenodytes*, penguins usually lay a clutch of two eggs (Gwynn, 1993), although some species (for example the African (*Spheniscus demersus*) and Gentoo penguins) can lay replacement clutches later in the season (Bost & Jouventin, 1990; Davis & Renner, 2003; Williams, 1990a). Furthermore, the crested penguins (*Eudyptes* species) lay two eggs – referred to as the A and B egg – which vary markedly in size (Massaro & Davis, 2005; Poisbleau et al., 2008; Williams, 1990a). The A egg, laid first, is significantly smaller than the B egg, and in Macaroni penguins (*Eudyptes chrysolophus*), where the size discrepancy is most pronounced, roughly half of A eggs are lost before the B egg hatches (Williams, 1990a).

Another example of inter-species variation comes during the post-guard (or 'crèche') phase, which occurs once adults leave their young unattended for the first time (Davis, 1982). In some species, for example the Adélie, Chinstrap (*Pygoscelis antarcticus*), Gentoo, and King (see Fig. 3) penguins, chicks form tight aggregations during this period, known as huddles or crèches (Black et al., 2016). Huddling has thermoregulatory benefits (famously associated with the behaviour in Emperor penguin adults (Ancel et al., 2015; Le Maho,

1977)), and also provides protection against both predators and aggressive adults (Black et al., 2016; Gilbert et al., 2006; Le Maho, 1977). As such, the degree of huddling will depend on predation risk (Wilson, 2009), intensity of adult aggression (Le Bohec et al., 2005; Penteriani et al., 2003), and weather conditions (Black et al., 2016; Wilson, 2009), although Black et al. (2016) show that huddling behaviour can also vary with location. Furthermore, the primary driver can depend on species – huddling in Gentoo penguins was found to be mainly associated with wet, cold conditions (Black et al., 2016), while adult aggression also seems to be a key driver for King penguin chicks (Le Bohec et al., 2005). In the *Eudyptula*, *Spheniscus*, and *Megadyptula* genera (which include species such as the Magellanic, Yellow-Eyed (*Megadyptes antipodes*) and Humboldt penguins), there is no evidence for any huddling behaviour (Wilson, 2009). Since the formation of aggregations is thought to be an ancestral behaviour in penguins, this trait must have been lost in these lineages (Wilson, 2009).

Once chicks have moulted their down feathers and fledged, the adult birds undergo hyperphagia (an intensive foraging effort to increase body fat) before their annual catastrophic moult. They then either remain close to the breeding site (sedentary species; e.g. the Yellow-Eyed penguin) or travel great distances (migratory species; e.g. the Macaroni penguin) during the winter period – a final example of inter-species variation in life history strategies (Black et al., 2017). For example, Magellanic and Chinstrap penguins can travel vast distances during the winter migration, with tracking studies recording journeys of 2,700km and 1,300km for the two species, respectively (Pütz et al., 2000; Trivelpiece et al., 2007). It should be noted that migration distance can vary with year and location. For example, Trivelpiece et al., (2007) found that Chinstrap penguins from two colonies on the South Shetland Islands (Cape Sheriff and Admiralty Bay) predominantly

foraged inshore during the winter of 2000, but mainly foraged offshore in 2004. They suggest that fewer food resources were available inshore during the 2004 winter, forcing the penguins further from the breeding colony. Towards the northern limits of the Adélie penguin range, individuals may visit breeding sites throughout the year, whereas those that breed at the highest latitudes of their distribution (which extends to almost 78°S) migrate north to avoid the harsh winter conditions of this habitat (Ballard et al., 2010). Adélies from Ross Island are thought to undertake the longest migrations of the species, with Ballard et al., (2010) reporting a total journey distance of 17,600km for one individual.

iii) The Pygoscelis penguins

The *Pygoscelis*, or "brush-tailed", penguin genus comprises the Adélies, Chinstraps and Gentoos (Fig. 4). They diverged from other Spheniscidae taxa approximately 38 million years ago (Ma) (31.6-44.7Ma), in the Eocene epoch (Baker et al., 2006). The most closely related species of this genus are the Chinstraps and Gentoos, with a divergence date estimated at 14.1Ma (10.8-18.3Ma), while the Adélie penguin became a distinct species approximately 19Ma (15.4-23.9Ma) (Baker et al., 2006). Following the climatic cooling event which took place between 34Ma and 25Ma, the *Pygoscelis* penguins remained in Antarctica, the only extant penguin group to do so other than the *Aptenodytes* (Emperors and Kings) (Trivelpiece & Trivelpiece, 2013).

The three *Pygoscelis* species are partially sympatric in their distribution, with overlap occurring over regions such as the Western Antarctic Peninsula, South Shetland Islands, South Orkney Islands and South Sandwich Islands (Croxall & Kirkwood, 1979; Dunn et al., 2016). However, their latitudinal boundaries vary considerably, with the southernmost

Adélie colony located at 77°S (Cape Royds, Ross Sea), compared to 64°S and 65°S in Chinstraps and Gentoos, respectively (Lynch, 2013; Trathan & Ballard, 2013; Trivelpiece & Trivelpiece, 2013). Gentoos are the most northerly species, with breeding colonies situated on the Falkland Islands (52°S) and Crozet Islands (46°S), and – like the Adélies – have a circumpolar range (Lynch, 2013; Trathan & Ballard, 2013). Populations of Chinstrap penguins are found on the South Sandwich Islands, and although 90% of their population breeds in the Scotia Sea, they have a nonbreeding range that extends to the edge of the upper Ross Sea (Lishman, 1985; Trivelpiece & Trivelpiece, 2013).



Figure 4: From left to right – Adélie penguins (*Pygoscelis adeliae*), Andresen Island; Chinstrap penguins (*P. antarcticus*), Orne Harbour; Gentoo penguins (*P. papua*), Madder Cliffs. Photos: F. M. Jones.

Adélie and Chinstrap penguins can breed in vast numbers, with the largest recorded Adélie colony, situated at Cape Adare in the Ross Sea, estimated to comprise 428,516 nests (Cimino et al., 2016; Humphries et al., 2017). Approximately 1.3 million breeding pairs of Chinstrap penguins are estimated to inhabit the South Sandwich Islands, representing

roughly half of the global population (Hart & Convey, 2018; Lynch et al., 2016). In contrast, Gentoo colonies are relatively small, with the largest numbering only 6000 breeding pairs (Lynch, 2013).

In addition to colony size, there are a number of other traits which differentiate the Gentoo penguin from its sister taxa. For instance, Gentoos are highly variable spatially, both in terms of their morphology and behavioural ecology. As summarised by Lynch (2013), there are variations in biometric characteristics such as body mass, bill dimensions and flipper area between individuals inhabiting different areas across the Southern Ocean. Compared to the other *Pygoscelids*, Gentoos are relatively sedentary (particularly to the north), mainly foraging inshore, which may explain their genetic and phenotypic variability, as they become increasingly well-adapted to their local environment (Bost & Jouventin, 1990; Lynch, 2013). Two subspecies of Gentoo – *P. papua ellsworthi* and *P. papua papua* – are officially recognised (although recent genetic analyses points to four clades (Vianna et al., 2017)), while the Adélies and Chinstraps exhibit high levels of genetic homogeneity (Freer et al., 2015; Lynch, 2013; Trathan & Ballard, 2013). Levy et al. (2016) suggest that the Polar Front acts as an important extrinsic barrier to gene flow, working alongside intrinsic factors such as natal philopatry, to maintain genetic differentiation between the northern (*P. papua papua*) and southern (*P. papua ellsworthi*) subspecies.

The Adélie and Chinstrap diet mainly comprises Antarctic krill (*Euphausia superba*, hereafter "krill"), a small crustacean that is a keystone species in the Southern Ocean ecosystem (Lishman, 1985; Lynnes et al., 2004). While krill also form an integral part of the Gentoo diet – particularly towards the southern extent of their range (McMahon et al., in review) – other prey species, including fish and cephalopods, are important (Lescroël et

al., 2004). As with their morphology, Gentoos thus exhibit greater variability than their congeners, with the former being generalist, opportunistic foragers (Lescroël et al., 2004; Lynch, 2013). Indeed, the spatial variability observed in the Gentoo diet is a logical consequence of their feeding ecology, with individuals relying on near-shore prey items which may only be found locally (Lescroël et al., 2004).

Predator species are similar amongst the *Pygoscelid* penguins, with skuas (*Catharacta antarctica* and *C. lonnbergi*), giant petrels (*Macronectes giganteus*) and kelp gulls (*Larus dominicanus*) amongst the main avian predators (Emslie et al., 1995; Lynch, 2013; Trathan & Ballard, 2013; Young, 1994). Although young individuals – and sometimes even adults – can be at risk from these birds, eggs and chicks are the primary target, with lone chicks being particularly vulnerable (Lynch, 2013; Trathan & Ballard, 2013). Once fledged, Leopard Seals (*Hydrurga leptonyx*) will predate on all three *Pygoscelis* species (Ainley et al., 2005; Lynch, 2013; Penney & Lowry, 1967; Rogers & Bryden, 1995; Trathan & Ballard, 2013). Ainley et al. (2005) suggest that penguins leaving (or returning to) intermediate sized colonies are most at risk from Leopard Seal predation. When "penguin traffic" was slower than 250 individuals per hour, Leopard Seals were not persistent in their predation attempts (presumably because their return was not sufficient to outweigh their predation costs), while the flow at large penguin colonies was enough to "'swamp' the seals' predatory efforts" (Ainley et al., 2005). The risk of predation from killer whales (*Orcinus orca*) in Antarctic waters is unclear, although Pitman & Durban (2010) report sightings of a predation event on Chinstrap and Gentoo penguins in the Gerlache Strait. Observations by Ballard & Ainley (2005) suggest that Adélies are not usually targeted, but the authors posit that individuals may be used to 'train' juveniles in hunting techniques.

As previously mentioned, the stages in the penguin annual cycle are relatively consistent between species (Fig. 2). However, even within the *Pygoscelis* genus, marked variation exists in key aspects of the yearly cycle, such as phenological timings (e.g. date of chick fledge), both within and between species (Black, 2016; Lynch, 2013; Trathan & Ballard, 2013; Trivelpiece & Trivelpiece, 2013). For example, Gentoos inhabiting Possession Island (Crozet Islands; 46°24'S, 51°46'E) have a reported mean egg lay date (for first clutches) of 23 July ($\pm$7 days; with replacement clutches laid until late November) – more than three months earlier than on Bird Island, South Georgia (54°0'S,38°3'W), where the mean date is 26.5 Oct ($\pm$ 3.7 days) (Bost & Jouventin, 1991; Lynch, 2013; Williams, 1990b). Furthermore, the duration of each stage varies – for example, the "guard" period is generally longer in Chinstraps and Gentoos than in Adélie penguins (see Table 1) (Black, 2016 (and references within); Lynch, 2013; Trathan & Ballard, 2013; Trivelpiece & Trivelpiece, 2013).

As colonial species, the *Pygoscelis* penguins breed in relative synchrony within each nesting site (Black, 2016). However, there are interspecies differences in the degree of this synchronicity, with Gentoos being the most asynchronous, and also the most variable in this trait (Bost & Jouventin, 1990). In Gentoo colonies on the Antarctic Peninsula, the egg laying period may be restricted to 28 days, and on the South Shetlands Islands can even be comparable to the Adélies (16 days) and Chinstraps (10 days), at 16 days (Bost & Jouventin, 1990; Trivelpiece et al., 1987). However, on the Crozet Islands, the egg laying period may last up to 154 days, partly owing to the ability of this species to lay replacement clutches, a phenomenon which is significantly more pronounced in the milder subantarctic than in Antarctic colonies (Bost & Jouventin, 1990).

While *Pygoscelid* phenology has been studied for a century, substantial gaps remain in our understanding (Black, 2016). This is especially true for certain combinations of life history stage, location, and species. For instance, studies of Chinstrap phenology have largely been restricted to the South Shetland Islands and South Orkney Islands, meaning spatial variation in the timing of Chinstrap life history stages is poorly understood (Black, 2016). A summary of the similarities and differences in the annual life cycle of Adélie, Chinstrap and Gentoo penguins is provided in Table 1.

iv) Key threats to Southern Ocean penguins

As with other seabird groups, penguins are at threat from a number of direct and indirect anthropogenic stressors, including fishing activities (leading to resource competition and bycatch), habitat degradation and pollution (Trathan et al., 2015). In the Southern Ocean – home to nine penguin species (of the *Aptenodytes*, *Pygoscelis* and *Eudyptes* genera) – the main threats are arguably climate change, an expanding krill fishery and human visitation (in the form of scientific research, national programmes or tourism).

With a rate of warming equivalent to 3.5°C±0.8°C per century, the Antarctica Peninsula is at present one of the fastest warming regions on the planet (Mulvaney et al., 2012). Alongside this increase in temperature comes a plethora of other climatic effects. Precipitation data from the Faraday/Vernadsky research station (65.14°S, 64.15°W) shows an increase of 13 precipitation events, or 5.7 precipitation days, per decade, based on data from 1960 to 1999 (Kirchgäßner, 2011). Heavy precipitation can have severe deleterious effects on penguin breeding colonies, with nest flooding from pooled rainwater or

	Adélie (*Pygoscelis adeliae*)	Chinstrap (*Pygoscelis antarcticus*)	Gentoo (*Pygoscelis papua*)
Winter migration	Migratory – at sea between May and August; can swim thousands of kilometres away from the breeding site; stays north of ~73°S (to use light); can forage in pack-ice (up to ~80% cover).	Migratory – at sea between May and September; spends the winter in waters north of the pack-ice.	Sedentary compared to the other *Pygoscelis* species.
Return to colony (start of breeding)	September to October.	Early October and through November.	*Northern populations*: breeding begins in early winter, with the breeding period lasting nearly 5 months in some locations. *Southern populations*: breeding begins in early spring and lasts only 2 to 3 weeks.
Partner fidelity estimate	60% to 80%	82%	49% to 90%, site dependent.
Site fidelity estimate	60% to 90%	94% (males), 82% (females)	60% to 100%
Egg laying	October (can extend into November).	October and November	Highly variable, see above.
Number of eggs	2	2	2
Incubation	32 to 34 days; both parents incubate in shifts, starting with the male.	33 to 36 days; both parents incubate in shifts, starting with the female.	33 to 37 days; both parents incubate in shifts, starting with the female.
Guard period	22 days post hatching.	~28 days post hatching.	25 to 30 days post hatching.

Crèche	Provisioned every 1 to 2 days by parents.	Provisioned every day by both parents.	Provisioned every 1 to 2 days by parents.
Chick fledging	February, aged 50 to 60 days.	Aged ~49 to ~56 days	Aged 80 to 105 days; evidence that chicks can be provisioned post-fledging.
Adults leave colony	Post chick-fledging, adults go to sea to feed.	Post chick-fledging, in late February, adults go to sea to feed.	Post-chick fledging, the adults go to sea to feed.
Adult postnuptial catastrophic moult	Late February or early March, roughly 4 weeks after leaving breeding site; lasts 2 to 3 weeks. Proximity to colony varies (moult can occur on sea ice away from the colony, or adults may return to the breeding site, or close by).	March; lasts approximately 3 weeks.	Lasts an average of 19.5 days. Dates vary widely between locations; moulting is complete between mid-December and mid-February, on average.
References:	(Black, 2016; Trathan & Ballard, 2013*)	(Black, 2016; Trivelpiece & Trivelpiece, 2013*)	(Black, 2016; Lynch, 2013*; Williams & Rothery, 1990)

Table 1: A comparison of phenological stages in the Adélie (*Pygoscelis adeliae*), Chinstrap (*P. antarcticus*) and Gentoo (*P. papua*) penguin species. * In Borboroglu & Boersma, 2013.

snowmelt (Youngflesh, 2018) leading to drowning of eggs and young chicks (Cimino et al., 2014; Massom et al., 2006; Thyen & Becker, 2006). Chicks are also at risk of hypothermia via wetting, requiring growth of juvenile plumage – in replacement of their down – before they can withstand intense rain (Boersma, 2008). Reduced chick growth rate has been attributed to increased precipitation in a number of bird species (e.g. the Little Auk *Alle alle* (Konarzewski & Taylor, 1989), Common Tern *Sterna hirundo* (Dunn, 1975) and Black-headed Gull *Chroicocephalus ridibundus* (Thyen & Becker, 2006)), and model simulations suggest that a wetting of just 10% is enough to reduce chick survival probability via a reduction in fledging weight (Chapman et al., 2011).

Studies also highlight the acceleration of ice shelf break-up, catalysed by a longer summer melt season and increased surface 'ponding' (Abram et al., 2013; Van den Broeke, 2005). Of particular importance for penguin populations is the extent and duration of sea ice cover. Long-lasting sea ice, such as that observed in the Terre Adélie region during the 2013-2014 breeding season, can prevent adults from adequately provisioning chicks, leading to mass mortality events (Ropert-Coudert et al., 2015). On the other hand, insufficient sea ice can be equally dangerous to species dependent on sea ice to breed (i.e. the Emperor (Fretwell & Trathan, 2019)) or provide a crucial winter habitat (e.g. the Adélie (Croxall et al., 2002)). In the Antarctic Peninsula and Bellingshausen Sea region, the ice-free season increased by more than three months between 1979/1980 and 2010/2011, making it comparable to observed Arctic changes (Ducklow et al., 2013; Maksym et al., 2012; Stammerjohn et al., 2012).

Krill are a keystone species in the Southern Ocean ecosystem, providing a lipid-rich food source for predators across trophic levels. They are also the focus of the largest (by

tonnage) fishery in the Southern Ocean, with an annual catch limit of 620,000 tonnes set by the Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR – see section *viii*) for the region currently fished (Subareas 48.1 to 48.4) (CCAMLR, 2018; Nicol et al., 2012). While considered "under-exploited" at present, growing interest in the fishery – indicated by a rise in patents for krill-related products – and advances in krill fishing technologies, mean that cautious management is essential (Nicol et al., 2012). Furthermore, while overall catch limits may be observed, there are currently no restrictions placed on fishing intensity (i.e. how much of this limit is extracted at a certain location, during a certain time period), with clear "bottom up" ramifications for local predator populations if local stocks are depleted (Daunt et al., 2008; Sherley et al., 2015). Increased monitoring of predator populations in areas of known fishing activity will enhance our understanding of local fisheries' impact, and help inform future guidelines.

More than half of the circumpolar krill stock is found in the southwest Atlantic sector (20°W-80°W), which has also been the focus of the krill fishery (Atkinson et al., 2019; Kawaguchi et al., 2009). However, recent analyses indicate a substantial contraction southward within this region (by approximately 440km over 90 years), with the majority of the change occurring since 1970 (Atkinson et al., 2019). Alongside this change in spatial distribution comes a shift in the size demographic of the krill population, with current individuals 6mm longer on average than those sampled in 1970 (Atkinson et al., 2019). This appears to translate to a decline in recruitment, seemingly driven by an increasing frequency of positive Southern Annular Mode (SAM) anomalies, bringing warmer and windier conditions, with less sea ice (Atkinson et al., 2019). While the relationship between sea ice and larval krill development has been difficult to elucidate, Meyer et al. (2017) suggest that the complex underside topography of the marginal ice zone offers juveniles

an attractive feeding ground, providing both protection from predators and a high density of ice-associated phytoplankton, relative to the surrounding area. A reduction in sea ice thus has the potential to affect penguin populations in multiple ways – directly (i.e. by providing more/less favourable habitat for ice-loving/ice-avoiding species) and indirectly (i.e. via the impact on a primary food resource). Furthermore, the trends of declining sea ice and a growing krill fishery are synergistic, with non-orthogonal effects: a lengthening of the summer ice-free season offers more time for the krill fishery to operate. In the past, krill have been protected by the winter sea ice cover – from both predators and the fishery (Kawaguchi et al., 2009). Year-round harvesting of krill, made possible by the opening of new shipping routes, would require careful management, and more extensive monitoring of predator populations.

Studies examining the impact of site visitation on Southern Ocean penguin populations have drawn inconsistent conclusions. As discussed by Lynch et al. (2010a) this is likely owing to the wide variation in study types (for example, testing physiological response (Lynch et al., 2019; Regel & Pütz, 1997; Wilson et al., 1991) versus analysing breeding success and/or population trends (e.g. Giese, 1996; Trathan et al., 2008; Woehler et al., 1994; Young, 1990), differences in study period (i.e. single versus multiple seasons), species examined, and visitation type (e.g. scientific observation or recreational). However, studies have produced contradictory results even when examining a specific physiological indicator such as heart rate (e.g. Culik et al., 1990; Culik & Wilson, 1995; Nimon et al., 1995), or a single parameter such as breeding success (e.g. Carlini et al., 2007; Woehler et al., 1994).

Bricher et al. (2008) found that proximity of nesting Adélie penguins to Casey research station was an important predictor of population trends on Shirley Island, outweighing the effects of wind exposure and snow accumulation. This is in contrast to the work of Carlini et al. (2007) who found no difference in the population trend of breeding pairs of Adélie penguins in areas of low and high disturbance near the Esperanza research station at Hope Bay. In fact, during the 1996-1997 season, breeding success was greater in the highly disturbed location, an effect also reported at other sites (e.g. Petermann Island, where Adélie (but not Gentoo) penguin breeding success was found to be higher in "frequently visited" colonies than "occasionally visited" colonies (Lynch et al., 2010a)). One explanation for these observations is habituation, where penguins become desensitised to human presence through repeated exposure to stimuli (Rankin et al., 2009). For example, Holmes et al. (2006) reported significantly stronger behavioural responses (e.g. high vigilance and low threat/display behaviours) to human approach off-site (i.e. away from the Australian Antarctic Program station on Macquarie Island) than on-site. Furthermore, Viblanc et al. (2012) detected attenuated heart rate responses to common human stimuli (sounds or human approach), but not infrequent situations (such as capture) in King penguins close to the Alfred Faure station, Possession Island.

Trathan et al. (2008) suggest that habituation to human activities may explain the population trends observed at Goudier Island, Port Lockroy – one of the most popular tourism sites in Antarctica (Fig. 5). They report no decline in breeding success (concordant with the findings of Cobley & Shears (1999)), instead attributing local population declines to a reduction in nest number. They suggest that older birds have habituated to the presence of tourists – perhaps having reached breeding age when visitation levels were lower – and therefore continue to nest at the same sites, while younger (i.e. unhabituated)

birds seek out unvisited sites (see Woehler et al. (1994)). It should be noted that habituation is not always observed, with Wilson et al. (1991) reporting behavioural and physiological stress responses to both human approach and aircraft disturbance, despite continued exposure.



Figure 5: Port Lockroy, Goudier Island, Antarctic Peninsula (64°49'S, 63°29'W) is one of the most popular tourist destinations in Antarctica, receiving 19,834 visitors over 112 days during the 2015-2016 season (Dunn et al., 2019). Photo: F. M. Jones.

Regardless of the results of these, or similar, studies, it is clear that continued regulation of human activities in the vicinity of Antarctic and sub-Antarctic penguin colonies is essential. As highlighted by Wilson et al. (1991) and Coetzee and Chown (2016), an absence of behavioural stress response does not necessarily equate to an absence of physiological response, or subsequent changes in breeding success or population trends, emphasising the need for long-term monitoring. Tourism in the region has increased exponentially over the last three decades (Bender et al., 2016; Lynch et al., 2010b), with

implications not only for terrestrial environments, but also for sensitive marine areas, with a marked rise in ship traffic and small boat activities – particularly around tourism 'hotspots' (Lynch et al., 2010b). Colonies of all three *Pygoscelis* species are regularly visited by tourists, with >73%, 20% and 16% of landings overlapping with Gentoo, Chinstrap and Adélie colonies, respectively (Bender et al., 2016; data pertain to the 50 most popular sites and the 2013-2014 season). The observed decrease in landed passengers at Hannah Point (home to a sensitive colony of >120 nesting southern giant petrels (*Macronectes giganteus*)) since the adoption of Site Guidelines (Antarctic Treaty Consultative Parties (ATCP), 2006) and voluntary restrictions, shows that such measures can be effective, and similar action can be taken to protect penguin colonies if necessary (Lynch et al., 2010b).

v) *Penguin population monitoring – from painted stones to UAVs*

Between January 12th 1921 and January 13th 1922, geologist Thomas Bagshawe and navigator Maxime Charles Lester lived in an upturned waterboat on the Danco Coast of the Antarctic Peninsula (Bagshawe, 1938). They had departed from England as part of the 'British Imperial Expedition' which, led by John Lachlan Cope, had grand aims – including a complete circumnavigation of Antarctica and continued exploration of the Weddell Sea coast (the latter following on from the earlier work of Nordenskjöld) (Cowan, 2001). However, after some "unanticipated difficulties" (mainly resulting from a lack of funds), Bagshawe and Lester found themselves as a party of two, in the spot they were to name 'Water-boat Point' (64°48'S, 62°43'W; (Bagshawe, 1938; Cowan, 2001)). During their year on the Peninsula they made extensive and detailed tidal and meteorological observations, and – though self-described "novices" – also compiled comprehensive notes on the

physical characteristics, mode of life and breeding behaviour of the penguins that surrounded them (mainly Gentoos and Chinstraps; (Bagshawe, 1938)). They recorded phenological timings (e.g. date of egg laying and chick hatching, and incubation time), performed counts during the "reoccupation of rookeries", and – over a period of 38 days – made hourly observations, day and night, of the number of penguins on each nest (Bagshawe, 1938). They marked penguins using Indian ink, and identified nests using stones painted with enamel, or sketches of nest locations relative to landmarks (Bagshawe, 1938). Although neither was a naturalist, these detailed observations are a valuable resource, not least because of the high temporal resolution of the data.

In order to assess the current or potential impact of environmental change and direct anthropogenic threats on Southern Ocean penguins (see section *iv*), high resolution monitoring of colonies, both temporally and spatially, is essential. This includes collecting data on colony location (and identification of new or newly extinct colonies), size (i.e. population censuses), breeding success and phenology. In some respects, the ultimate aim is to undertake large-scale studies as comprehensive as those of Bagshawe and Lester, without the logistical, practical and financial challenges associated with long-term manned monitoring projects, and to do so in a way which does not cause colony disturbance.

Population counts

Population censuses have traditionally been undertaken through 'ground counts', where researchers are positioned at a colony and tasked with manually counting nests as a proxy for the number of breeding pairs (Crawford et al., 2003; Lescroël & Bost, 2006). To reduce error, counts are usually repeated (see e.g. Bingham, 1998; Taylor et al., 1990). For example, Naveen et al., (2012) used rope to divide a large Chinstrap colony into

increasingly smaller sections, until three counts of a section were within 5% of the mean. If a colony is large, it may be necessary to extrapolate counts based on the nest density of a smaller area, although the error associated with this strategy is larger than for direct counts (e.g. within $\pm$ 10% compared to within $\pm$ 5% (Bingham, 1998)). These methods are labour-intensive, limiting the number of sites that can be surveyed by an individual research team in one season. In an effort to increase the spatial coverage of count data, and therefore aid implementation of the 1991 'Madrid Protocol' (see section *viii*), the Antarctic Site Inventory (ASI) was launched in 1994 (Naveen et al., 2000). As contributors to the ASI, researchers are placed on tour boats, and carry out censuses as part of scheduled tourist landings or Zodiac (small boat) cruises. Additional site visits and aerial surveys (see below) have been facilitated through cooperation with the Royal Navy ice patrol ship *HMS Endurance* (Naveen et al., 2000). As such, they are able to survey a wide range of locations, and between 1994 and 2015 (across 22 field seasons) researchers from the ASI undertook 1713 visits, retrieving data from 223 sites (Oceanites, 2019). Owing to the opportunistic nature of this work, surveys do not always correspond to the optimal time to carry out a population census (e.g. during egg laying), but this issue is arguably outweighed by the increased coverage the ASI provides (Lynch et al., 2008).

Aerial counts – where photographs are captured using light aircraft (Kooyman & Mullins, 1990; Taylor et al., 1990), helicopters (Haftorn et al., 1981) and more recently Unmanned Aerial Vehicles (UAVs; Fig. 6 (Goebel et al., 2015; Ratcliffe et al., 2015)) or kites (Fraser et al., 1999) – have proved to be a reliable alternative, or addition, to ground counts. They are particularly useful for surveying large geographic areas, or sites that are inaccessible during part or all of the year – for example, many penguin colonies in the Ross Sea region are out of helicopter range of bases, and are not visited by ships during the incubation

period (Taylor et al., 1990). Under CCAMLR guidelines, counts must be performed at the peak of egg laying (for breeding pairs) or crèche (for chicks), and aerial surveys can therefore assist with these requirements (Lynch et al., 2008).



Figure 6: Unmanned aerial vehicle (UAV; DJI Mavik Air) image of an Emperor penguin (*Aptenodytes forsteri*) colony at Snow Hill Island, Antarctic Peninsula (64°28'S, 57°12'W). Photo: T. Hart and F. M. Jones.

Being cheaper and more accessible than manned aircraft, UAVs (or drones) are rapidly becoming the tool of choice for many ecological researchers (Crutsinger et al., 2016; Ivošević et al., 2015; Koh & Wich, 2012). UAV images (Fig. 6) can be captured at lower altitude than conventional aerial photographs (since light aircraft cause greater disturbance at equivalent heights), producing high-resolution imagery where every bird in a colony is potentially resolved (Ratcliffe et al., 2015). Furthermore, these images can be digitally stitched together and combined with Global Positioning System (GPS) data in the

field, producing detailed geo-referenced maps of colony habitat (Koh & Wich, 2012; Ratcliffe et al., 2015).

A final censusing strategy comes from commercial satellite imagery, where the potential for ecological monitoring applications has been realised for a number of years (Schwaller et al., 1984). Satellite data have played a crucial role in the identification of new Emperor penguin colonies, where the difference in reflectance spectra between sea ice and guano stains has been used to locate sites (Fretwell & Trathan, 2009). In addition, high-resolution panchromatic images (sometimes combined with multispectral data) have been used to assess the relative size of colonies (Barber-Meyer et al., 2007), or even produce population estimates (Fretwell et al., 2012; Lynch et al., 2012a; Naveen et al., 2012). These studies employ linear regression of known nest densities (and how these relate to pixel densities) to generate estimates of nest densities in satellite images. Count data produced using this method have proved reliable; for example, Lynch et al. (2012a) calculated a size estimate of 115,673 breeding pairs (95^{th} percentile CI = 99,222 – 127,203) for an Adélie colony on Paulet Island, which is comparable to an estimate of 95,000 – 105,000 ($\pm$25%) generated using data from the ASI (Naveen et al., 2000).

Each censusing method is associated with its own challenges, some of which have already been discussed. While a safe operating environment is essential for fieldwork, poor weather conditions do not automatically prohibit ground counts (unless they prevent ship-based researchers from reaching the shore), whereas high winds and precipitation can cancel scheduled UAV flights, and cloud cover can obscure satellite imagery (Lynch et al., 2012a; Ratcliffe et al., 2015). Disturbance is also a concern when monitoring, whether via direct entrance to a colony, or through noise and visual disturbance from above. Ongoing

research into the effects of UAVs at different altitudes is required, and it is important that researchers adhere to the latest guidelines, whatever their chosen monitoring method.

Monitoring phenology and breeding behaviour

Traditionally, examination of penguin phenology, reproductive success and breeding behaviour has involved season-long on-site monitoring projects, often in close proximity to research bases for logistical ease. In order to obtain temporally precise phenological data, daily colony visits are often required, particularly prior to transitionary phases such as egg lay or chick crèche (Crawford et al., 2003; Lynch et al., 2010a; Moreno et al., 1994). As such, these studies are labour-intensive, and spatial coverage (in terms of the number of nests or colonies that can be analysed) is limited by time-constraints, team size, and access to support facilities. Furthermore, while some researchers (e.g. Moreno et al., (1994)) report that their observations – which can require manual checking of nests for eggs and chicks – had no significant impact on survival, others (e.g. Crawford et al., (2003)) found their activities were causing "considerable" disturbance. In response, the team switched to an alternative monitoring strategy, observing nests from a distance using binoculars (Crawford et al., 2003).

An attractive alternative to these conventional monitoring studies can be found in remote camera technologies, and specifically time-lapse imaging techniques. Time-lapse footage (albeit in a primitive form) has been used to aid penguin research for many decades. In the early 1960s, in response to the "time-consuming, meticulous, and demanding task" of nest data collection, Milton Weller began to develop an 8mm time-lapse video camera, solely for the purpose of "recording events at nests" (Weller & Derksen, 1972). In the 1969-1970 breeding season, Weller and Derken deployed this camera (a Nikkorex Zoom 8 with a time-

lapse setting of 2.47 minutes), and two Nizo S-80 cameras (with inbuilt time-lapse functions set to one image per minute), at an Adélie colony at Hallet Station (Weller & Derksen, 1972). In case the internal time-keeping mechanisms failed, a clock was placed in the field of view, as shown in Fig. 7.

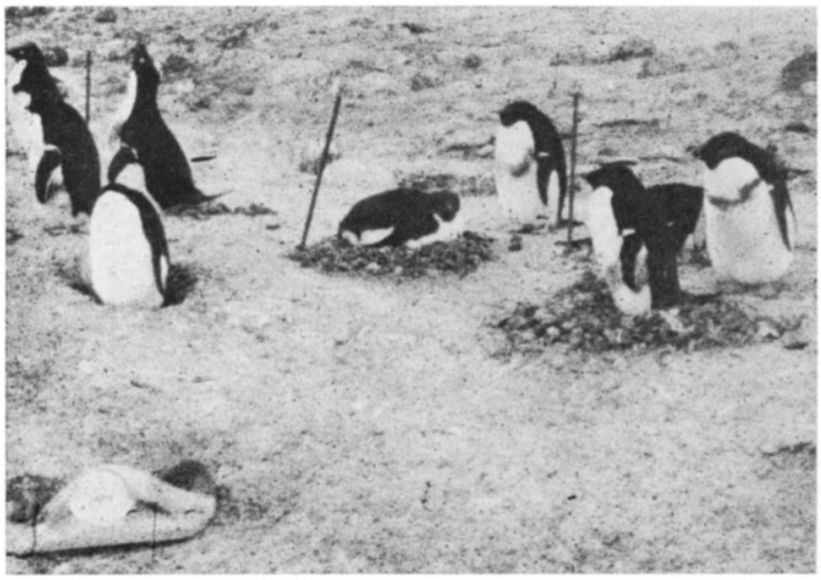


Figure 1. Colony of marked and banded Adélie Penguins showing nest-marking stakes and clock. Chest mark shows bird at right is a male, and a marked female is third from right. Male third from left is starting a head-up ecstatic display.

Figure 7: An early example of the use of time-lapse imagery in penguin monitoring studies, taken from Weller & Derksen, 1972.[1]

During the course of this study, they exposed 2,100 ft. of Super 8 film, producing almost 151,200 individual frames with which to examine incubation behaviour. A further 108,000 frames (over 1,500 ft. of Super 8 film) were collected in a follow-up study (Derksen, 1977). Other early examples come from Penney (1968), who captured a 16mm photograph every ten minutes throughout the 1959-1960 and 1960-1961 breeding seasons to examine patterns of daily activities and the breeding cycle of Adélie penguins, and Haftorn (1986)

[1] Weller & Derksen, Use of Time-Lapse Photography to Study Nesting Activities of Birds, *The Auk*, 1972, 89, 1, pp. 196–200, DOI: N/A. Reprinted by permission of Oxford University Press on behalf of the American Ornithological Society. Exempt from thesis copyright licence - reprint only with permission.

who used time-lapse imagery to study sex-based division of labour of Chinstrap and Macaroni penguins.

These early studies were restricted by technological constraints, in particular with regard to battery life and film capacity (Weller & Derksen, 1972). For example, one cartridge of Super 8 film can store approximately 1,500 frames, meaning – if images were captured at a frequency of one frame per minute – replacement was necessary every 60 hours (Weller & Derksen, 1972). Weller and Derksen (1972) took measures to increase battery life (e.g. storage in a heated insulated box or observation hut), but even so, batteries also required replacement every 60 hours.

The exceptional advancement in camera and software technologies that has occurred during the 21st Century means that time-lapse monitoring now represents an efficient, affordable, and practical means of ecological monitoring (Appendix 1 (Hart et al., in review)). Moreover, autonomous camera systems, combined with long-lasting batteries and high-storage memory cards, facilitate remote monitoring projects with little requirement for human input. In recent years, remote time-lapse cameras have been used to examine penguin behaviour (e.g. huddling (Black et al., 2016) and movement during the winter months (Black et al., 2017)), phenology (Black et al., 2018; Hinke et al., 2018; Southwell & Emmerson, 2015) and breeding success (Hinke et al., 2018; Southwell & Emmerson, 2015). The potential for remote camera technologies in ecological studies is discussed further in section *vi*.

Please note that much has been learnt about the at-sea behaviours and migration routes of penguins through the attachment of GPS tracking devices (see e.g. Bost et al., 2009;

Takahashi et al., 2018). However, this work is beyond the scope of this thesis and therefore not discussed here.

vi) Remote camera technology: a tool for ecological monitoring

> *"For…detailed knowledge of the appearance and behavior of penguins we have to thank not only the publications of naturalists…but also those new adjutants of exploration, the camera-men" (Robert Cushman Murphy, 1936).*

Certain phenomena of interest to ecologists, for example rare animal behaviours or shifts in phenological characteristics, can be difficult to observe using traditional monitoring techniques. In the case of field observations, visits are labour-intensive, and their frequency may be insufficient to capture unusual behaviours, or to effectively document ongoing population changes (Collins et al., 2014; Nagai et al. 2014). Furthermore, the newer technique of satellite remote sensing can be inadequate in some cases, owing to low spatial resolution, potential cloud disruption, and the need for atmospheric correction (Richardson et al., 2013). In addition, this latter approach is clearly unsuitable for certain projects, for instance where a single nest or small patch of vegetation requires monitoring.

Ground-based remote camera technologies offer a solution to this problem, allowing sites to be monitored frequently throughout the year, without the need for continuous human input (Southwell & Emmerson, 2015). Cameras may capture images only when triggered by wildlife (see Holden et al. (2003) and Pettorelli et al. (2010) for examples of studies using 'camera traps'), or may take a photograph at pre-programmed intervals. Time-lapse cameras (see Fig. 8) have been deployed in a range of environments, including Antarctic, marine and alpine habitats, and have the potential to provide detailed information about

populations in remote locations (Ide & Oguma, 2013; Southwell & Emmerson, 2015; Peirano et al. 2016). Theoretically, a population need only be visited twice (once to deploy the camera, and once to retrieve it), to obtain many months of footage – the exact time dependent upon battery life.



Figure 8: Example of a remote time-lapse camera (part of the *Penguin Watch* network – see section *ix*). Photo: F. M. Jones.

Research studies in polar regions are expensive (Mallory et al. (2018) suggest an Arctic project costs approximately eight times as much as an equivalent study in a temperate location). By reducing the requirement for field observation, automated time-lapse cameras facilitate large-scale projects – both temporally and spatially (Southwell & Emmerson, 2015), and the ability to use 'off-the-shelf' devices, combined with reduced

fieldwork hours, has economic benefits (Ide & Oguma, 2013). Southwell and Emmerson (2015) estimate that a camera network covering 20 locations across East Antarctica for ten years could be maintained for one-tenth of the cost of a comparable traditional polar monitoring study. In addition, studies employing time-lapse cameras in novel ways, for example as "inexpensive multispectral sensors" (recording red, green and blue wavelengths), show that they can also be used in place of higher cost alternatives (Ide & Oguma, 2013).

By deploying cameras that take photographs at regular intervals, regardless of what is in the field of view, observer bias can be reduced (Nagai et al., 2014). Of course, post hoc image processing is required, and efforts must be made to avoid sampling bias at this stage. Image analyses can be time-consuming, and researchers may seek ways to improve the efficiency of this process, such as citizen science or machine learning. Furthermore, while limiting observer bias, there are disadvantages to the use of a stationary device operating with minimal human input. Firstly, a subject of interest may fall outside the field of view. Studies examining phenology may be dependent on identifying the first egg or fruit of the season, and these could easily be hidden by another object (Laskin & McDermid, 2016). Furthermore, the camera lens could be covered, for example by snow, obscuring the image (Southwell & Emmerson, 2015). While this may be temporary, other problems – such as movement of the camera (e.g. by fauna or wind), movement of the subject (e.g. through plant growth or repositioning of a nest) and theft or vandalism – could render a dataset unusable (Laskin & McDermid, 2016; Merkel et al., 2016).

In order to assess the reliability and accuracy of measurements made using camera systems, Southwell and Emmerson (2015) examined data collected from a network of

remote cameras in East Antarctica, positioned to monitor breeding phenology in Adélie penguins. A comparison was made between direct (i.e. field-based) measurements and camera-derived data (one image per day) for approximately 2000 breeding pairs at Béchervaise Island; overall camera performance was also considered. A strong correlation was found between estimates of breeding success made using the two methods over eight breeding seasons ($r \geq 0.95$, df=6, $p \leq 0.0003$). There was a slight negative bias associated with the camera measurements – i.e. a tendency for estimates to be lower – when chick counts were made earlier in the season (no bias was detected in late January). The authors ascribe this to "low detectability of chicks when they are small and closely guarded by the brooding parent". First arrival was recorded as zero to two days later using the camera system, while the first egg was recorded between two and six days later. However, since there is little annual variation in the time lag, the system is appropriate for the detection of change over time (Southwell & Emmerson, 2015).

While adjustments to the camera structures were required (e.g. to protect the lens from abrasion), the authors note an absence of "major malfunctions" over multiple years. Of 5000 images, 99.3% were usable for research, and 71% of unusable images were captured during blizzard conditions. The authors conclude that the remote camera network can provide valuable phenological data in a cost-effective manner, and suggest that the system is adopted and developed by those examining other seabird species.

A number of studies have employed remote time-lapse cameras to examine vegetative phenology (e.g. Ide & Oguma, 2013; Nagai et al., 2014). Laskin and McDermid (2016) used a network of cameras to monitor the structural changes of green-leaf understory, and reproductive phenology of Canadian buffaloberry (*Shepherdia Canadensis*) and alpine

sweetvetch (*Hedysarum alpinum*), over a single growing season in the Rocky Mountains of Alberta, Canada. Measurements made over three spatial scales – individual plants, an area with a ten metre radius, and a 250 metre-square plot – were compared to equivalent field observations (which totalled approximately 4000 over the course of the season). High agreement between measurements at all spatial scales was found, with a weighted kappa statistic showing "near perfect agreement between the two observation methods" (kappa > 0.9). The phenology of a single plant was found to be indicative of the development of the surrounding 250 metre-square plot, meaning that, with strategic placement of devices, a small camera network could provide large spatial coverage.

Alongside observation of rare or poorly-understood behaviours (for example the predation of Black-legged Kittiwake (*Rissa tridactyla*) chicks by Peregrine Falcons (*Falco peregrinus*; Collins et al., 2014) or nonbreeding strategies of Gentoo penguins (Black, Rey, & Hart, 2017), studies such as the above highlight the potential for remote camera networks in ecological research. While not suitable for all study species (for example, fast growth of certain plant species would require constant camera readjustment (Laskin & McDermid, 2016)), time-lapse cameras offer an economic method of large-scale environmental monitoring, and as techniques to aid post-hoc image processing advance, it is likely that the number of studies incorporating this technology will increase.

vii) Image data processing

The ecological imaging techniques discussed in sections *v* and *vi*, for example aerial and time-lapse photography, require analysis and annotation before they can be fully useful for research purposes. Traditionally, image processing has been carried out manually by

research teams – for example, individual birds on aerial photographs have been counted by hand using manually tiled, magnified prints (McNeill et al., 2011; Trathan, 2004). However, such annotation is labour-intensive and time-consuming, and a certain level of human error is inevitable. More recently, computer software packages with a 'counting tool' have allowed the process to be semi-automated in some cases, reducing workload and presumably error-rate. For example, Southwell & Emmerson (2015) used the open-source 'Penguin Nest Picture Analyser' software (Newbery, 2016), written in Java Script, to delineate nest zones and count adults and chicks in time-lapse photographs of Adélie penguin colonies. Slightly more complex approaches have been adopted by Trathan, (2004) and McNeill et al., (2011), both of whom employed computational techniques to automatically separate recognisable background regions (e.g. snow and rock) from penguin colonies in aerial images. Trathan (2004) applied various colour bands to scanned images in order to determine which one could best distinguish between penguins and background, and then used a coarse and fine-tuned threshold filter to eliminate the background region. He then 'sieved' the markings to discard any too large or small to be penguins, before using a 'counting tool' to generate a populate estimate. McNeill et al. (2011) also used colour properties (hue, saturation and lightness colour space) to build a semi-automated counting software in MATLAB (R2010b). The program allows a user to eliminate background regions of images, before identifying individual Adélie penguins through detection of 'regions of local brightness minima'. Researchers can then fine-tune the population estimate through deletion of obviously erroneous counts (e.g. shadows) and addition of missed birds (usually found away from, or on the periphery of, the main colony).

These semi-automated methods significantly reduce the image processing workload for research teams, while keeping error at an 'acceptable' level (McNeill et al., 2011). However, they still require a certain amount of manual annotation. Two further methods of image processing, which – once up and running – require minimal researcher input are computer vision (via machine learning) and citizen science. Machine learning approaches have enormous potential in the field of ecology (Peters et al., 2014), and state-of-the-art deep learning neural networks are already being employed to identify and count animals in time-lapse images (Arteta et al., 2016; Norouzzadeh et al., 2018). Crucially, however, these algorithms require large, detailed training sets before they can function independently. Citizen science projects can provide these, as well as being effective data-processing tools in their own right, as discussed below.

Citizen Science

'Citizen science' is the practice of actively engaging members of the general public in research. This could take the form of data collection, for example, tasking volunteers with capturing and sharing photographs of plants or animals (e.g. *Orchid Observers* (Shuttleworth et al., 2017) and *Happy Whale* (Happywhale, 2019)), or assistance with data processing (e.g. *Stall Catchers* (EyesOnALZ, 2019) and *Chimp & See* (Kalan et al., 2019; Tagg et al., 2018)). The recruitment of non-expert volunteers to aid research is not a new idea, with studies in the fields of ornithology and astronomy being prominent early examples of the concept (Dickinson et al., 2010). A Finnish investigation into the timing of avian migration began as early as 1749, involving amateur naturalists in the collection of data, and the first large-scale citizen science project is arguably the *National Audubon Society's Christmas Bird Count*, which was initiated in 1900 (Dickinson et al., 2010; Silvertown,

2009). Run annually, this census is a prime example of a study which utilises the wide geographical spread of volunteers. The potential for research projects with wide spatial breadth is undoubtedly one of the major attractions of collection-based citizen science to ecologists, particularly since phenomena such as climate change-driven migration may only become apparent at these scales (Dickinson et al., 2012).

The recent marked increase in the number of citizen science initiatives (Gura, 2013) has undoubtedly resulted from advancements in computing technology and the rise of the internet, with more than 4.33 billion people (roughly 56% of the global population) actively using the internet in July 2019 (Clement, 2019; Land-Zandstra et al., 2016). Online projects, many of which ask members of the public to help classify large datasets, have become extremely popular in recent years. For instance, the Zooniverse platform (www.zooniverse.org), which has a registered volunteer base of nearly 1.8 million people, currently hosts 104 projects, covering topics ranging from 'Space' to 'Medicine' (Zooniverse, 2019). These so-called 'virtual citizen science' projects (Jennett et al., 2014) include *Snapshot Serengeti* (Swanson et al., 2015; Swanson et al., 2016), which tasks volunteers with annotating animals in camera trap images, and *Notes from Nature* (Hill et al., 2012), which requires users to transcribe information from museum specimen labels. Such projects allow datasets to be processed at a much faster rate than could otherwise be achieved, and in some cases may even make a previously inviable project possible.

Other projects utilise smartphone technologies, asking members of the public to take and upload photographs (e.g. *Project Noah*, where volunteers share wildlife images (Project Noah, 2019)) or even use their phone as a measurement device (Newman et al., 2012). Examples of the latter include *Project NoiseTube* (NoiseTube, 2017), which aims to monitor

noise pollution, and the *iSPEX* project (iSPEX, 2019), where volunteers use an app and add-on camera filter to examine atmospheric aerosol concentrations (Land-Zandstra et al., 2016). The ability for data to be automatically geo-referenced using a mobile phone's internal GPS further increases the potential for such projects (Land-Zandstra et al., 2016).

In the field of ecology, citizen science projects have helped to identify phenological changes (for example, alterations in flowering time (Gonsamo et al., 2013) and lay date of eggs (Dunn & Winkler, 1999)), migration behaviour (Howard & Davis, 2004), geographical range shifts (Wilson et al., 2013), and population declines (Hof & Bright, 2016). However, in addition to the advancement of scientific understanding, there are public benefits to citizen science (Bonney et al., 2009; Dickinson et al., 2012). Through collaborative projects, it is likely that the relationship between scientists and the general public is strengthened, and in the case of environmental studies, it has been suggested that citizen science encourages Earth stewardship (Dickinson et al., 2012). Another key benefit is education. Through engagement in scientific research, members of the public may be introduced to the scientific method, and – if suitable materials are produced – have the opportunity to consider the theory and scientific facts behind their contribution (Dickinson et al., 2012; Masters et al., 2016). Indeed, Masters et al. (2016) found a positive correlation between level of engagement in a project and performance in a subject-specific quiz, using five Zooniverse projects as case studies. Although a causative link is difficult to establish, a study by Trumbull et al. (2000) also found that participants in a Cornell Lab of Ornithology citizen science project had begun to think in a manner similar to scientific researchers.

The motivation behind volunteers' involvement in citizen science has also been examined. In an online survey, participants of *Galaxy Zoo* – a Zooniverse project asking volunteers to

classify the shape of galaxies from images – were asked to provide their main reason for involvement (Fig. 9; Lintott et al., 2008; Raddick et al., 2013). Out of 13 possible answers, including *Learning* ("I find the site and forums helpful in learning about astronomy") and *Fun* ("I had a lot of fun categorizing the galaxies"), the most popular response was *Contribute* ("I am excited to contribute to original scientific research"; 39.8%, n = 10532).

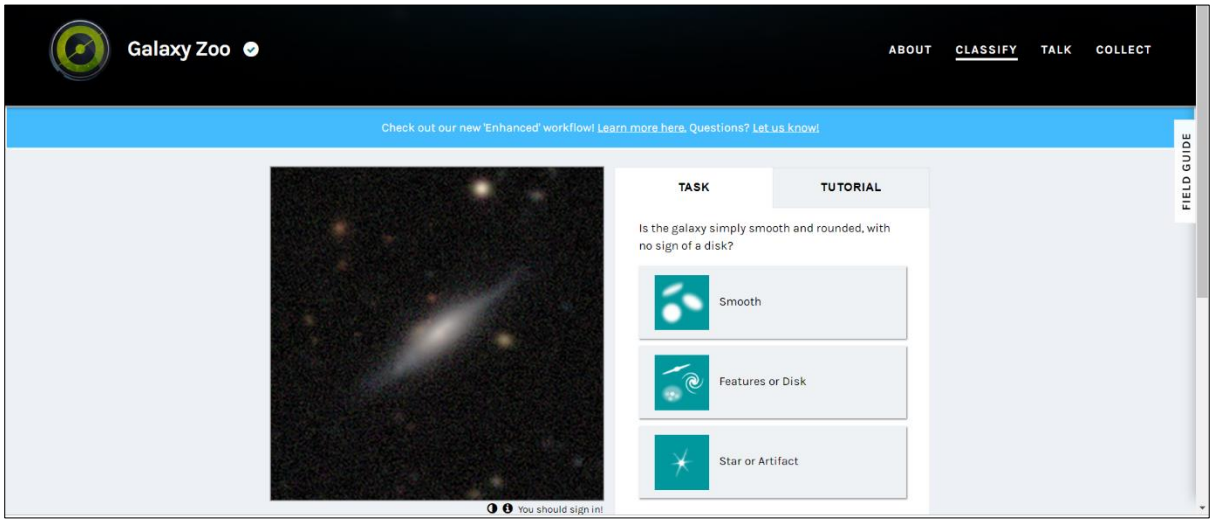


Figure 9: The *Galaxy Zoo* interface (hosted by the Zooniverse). This is an example of an online citizen science project, where volunteers are asked to perform an image annotation task. Here this involves classifying the shape of galaxies in images captured by the Dark Energy Camera Legacy Survey (DECaLS). As with all Zooniverse projects, each image will be shown to multiple volunteers before data are extracted.

Source: https://www.zooniverse.org/projects/zookeeper/galaxy-zoo/classify.[2]

This was followed by *Astronomy* ("I am interested in astronomy"; 12.4%), and *Discovery* ("I can look at galaxies that few people have seen before"; 10.4%). Though focusing on a single project, these results suggest that an opportunity to contribute to original research

[2] Screenshot printed here by permission of Prof C. Lintott, Principal Investigator, *Galaxy Zoo*.

could be of more importance to citizen scientists than the subject matter itself. Further support for this can be found when looking at volunteer traffic on the *Snapshot Serengeti* site. Out-of-season, when volunteers are informed that their classifications will not be used for research purposes, traffic on the site "drops dramatically" (C. Lintott, pers. comms.).

As mentioned above, the potential to discover something new can be an important motivator for citizen scientists, and serendipitous discoveries have indeed been made. An example of this comes from *Galaxy Zoo* (Lintott et al., 2008). While taking part in this project, a volunteer named Hanny van Arkel drew attention to an unusual looking object – later to become known as 'Hanny's Voorwerp' – which was found to be a unique example of a quasar light echo, and thus of high significance and interest to the research and volunteer communities (Christian et al., 2012; Lintott et al., 2009). Human volunteers, as opposed to machine learning algorithms, will continue to draw attention to unusual data points (Christian et al., 2012), and there is even an argument that non-expert classifiers are more likely to notice something of significance, since they have less tendency to dismiss an anomaly as error or noise.

Alongside the benefits discussed, however, the use of citizen science presents a number of challenges to the researcher. Some of these are emergent – i.e. they are connected to the way modern, often technology-dependent studies are executed – while others relate to the vast amounts of data handled throughout citizen science projects. Sampling bias may result from the methodology of a project, or the approach of volunteers. An example of the former is the *North American Breeding Bird Survey* (BBS, 2018), where citizen scientists are asked to collect data from roadsides throughout North America, thus

excluding areas away from roads from the analyses (Dickinson et al., 2010). The need for innovative statistical methods, capable of dealing with large, highly variable, and perhaps biased datasets, has led to developments in areas such as exploratory data-mining and hierarchical modelling (Dickinson et al., 2010). In addition to novel statistical approaches, methods to minimise error – such as random assessments of volunteer accuracy and clear, unambiguous protocol – will be required to maximise data reliability.

Online projects have the advantage of allowing participation from a global volunteer base, yet they prohibit individual training of citizen scientists. It is not currently understood whether training via an online tutorial is as effective as one-to-one tuition, but there is concern that a lack of clear instruction could lead to higher levels of error and bias (Dickinson et al., 2010). Indeed, obtaining a good understanding of data quality is key, especially since volunteers vary in age, ability and reason for involvement. Data validation (for example, detailed comparisons between citizen science- and expert-derived data) is essential, and findings can be used to enhance the volunteer experience, in addition to informing researchers about data quality. For instance, it may be that volunteers find one particular task within in a multi-task project particularly confusing or challenging – knowledge of this would allow the research team to adjust, or even remove, this aspect of the project, creating an easier, more enjoyable user experience and more accurate results overall.

Some projects, such as those hosted by the Zooniverse, present the same data (e.g. image or audio clip) to multiple people. This allows a consensus to be extracted from volunteers' input, and the potential to filter out unreliable results. Importantly, this practice allows people of all ages and backgrounds to take part in the project, without fear of any potential

mistakes negatively impacting the results. Regular assessment of data quality, combined with an engaging research rationale, straight-forward task, clear tutorial and dedicated team of volunteers, make for a successful online citizen science project, which can facilitate efficient and reliable processing of vast amounts of data.

viii) The geopolitics of Antarctic conservation

"*Antarctica shall be used for peaceful purposes only*"

Antarctic Treaty, Article I.

The development of the Antarctic Treaty, signed in Washington on December 1st 1959 and entering into force on June 23rd 1961, was catalysed both by growing tensions amongst nations laying claim to Antarctic territories and the success of the 1957-1958 International Geophysical Year (Australian Antarctic Division (AAD), 2016). Considered one of the "most successful international agreements" (British Antarctic Survey (BAS), 2015b), the Treaty's main objectives are to:

- Demilitarise Antarctica – "There shall be prohibited, *inter alia*, any measures of a military nature, such as the establishment of military bases…[or] the testing of any type of weapons" (Article I).
- "…promote international cooperation in scientific investigation to Antarctica" (Articles II and III).
- Set aside matters of territorial sovereignty – "No acts or activities taking place while the present Treaty is in force shall constitute a basis for asserting, supporting or denying a claim to territorial sovereignty in Antarctica or create any rights of

sovereignty in Antarctica" (Article IV). As such, the status quo is maintained; territorial claims are shown in Fig. 10.

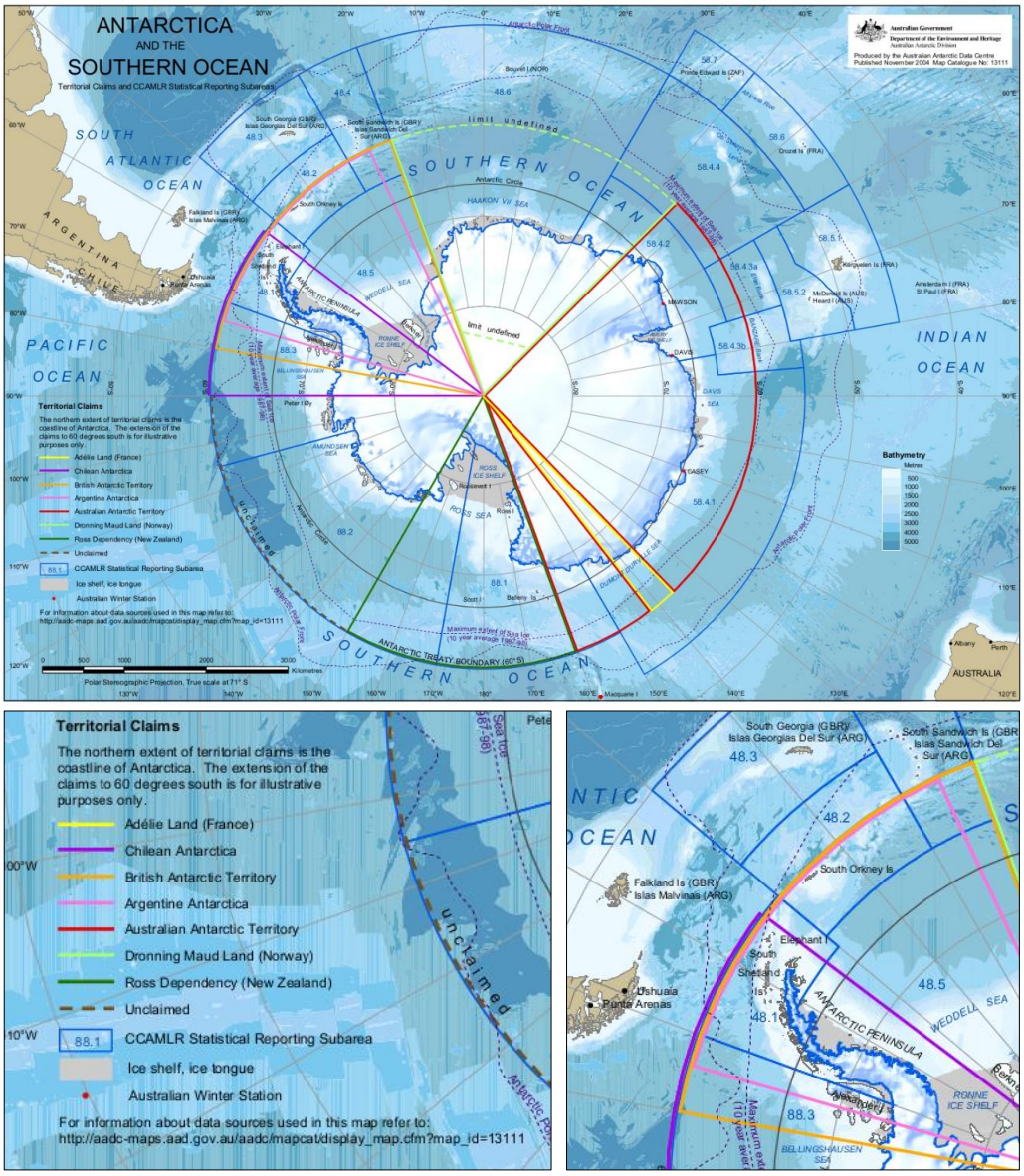


Figure 10: Claims of Antarctic territorial sovereignty made by seven nations. The list of territorial claims, and the Antarctic Peninsula and Scotia Sea region, are shown enlarged (bottom left and right, respectively).[3]

[3] Other than enlarging regions of interest, no alterations have been made to this map. Source: https://data.aad.gov.au/aadc/mapcat/display_map.cfm?map_id=13111; Department of the Environment and Heritage, Australian Antarctic Division (AAD). The map is held under a Creative Commons Attribution 4.0 Unported licence: https://creativecommons.org/licenses/by/4.0/ (it is therefore exempt from the licence covering this thesis – please refer to terms of the CC BY 4.0 licence).

The twelve original signatories of the Treaty were countries with active Antarctic scientific programs: Argentina, Australia, Belgium, Chile, France, Japan, New Zealand, Norway, South Africa, the United Kingdom, the United States of America and the Soviet Union. The Treaty can be acceded to by any member of the United Nations, and currently has 52 signatories, 28 of which are Consultative Parties (BAS, 2015).

Since the Treaty was ratified by its original signatories, a number of related agreements and organisations have been established (Fig. 11); together with the Treaty these form the Antarctic Treaty System (ATS).

One of the most notable additions is the "Protocol on Environmental Protection to the Antarctic Treaty" signed in Madrid in 1991 (giving it the alternative name of "The Madrid Protocol"). Amongst other things, this Protocol established the Committee for Environmental Protection (CEP), and placed an indefinite ban on Antarctic mining – legislation which was reaffirmed at the 2016 Antarctic Treaty Consultative Meeting, which marked the 25th anniversary of its ratification (AAD, 2019a). Also of particular note is the 1980 "Convention on the Conservation of Antarctic Marine Living Resources", which was brought into force through the "Commission for the Conservation of Antarctic Marine Living Resources" (CCAMLR) in 1982, which now has 25 full, and eight non-voting, members (AAD, 2019b; CCAMLR, 2019). The Convention aims to manage the marine living resources "south of 60° South latitude" and "between that latitude and the Antarctic Convergence which form part of the Antarctic marine ecosystem" (Convention Article I, 1980). CCAMLR adopts an ecosystem management approach, basing decisions not only on the particular species in question (with regard to, for example, harvesting), but on all species within the ecosystem (CCAMLR, 2019).

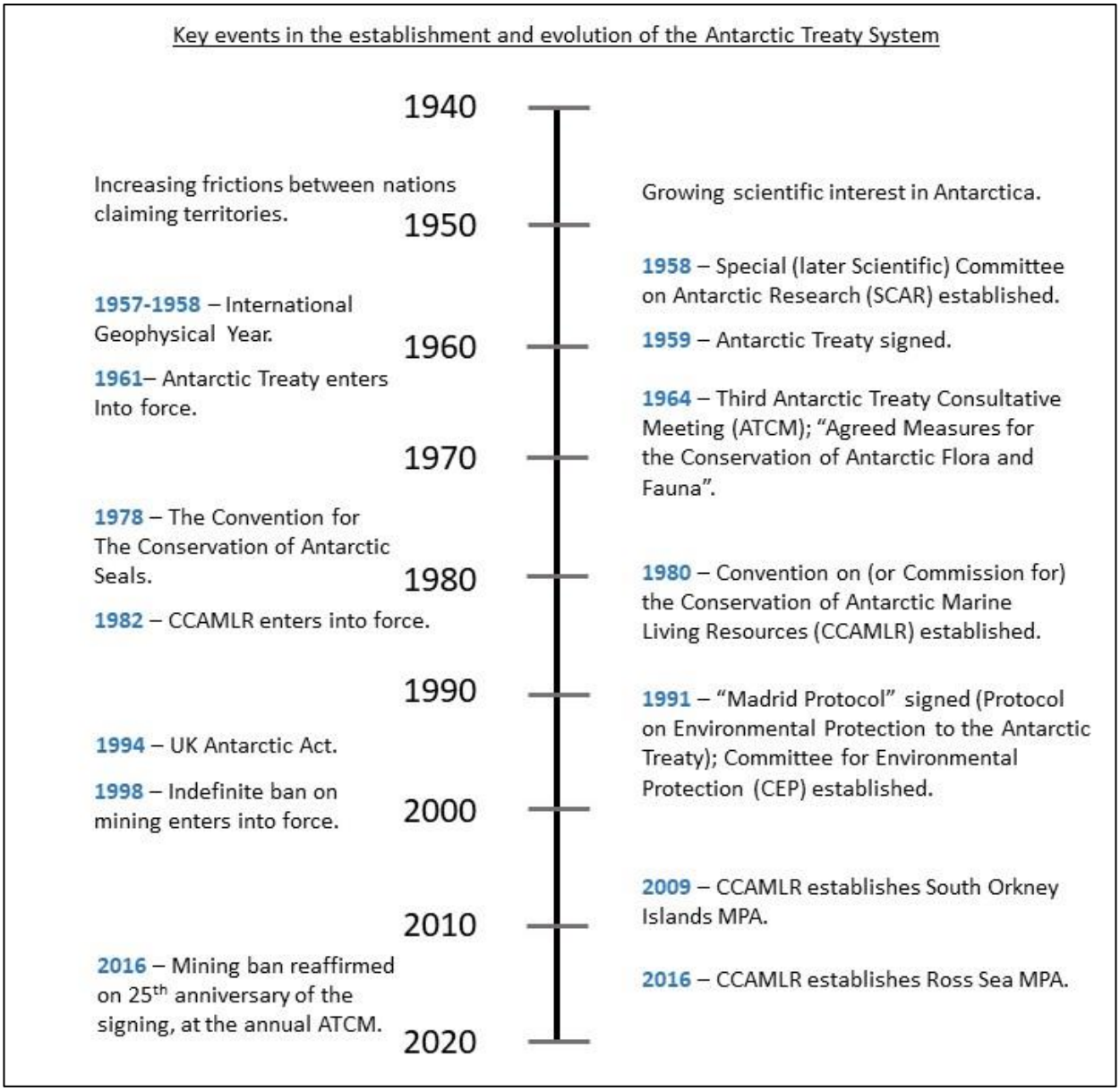


Figure 11: Timeline of key events in the history of the Antarctic Treaty System.

CCAMLR was established in an attempt to avoid a repeat of the historical cases of over-exploitation in the Southern Ocean, and prompted by the rising interest in the commercial harvesting of Antarctic krill (CCAMLR, 2019). Article II of the Convention states that, in relation to harvesting, there should be:

- "Prevention of decrease in the size of any harvested population to levels below those which ensure its stable recruitment. For this purpose its size should not be allowed to fall below a level close to that which ensures the greatest net annual increment".
- "Maintenance of the ecological relationships between harvested, dependent and related populations of Antarctic marine living resources and the restoration of depleted populations to the levels defined [above]".
- "Prevention of changes or minimisation of the risk of changes in the marine ecosystem which are not potentially reversible over two or three decades, taking into account the state of available knowledge of the direct and indirect impact of harvesting, the effect of the introduction of alien species, the effects of associated activities on the marine ecosystem and of the effects of environmental changes, with the aim of making possible the sustained conservation of Antarctic marine living resources."

Country-specific legislation relating to the Antarctic Treaty is also introduced, such as the United Kingdom's Antarctic Act (1994) and Antarctic Regulations (1995; (BAS, 2015a). For example, the Antarctic Act (1994) states that:

"No United Kingdom national (and no non-national on a British expedition) may in Antarctica–

(a) intentionally kill, injure, capture, handle or molest any native mammal or native bird,

(b) while on foot intentionally disturb a breeding or moulting native bird, or a concentration of native mammals or native birds,

(c) use a vehicle, vessel or aircraft in a manner that disturbs a concentration of native mammals or native birds, …

… (g) do anything that is likely to cause significant damage to the habitat of any native mammal, bird, plant or invertebrate,

except in accordance with a permit granted under section 12 or under the written authorisation of another Contracting Party." (Chapter 15, Part II, Section 7).

The Antarctic Treaty Consultative Meeting is held annually, and is attended not only by Treaty parties, but by associated organisations such as the Scientific Committee on Antarctic Research (SCAR), the Council of Managers of Antarctic National Programs (COMNAP), the International Association of Antarctica Tour Operators (IAATO) and the International Union for Conservation of Nature (IUCN) (BAS, 2015b).

ix) Thesis aims

As discussed in section *iv*, Southern Ocean penguins currently face numerous threats, including climate change (Ropert-Coudert et al., 2015; Youngflesh, 2018), an expanding krill fishery (Kawaguchi et al., 2009; Nicol et al., 2012), and human visitation (Bricher et al., 2008; Lynch et al., 2010b). As marine sentinels (section *i*; Boersma, 2008), the trends observed in penguin populations can be highly informative of overall ecosystem health, making large-scale monitoring projects a priority. However, as discussed, high-resolution on-ground monitoring studies are expensive and logistically challenging (section *v*). Furthermore, remote camera solutions (section *vi*) can produce an overwhelming quantity of image data, making it difficult to bridge the gap between capturing a photograph and

obtaining useful biological information. A potential solution to these challenges is the *Penguin Watch* project, which is outlined below and forms the basis of this thesis.

Penguin Watch

The *Penguin Watch* project is a monitoring initiative based in the Department of Zoology, University of Oxford. Launched in 2010, it aims to fill the data gap in penguin monitoring by positioning remote time-lapse cameras in colonies across the Scotia Arc. The network currently comprises ~80 devices, in locations primarily on the Falkland Islands, South Georgia, the South Sandwich Islands, and the Antarctic Peninsula (Fig. 12; see Appendix 2 for a complete camera list). These cameras typically take photographs once per hour during a set daytime period, and are serviced once a year during the austral summer.

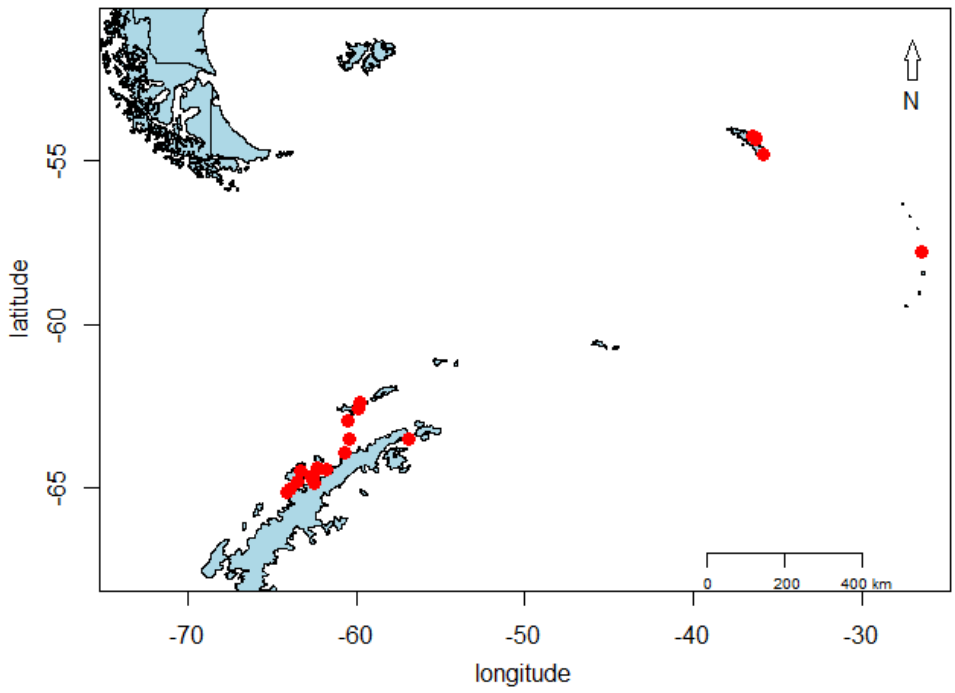


Figure 12: The *Penguin Watch* network is mainly positioned across the Falkland Islands, South Georgia, the South Sandwich Islands and the Antarctic Peninsula. The red dots show the cameras which provided data for this thesis.

Each time-lapse camera has the potential to capture and store thousands of images (e.g. Fig. 13) per year. As such, manual image processing is impossible for the small *Penguin Watch* team. Instead, we employ two alternative methods of analyses: 1) a computer vision algorithm (*Pengbot*), developed by the Visual Geometry Group (Department of Engineering, University of Oxford), and 2) the *Penguin Watch* online citizen science project (www.penguinwatch.org), launched on the Zooniverse platform in September 2014 (see section *vii* for an introduction to citizen science). To date, over 6.5 million images have been classified by nearly 50,000 registered volunteers and a wealth of anonymous participants.

I use the *Penguin Watch* network of remote time-lapse cameras to examine populations of *Pygoscelis* penguins (Adélies, Chinstraps and Gentoos) in the South Atlantic sector of the Southern Ocean. Although members of a single genus, these three species have marked differences in key life-history traits (see section *iii*), and evidence suggests that they are responding differently to climatic changes in the region. For example, populations of Gentoo penguins are currently increasing and expanding their range southward in the Antarctic Peninsula region, with breeding colonies now observed as far south as Cape Tuxen (65°16'S) (Ducklow et al., 2007; Forcada et al., 2006; Lynch et al., 2008; Lynch et al., 2010a). This is in contrast to the Adélie, which has shown marked declines in recent years, particularly in the south-western subarea of the Antarctic Peninsula (Lynch et al., 2008; Lynch et al., 2012b).



Figure 13: Example of a time-lapse photograph from the *Penguin Watch* network. This image shows a Gentoo penguin colony at George's Point, Rongé Island, Antarctic Peninsula (64°40'S; 62°40'W) on 27th December 2013 at 7pm. Penguins in these images can be counted using the *Pengbot* computer vision algorithm, or annotated via the *Penguin Watch* citizen science project (www.penguinwatch.org).

However, although Gentoos currently appear to be 'climate change winners' (Clucas et al., 2014; Hinke et al., 2007), it is difficult to predict how all three species will respond to further climatic changes, in addition to an expanding krill fishery and rising levels of tourism. It is therefore important to be able to monitor *Pygoscelis* species effectively, which includes bridging the gap between image collection and extraction of biological information.

In this thesis, I aim to comprehensively explore the use of the *Penguin Watch* online citizen science project as a data processing tool, and devise new methodologies with which to extract useful biological metrics from these data. Furthermore, I aim to use the camera network to make a contribution to our knowledge of Chinstrap phenology, and investigate the impact of key anthropogenic stressors on breeding success in Gentoo penguins. The position of this work, in the context of a broader research framework, is shown in Figure 14.

Key aims:

- To present the *Penguin Watch* camera network and citizen science project to the scientific community through peer-reviewed publications.
- To make a quantity of *Penguin Watch* photographs, citizen science classifications, metadata and population metrics freely available to the research community via open-source online databases.
- To devise new methodologies for extracting useful biological information – such as population counts, spatial distribution patterns, and phenological dates – from *Penguin Watch* citizen science data.
- To assess the reliability of the *Penguin Watch* citizen science project, and the *Pengbot* computer vision algorithm, as tools for generating count data. *How do counts derived via citizen science and computer vision compare to gold standard (i.e. expert-derived) count data? Are the same population trends observed in time-series generated using these three methods?*
- To assess the reliability of the *Penguin Watch* citizen science project for extracting phenological information from time-lapse imagery. *How do citizen science-derived*

estimates of adult arrival, chick hatch and adult departure compare to gold standard (i.e. expert-derived) dates?

- To investigate novel applications for *Penguin Watch* citizen science data. *Can citizen science data be used to help extract difficult-to-constrain dates (i.e. crèche and chick fledge)? Can citizen science data be used to detect huddling behaviour?*
- To examine Chinstrap phenology by determining key dates for a number of new colonies and seasons. *Do key dates vary substantially between sites (e.g. with longitude) and years?*
- To investigate the relationship between temperature and huddling behaviour in Chinstrap chicks. *Do chicks form tighter aggregations on colder days?*
- To investigate Gentoo breeding success through a virtual mark-recapture study, using *Penguin Watch* time-lapse imagery. *When is mortality rate greatest (i.e. during the egg, young chick, or old chick stage)? How do three key anthropogenic stressors – precipitation, krill fishing and tourism – effect chick survival? What mechanisms may explain the observed relationships?*

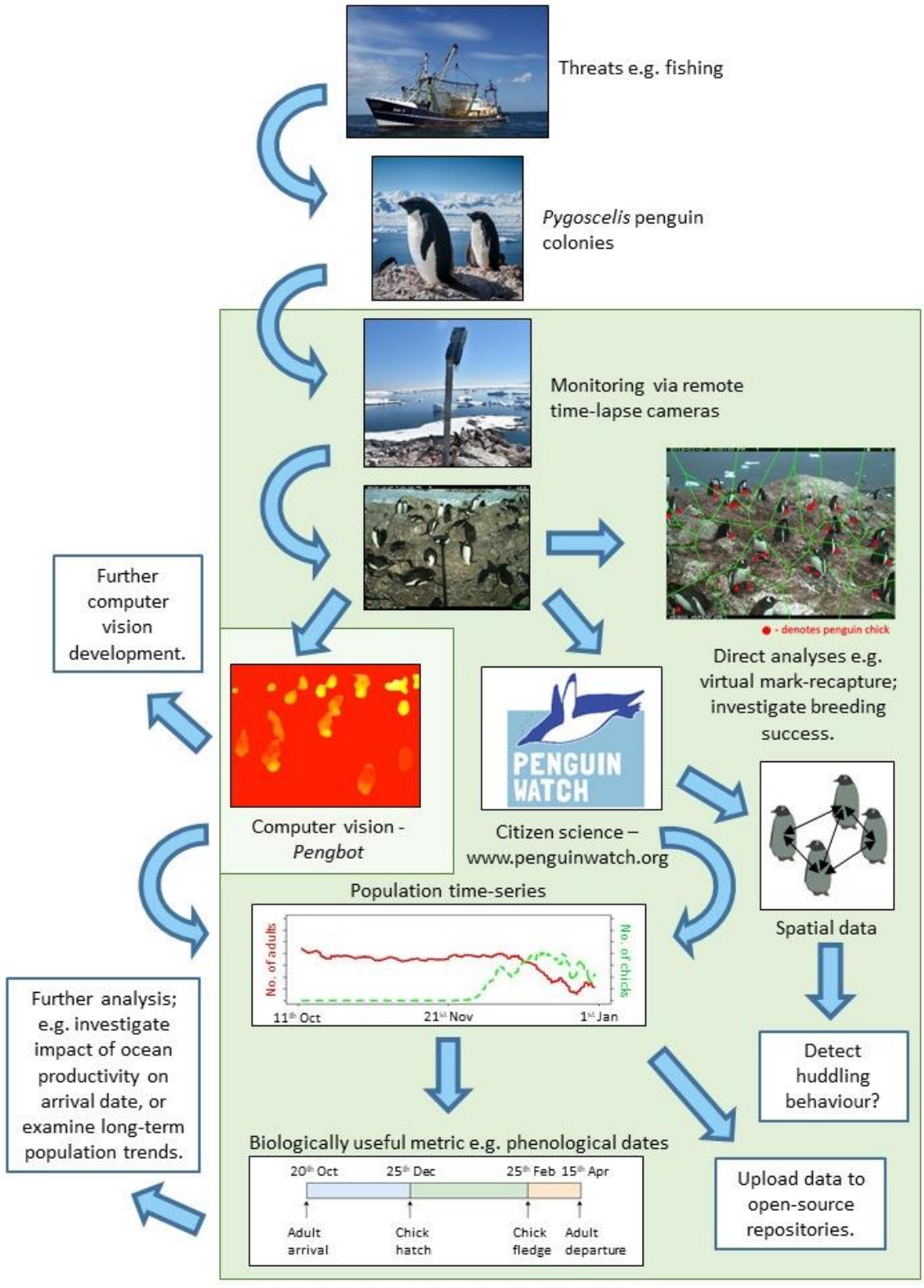


Figure 14: The context of this thesis (green box), as part of the wider research framework. Computer vision is highlighted in paler green, as the main image annotation tool used is citizen science. Fishing image: pixabay.com

x) *Ethics and permits relating to fieldwork and citizen science*

Fieldwork

Fieldwork trips took place between the following dates: 20th December 2016 to 23rd January 2017 (Antarctic Peninsula, Falkland Islands, South Georgia); 6th November 2017 to 26th November 2017 (Antarctic Peninsula); 16th October 2018 to 13th December 2018 (Antarctic Peninsula, Falkland Islands, South Georgia). These dates include travel from the United Kingdom, and time in Ushuaia, Argentina prior to sailing and between excursions.

All excursions were facilitated by a collaboration with Quark Expeditions. Fieldwork was carried out with the primary aim of adding to the *Penguin Watch* network of remote time-lapse cameras (see Figs. 8 and 15), and servicing as many devices as possible (the exact number being dependent upon weather and ice conditions). The wider *Penguin Watch* team aims to service the entire network over the course of each austral summer, with the exception of difficult-to-access locations such as the South Sandwich Islands. The camera system used in this thesis has been devised and refined from an approach used by Newbery & Southwell (2009) and Southwell & Emmerson (2015). Construction of a new camera requires two people, a camera system (e.g. a Reconyx HC500 Hyperfire Trail Camera; commercially available from Reconyx Inc., Holmen, WI, USA), and access to an appropriate area of colony (i.e. a high point, overlooking part of a colony with nests). Servicing involves retrieving the past year's data (on an SD card), changing the batteries and carrying out any necessary repairs. In some cases, the metal support structure may be modified to make it more robust. Similar cameras have been deployed in Antarctica for

roughly ten years, by five different national programs. The *Penguin Watch* team has been actively constructing cameras since the 2010/2011 season.



Figure 15: Myself and Tom Hart constructing a remote time-lapse camera on Andresen Island, Antarctic Peninsula (66°53'S; 66°40'W). Photo: D. Allcorn.

While the *Penguin Watch* network is a non-invasive monitoring strategy, any research that involves animals must adhere to strict guidelines and is subject to ethical review. Research in the Falkland Islands, South Georgia, the South Sandwich Islands and Antarctica does not require a Home Office Licence, but instead local permits must be obtained and adhered to. In the case of the Falkland Islands, South Georgia and the South Sandwich Islands, permits are granted by the appropriate local government body (and have been issued to *Penguin Watch* since 2010), while for Antarctica permits are obtained from the Foreign and Commonwealth Office (and issued to *Penguin Watch* since 2009). Each of these licences requires its own ethical review, and Antarctic research must also follow the

guidelines laid out by the Scientific Committee for Antarctic Research (SCAR), and protocols within the legislation of the Antarctic Act (1994; in particular Chapter 15, see section *viii*) and Antarctic Regulations (1995). Up to and including the 2011/2012 fieldwork season, all procedures in this thesis were covered by the Zoological Society of London's ethical review, and from the 2012/2013 season onwards have been covered by the University of Oxford's ethical review process. All data have been collected under permit (please see Appendix 3 for a list of relevant permits obtained).

Citizen science

All *Penguin Watch* volunteers freely participate in the project. Personal details are not stored (other than an IP address) unless a volunteer creates a Zooniverse account. A copy of the data and privacy agreement common to all Zooniverse projects can be found here: https://www.zooniverse.org/privacy. By registering with Zooniverse, volunteers are able to keep track of their input, as well as receiving project updates. As part of this thesis, some *Penguin Watch* data have been made publically available via the Dryad Digital Repository – all usernames were anonymised prior to upload.

A central philosophy of the Zooniverse team is that no volunteer input should be wasted. As such, volunteers begin to contribute data to a project as soon as they start classifying (i.e. without a training period). Projects also have a 'Talk' forum, managed by a team of moderators, where volunteers can ask questions and discuss interesting findings. The *Penguin Watch* 'Talk' forum is moderated by Joan Arthur, Holly Torsey and Zuzana Ročkaiová (nee Macháčková).

Chapter 2: Time-lapse imagery and volunteer classifications from the Zooniverse *Penguin Watch* project[4]

Authors

Fiona M. Jones[1], Campbell Allen[2], Carlos Arteta[3], Joan Arthur[2], Caitlin Black[1], Louise M. Emmerson[4], Robin Freeman[5], Greg Hines[2], Chris J. Lintott[2], Zuzana Macháčková[2], Grant Miller[2], Rob Simpson[6], Colin Southwell[4], Holly R. Torsey[2], Andrew Zisserman[3], Tom Hart[1]

Affiliations

1. Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, UK
2. Zooniverse, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH, UK
3. Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, UK
4. Department of the Environment and Energy, Australian Antarctic Division, 203 Channel Highway, Kingston, Tasmania 7050, Australia
5. Institute of Zoology, Wellcome Building, Outer Circle, Regent's Park, London, NW1 4RY, UK
6. Google UK, Belgrave House, 76 Buckingham Palace Road, London, SW1W 9TQ, UK

Corresponding authors: F. M. Jones (fiona.jones@zoo.ox.ac.uk)

[4] A version of this chapter is published as Jones et al., 2018 and 2019 - copyright with Authors.

T. Hart (tom.hart@zoo.ox.ac.uk)

Abstract

Automated time-lapse cameras can facilitate reliable and consistent monitoring of wild animal populations. In this report, data from 63,070 images taken by 14 different *Penguin Watch* cameras are presented, capturing the dynamics of penguin (Spheniscidae; *Pygoscelis spp.*) breeding colonies across the Antarctic Peninsula, South Shetland Islands and South Georgia (03/2012 to 01/2014). Citizen science provides a means by which large and otherwise intractable photographic datasets can be processed, and here we describe the methodology associated with the Zooniverse project *Penguin Watch*, and provide validation of the method. We present anonymised volunteer classifications for the 63,070 images, alongside the associated metadata (including date/time and temperature information). In addition to the benefits for ecological monitoring, such as easy detection of animal attendance patterns, this type of annotated time-lapse imagery can be employed as a training tool for machine learning algorithms to automate data extraction, and we encourage the use of this dataset for computer vision development.

Background and Summary

Camera-traps have proved an invaluable asset to numerous ecological studies, allowing large datasets to be obtained with relative ease and constant effort (Black et al., 2017; Holden et al., 2003; Swanson et al., 2015). Indeed, the potential applications of camera-trap research are ever-expanding, mainly owing to significant advancements in both camera technologies and the software packages available to process their output (Krishnappa & Turner, 2014; Steenweg et al., 2017). When combined with other methodologies – for example niche analysis (Pettorelli et al., 2010) – image datasets can be used to reveal fundamental information about elusive species, such as the spatial distribution of a population (Pettorelli et al., 2010), abundance estimates (Silver et al., 2004), or the nature of predation events (Bolton et al, 2007). In addition to motion-triggered camera-traps, however, it is becoming increasingly clear that there is a need for *in situ*, automated, monitoring in ecological research (Turner, 2014). When cameras are used for this monitoring, they can be programmed to capture images at regular intervals (e.g. once per hour) irrespective of whether animals are present (unlike camera-traps), and thus provide a near-continuous, regular record of occupancy and observations of a population or habitat over time (Black et al., 2016). As with camera-traps, automated time-lapse cameras have benefitted from advances in digital technology, meaning it is now possible for them to be deployed in remote areas over long periods of time (Turner, 2014). Furthermore, citizen science platforms provide a means by which large image datasets can be rapidly processed, reducing the need for time-consuming manual annotation (Swanson et al., 2015).

Due to its remote, harsh environment, large-scale on-ground monitoring studies in Antarctica are challenging and therefore rare (Lynch et al., 2012b). As such, the majority of studies investigating penguin population dynamics have been focused on a specific location, or collectively several locations, with subsequent extrapolation of local data to cover a wider region or regions (Lynch et al., 2012b). However, this approach is insufficient when it comes to understanding numerous wide-spread populations, especially since threats, such as over-fishing, may vary markedly between regions. Automated cameras overlooking penguin colonies were pioneered by Newbery & Southwell (2009), and a large-scale camera network was subsequently established in East Antarctica by the Australian Antarctic Division. The method has since been adopted by researchers from several nations (the United Kingdom (including the Oxford University project *Penguin Watch*), Argentina, Poland, Spain, Ukraine and the United States) working in other regions of Antarctica, and has received support from the Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR). There are now over 150 cameras on penguin colonies in operation on the Antarctic continent and selected Southern Ocean islands. The construction of a collaborative remote camera network and a growing number of projects in Antarctica should address the knowledge gap of limited geographic coverage for aspects of population monitoring, and thus enhance our understanding of breeding phenology and reproductive success of penguins and other seabirds. Ultimately, a greater understanding of penguin population dynamics, reproductive success and phenology will allow the impact of threats, such as climate change and over-fishing, to be monitored and disentangled, paving the way for more effective conservation measures.

An automated camera network provides a solution to the impracticalities of year-round, large-scale manned monitoring, but the wealth of photographic information produced can

be problematic for research teams to process. Citizen science can be used to dramatically increase the efficiency of data processing, thus helping to alleviate the problem of time-consuming image annotation (Swanson et al., 2015). As such, we created the website *Penguin Watch* (www.penguinwatch.org), a citizen science project hosted by the Zooniverse (www.zooniverse.org). Within *Penguin Watch*, volunteers are asked to classify images by tagging individuals and labelling them as 'adult', 'chick' or 'egg' (for penguins) or 'other' (which can be used to indicate the presence of other animals, humans and ships). This level of annotation is integral for down-stream analyses, as it allows detection of important phenological stages such as chick hatching. To date, over six million images have been classified by nearly 48,000 registered volunteers and a wealth of anonymous (unregistered) participants. This report provides a detailed overview of *Penguin Watch*, from the initial collection of data via remote cameras to the production of metadata through Zooniverse; we also present analyses of data reliability (Tables 1-3).

Citizen Science and Computer Vision

In addition to aiding a wide variety of research programmes, citizen science can be considered an effective strategy for public engagement and education (Bonney et al., 2009). In a survey carried out by the Zooniverse in 2014, 90.6% of participating volunteers stated that they "like to contribute to scientific progress," while 84.7% of people were "fascinated by the projects" in which they were involved (Homsy, 2014). Individual responses also conveyed an enthusiasm for the projects and a desire to discover something new – whether it be an animal or asteroid – was a common motivation (Homsy, 2014).

A further benefit of citizen science is that the large quantity of information provided can be used to train machine learning algorithms to carry out the task automatically (Arteta et al., 2016; Beaumont et al., 2014). Computer vision techniques provide an alternative and complementary approach to photographic data processing, and there is potential to adopt both methodologies for future analyses of *Penguin Watch* data (see Arteta et al., 2016 for a machine learning approach).

Methods

Field Methods

Layout of remote camera network

The *Penguin Watch* camera network currently comprises 91 units spread across the Falkland Islands, South Georgia and the South Sandwich Islands, and the Antarctic continent. Each camera generally captures images once per hour year-round (exceptions include summer 'foraging cameras', which take photographs once per minute), typically for a period falling between 0700 and 2000 hours minimum (Table 4 (Appendix 4)). As processing and analyses fundamentally rely on the ability to visualise penguins within images, the exact nature of the camera type, set-up and imaging frequency is flexible, with images still being comparable between studies and remaining tractable for citizen science and machine learning approaches. For example, we make the assumption that there is no substantial downstream impact of cameras being placed at different heights, angles and distances from a colony, since we are interested in the presence of individuals and – in some cases – the relative distances between them (as opposed to absolute distances).

Cameras

The images published here originate from the camera network constructed under the *Penguin Watch* program at Oxford University. All images in this Data Descriptor were captured by Reconyx HC500 Hyperfire trail cameras (commercially available; Reconyx Inc., Holmen, WI, USA). Each unit costs approximately $US 550, with an additional cost of $US 180 for an (optional) external rechargeable lead acid battery. The use of an external battery with a solar cell negates the need for annual servicing and internal battery changes. If powered internally, 12x AA lithium-ion batteries are required. Camera units have a lifespan of approximately three years before they become unreliable or show signs of wear.

Each camera is attached to a vertical metal pole, which is supported either by a wire rock basket (Fig. 1a) or by additional metal poles, which are in turn secured by rocks (Fig. 1b). In most cases, maintenance occurs once a year and involves the removal of Secure Digital (SD) cards and the insertion of new batteries. With the addition of large external power supplies, cameras could be left in place for multiple years and, if necessary, the frequency at which images are captured may also be altered to monitor fine-scale behaviour.

The use of time-lapse cameras as opposed to camera-traps fitted with passive infrared (PIR) sensors is advantageous for a number of reasons. Firstly, the PIR detector may be insufficient to pick up heat signals from distant penguin bodies, while images from time-lapse cameras can show penguins in the far distance. These individuals may or may not belong to the colony being monitored; regardless, knowledge of their presence is useful ecological information. Furthermore, it is important to obtain data from constant time intervals to perform analyses (such as mark-recapture studies) on the data. Finally, it

would be difficult to ensure that a camera-trap has enough power to continue operating throughout an entire year using a PIR sensor. Increased function in the summer months could lead to camera failure before battery replacement, meaning the colony could go unmonitored during crucial stages of their annual cycle (e.g. the arrival of adults in the spring).



Figure 1: Two examples of remote camera structures. a) A wire rock basket, covered by further rocks, is used to support the metal scaffold pole (Orne Harbour, Antarctic Peninsula); b) multiple metal "legs" are fastened to the main structure for support; each "foot" is secured using rocks (Aitcho Islands, South Shetland Islands, Antarctic Peninsula). The design shown in 'b' is favoured for future constructions as the "legs" provide increased stability and are longer-lasting than the wire used in 'a', which becomes brittle after approximately three years. The cameras shown here are powered by internal batteries. Photo credit: F. M. Jones.

Data and Analytical Methods

Analysis-ready images (i.e. renamed and resized raw files) were prepared using loops in R (R Core Team, 2016). An Exchangeable Image File (EXIF tool; sourceforge.net/p/exiftool/code/ci/master/tree/) was used to extract metadata, such as date/time information and temperature, from each photograph. The image processing script can be found on GitHub and Figshare (see 'Code Availability'). It should be noted that the images published alongside this Data Descriptor (Table 4 (Appendix 4); Data Citation 1) have already been renamed, but have not been resized – they retain their original dimensions of either 1920x1080 pixels or 2048x1536 pixels. During the resizing process, aspect ratios are maintained: 1920x1080 pixel images are resized to 1000x562.5 pixels (aspect ratio = 0.5625), and 2048x1536 pixel images are resized to 1000x750 pixels (aspect ratio = 0.75). See Data Citation 1 for the raw images and associated metadata.

Image Annotation

The *Penguin Watch* citizen science project (*v 1.0* launched September 17th 2014) is the product of a collaboration with the citizen science platform Zooniverse (www.zooniverse.org). *Penguin Watch* operates on the Ouroboros Application Programming Interface (API), is scripted in Ruby on Rails, and asks volunteers to classify images through a JavaScript interface (via a regular web browser). Citizen scientists are presented with a single image, and follow the workflow outlined in Fig. 2.

As with all Zooniverse projects, each subject is viewed by multiple volunteers in order to increase data reliability. Specifically, if *any* volunteer identifies animals in an image, that image is shown to ten people by default. If the first four volunteers state that no animals

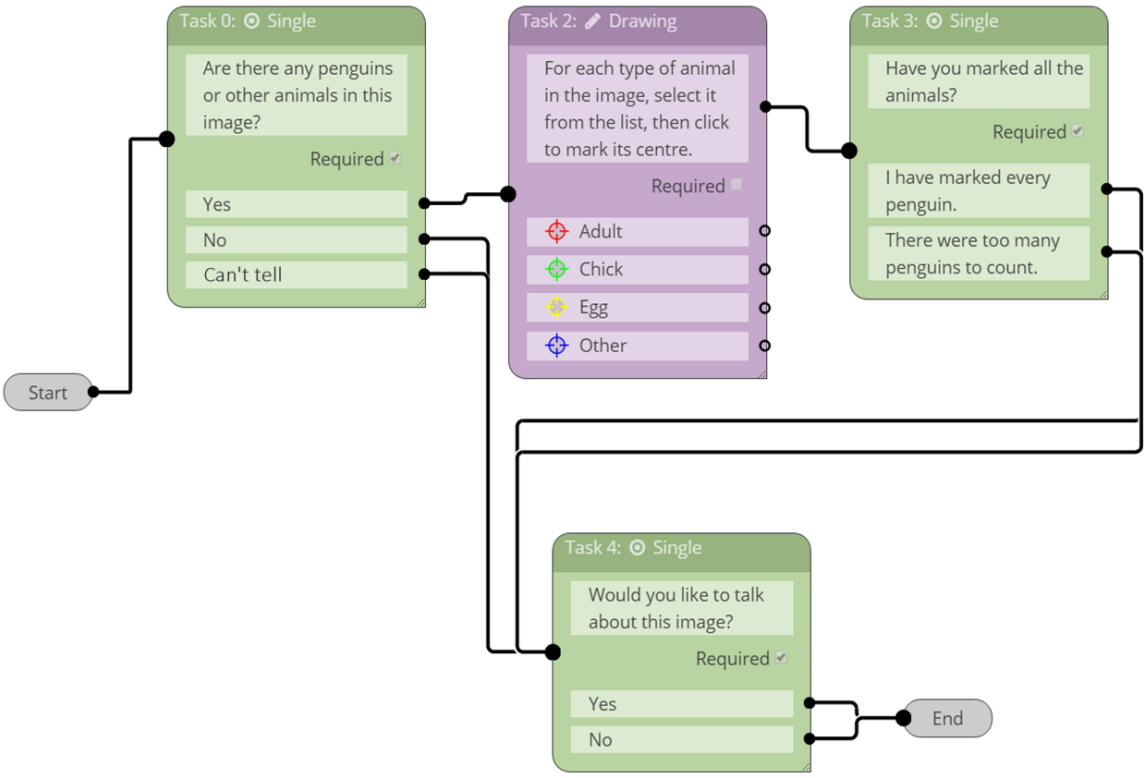


Figure 2: The *Penguin Watch* volunteer work flow. If animals are present in a given image, volunteers are asked to tag individuals by clicking on them, and classify them as 'adult', 'chick', 'egg', or 'other' (the latter can be used to identify other fauna, ships or humans). Once an image has been classified, volunteers are given the opportunity to 'talk' about it on a *Penguin Watch* forum. Green boxes indicate that volunteers must supply an answer, purple boxes indicate that a process must be carried out (such as clicking on penguins). Image source: www.zooniverse.org/lab.

are present, or they are unsure (indicated by an 'I can't tell' option), that image is then retired (assuming that nobody has identified any animals), meaning it is removed from the active dataset and not seen by any further volunteers. This practice of retiring blank images is also employed by the *Snapshot Serengeti* citizen science project (Swanson et al.,

2015; Swanson et al., 2016). Classification data are stored in MongoDB, a NoSQL database which records data in a BSON (Binary JavaScript Object Notation) format (Swanson et al., 2015). Volunteer information is listed for each annotation, as either an anonymised IP address hash (non-logged in volunteers) or a unique Zooniverse identification number (logged in volunteers). Source code for the *Penguin Watch* project can be accessed online (see 'Code Availability').

Clustering Algorithm

Since each *Penguin Watch* image is classified by multiple citizen scientists, a clustering algorithm is required to generate 'consensus click' data for each annotated object (penguin or 'other'). In other words, the coordinates of clicks made on an image (by multiple volunteers – see Fig. 3 for an example) are grouped by their spatial position, so that an average xy coordinate is produced for each tightly clustered group of markings.

This is achieved through agglomerative hierarchical clustering, using Ward's method to cluster based on the Euclidean distance between points. Agglomerative hierarchical clustering allows the user to choose the level of final clustering, and therefore the total number of clusters, using relevant data properties. Visually, this is represented by a dendrogram showing the number of clusters at each agglomerative step, with the 'leaves' of the tree being single-marking clusters, and the 'root' a cluster containing all markings. The final clusters for each image were chosen by traversing each branch of the dendrogram tree and selecting a level for that branch which maximized the number of clicks per cluster while also satisfying the requirement that no user clicked twice within the same cluster. This choice is founded in the logic that, while different volunteers' clicks on a given penguin may be shifted slightly from one another, each volunteer will click an

individual penguin only once, meaning that two separate clicks made by one person represent two penguins even if those clicks are near each other. When every marking has been assigned to a cluster, an average xy coordinate (using the median average, which is more robust to outliers than the mean average) – or 'consensus click' – can be calculated for each grouping. The aggregation script (written in Python v2.7) is publicly available (see 'Code Availability').



Figure 3: Image HALFb2013a_000051.JPG, with the 'raw clicks' of *Penguin Watch* volunteers overlaid. Each dot represents a single click, with colours specific to ten individual volunteers (images said to contain animals are shown to ten volunteers by default). Using the clustering algorithm (see 'Code Availability'), 'consensus clicks' are derived from each group of markings. The coordinates of 'raw clicks' and 'consensus clicks' can be found in the Dryad Digital Repository (Tables 5 and 6 (Appendix 4); Data Citation 1).

Output data (see Table 5 (Appendix 4)) are exported in comma-separated values (csv) file format and comprise of: 1) image name, 2) ‘consensus click’ coordinates, 3) estimated probability (P) values for adults, chicks and eggs, 4) the estimated probability (P) of a true positive, and 5) the number of markings that contributed to each classified individual (Data Citation 1). The probability value for an adult/chick/egg refers to the estimated probability that the corresponding individual is indeed an adult/chick/egg, based on the number of volunteers assigning it this label, as a proportion of the total number of clicks on that individual. Similarly, the probability of a true positive is the estimated probability that an individual is actually present at the location of the ‘consensus click’, given as the proportion of volunteers marking the area out of the total number of volunteers who indicated that animals were present in the image. Therefore, the number of volunteers who selected ‘yes’ when asked ‘Are there any penguins or other animals in this image?’ can be quickly calculated by dividing the number of markings by the probability of a true positive. The 'number of markings' is simply the number of clicks that contributed to the 'consensus click'. This allows the researcher to decide on a probability threshold for subsequent analysis (e.g. $P > 0.5$ is treated as a true positive), and to filter out classifications with less than a set number of markings. For example, if a cluster was formed from only one click, it is likely that it was made in error, and should not be carried forward.

Sometimes it is necessary to filter the raw clicks prior to performing the cluster analysis. For example, *Penguin Watch* volunteers are very occasionally shown the same image more than once. Following the logic of the aggregation algorithm (where clicks made by the same volunteer are placed in separate clusters), if duplicate markings are not removed, multiple clusters will be erroneously created, leading to an overestimate of penguin

numbers. In this dataset, 0.28% of raw clicks were not incorporated into 'consensus clicks', either owing to this duplication, or cluster failure (in rare cases clicks can be missed by the aggregation algorithm). Furthermore, when using a mobile phone to operate *Penguin Watch*, volunteers can sometimes drag the marker tool beyond the boundaries of an image. A click made in this area will be carried forward, but the coordinate values will lie beyond the limit of the image (i.e. x<0 or >1000, y<0 or >562.5 or >700) and can therefore be easily identified. These accidental markings can be filtered out prior to clustering, or – as is the case in the accompanying dataset (Data Citation 1) – can be removed during subsequent analysis (~97% of these markings are filtered out using the lowest threshold level of '>1 marking' required to form a 'consensus click').

Code Availability

All datasets are processed in R (currently v3.4.1) to produce analysis-ready image files. The 'rename and resize' script can be accessed via GitHub at https://github.com/zooniverse/Data-digging/blob/master/example_scripts/Penguin_Watch/Penguin_Watch_ImageProcessingScript.R, and a static version is archived on Figshare (Hart et al., 2017). The aggregation script (clustering algorithm) is written in Python (v 2.7) and can be found at https://github.com/zooniverse/aggregation/blob/master/penguins/aggregate.py. A static version of this script is also archived on Figshare (Hines, 2017). Source code for *Penguin Watch* is located at https://github.com/zooniverse/penguinwatch.

Data Records

A total of 63,070 raw image files retrieved from 14 *Penguin Watch* time-lapse cameras, collected between March 2012 and January 2014, are stored in the Dryad Digital Repository (Table 4 (Appendix 4); Data Citation 1). An example image can be seen in Fig. 4. The images are in JPEG format and range in number from 137 to 10,732 files per camera for a single year (Table 4 (Appendix 4)). This variation in quantity exists owing to differing



Figure 4: A sample image (DAMOa2014a_000028.JPG), taken by a Reconyx time-lapse camera at Damoy Point, Weincke Island, Antarctic Peninsula (64.82° S, 63.49° W). Date, time, moon phase and temperature information is shown at the top of the image. This is one of 63,070 photographs that can be found online in the Dryad Digital Repository (Data Citation 1).

operational durations and frequencies of image capture: for example, MAIVb2013c (137 photographs) contains hourly footage (0800-2000) from 11 days, whereas PETEc2014 (10,732 photographs) comprises images taken every half hour (0700-2130) from January 2013 to January 2014. While differing in duration, each image set is complete; i.e. no photographs have been discarded and there are no irregular temporal gaps in the datasets. Every image is uploaded to *Penguin Watch* – if visibility is poor (owing to low light levels or an obscured lens for example), volunteers can respond with 'I can't tell' when asked if any animals are present. *Penguin Watch* 'consensus click' data and metadata associated with each of the 63,070 images can also be found in the Dryad Digital Repository (Tables 5 and 6 (Appendix 4); Data Citation 1). Raw images are grouped into 22 separate folders based on camera and year. There are 32 'consensus click' folders and 32 corresponding metadata folders, as some camera/year combinations have been subdivided.

Explanation of terms

'Consensus click data'

'Consensus click data' are the data produced via the clustering algorithm discussed in 'Methods'. Each row contains the following information:

Name: Unique image reference for identification, in the format: SITExYEARx_imagenumber.csv; e.g. DAMOa2014a_000001.csv.

x_centre: x coordinate value (in pixels) for the 'consensus click' (i.e. the coordinate calculated by the clustering algorithm). The origin (point 0, 0) is located in the top left-hand corner of the image, meaning it may be necessary to reverse the y-axis of a plot in

order to overlay the ‘consensus clicks’ correctly. One coordinate value denotes one individual penguin/’other’.

y_centre: y coordinate value (in pixels) for the ‘consensus click’ (i.e. the coordinate calculated by the clustering algorithm). The origin (point 0, 0) is located in the top left-hand corner of the image, meaning it may be necessary to reverse the y-axis of a plot in order to overlay the ‘consensus clicks’ correctly. One coordinate value denotes one individual penguin/’other’.

probability_of_adult: Estimated probability that the corresponding individual is an adult – based on the number of volunteers classifying it as such, as a proportion of the total number of clicks on that individual.

probability_of_chick: Estimated probability that the corresponding individual is a chick – based on the number of volunteers classifying it as such, as a proportion of the total number of clicks on that individual.

probability_of_egg: Estimated probability that the corresponding marking indicates an egg – based on the number of volunteers classifying it as such, as a proportion of the total number of clicks on that area.

probability_of_true_positive: Estimated probability of an individual (of any ‘type’; ‘adult’, ‘chick’, ‘egg’ or ‘other’) being present at the location of the coordinates. This is the proportion of volunteers marking the penguin out of the total number volunteers who click ‘yes’ when asked ‘Are there any penguins or other animals in this image?’.

num_markings: The number of volunteer clicks that were aggregated to produce the 'consensus click' coordinate values (i.e. the number of individual clicks on a specific area of the image).

Anonymised raw classifications and metadata

These data include the coordinates for raw classifications, which provide the input data for the aggregation script (see 'Methods'). Environmental metadata, such as temperature, are also provided.

user_name (anonymised): Zooniverse unique volunteer name (or IP address), here anonymised.

subject_zooniverse_id: Unique identification code assigned to each image within Zooniverse.

lunar_phase: Moon phase when the image was captured (one of eight options: "full" (full), "new" (new), "newcres" (new crescent), "firstq" (first quarter), "waxinggib" (waxing gibbous), "waninggib" (waning gibbous), "lastq" (last quarter) or "oldcres" (old crescent)).

original_size_x: Image dimension (x axis) prior to resizing for Zooniverse; either 1920 or 2048 pixels (note: aspect ratios are maintained during resizing).

original_size_y: Image dimension (y axis) prior to resizing for Zooniverse; either 1080 or 1536 pixels (note: aspect ratios are maintained during resizing).

path: Folder pathway, which includes the image name (e.g. DAMOa/DAMOa2014a_000025).

temperature_f: Temperature (in degrees Fahrenheit, as recorded by the camera) at the time the photograph was taken.

timestamp: Time and date information for the image, in the format YYYY:MM:DD HH:MM:SS.

animals_present: Indication by volunteers that animals are present in a given image. In the format 'yes' (present), 'no' (absent) or 'can't tell'.

all_marked: Whether volunteers think that they have marked all the individuals (including all adults, chicks, eggs and others) in an image. In the format 'complete' or 'incomplete'.

value: Volunteer classification for the 'type' of individual: 'adult,' 'chick,' 'egg,' or 'other.'

x: x coordinate value (in pixels) for an individual click (note, not a 'consensus click' – see '*x_centre*'). The origin (point 0, 0) is located in the top, left-hand corner of the image.

y: y coordinate value (in pixels) for an individual click (note, not a 'consensus click' – see '*y_centre*'). The origin (point 0, 0) is located in the top, left-hand corner of the image.

Technical Validation

To examine the agreement between aggregated volunteer clicks and 'expert' annotations – and thus the reliability of *Penguin Watch* as a data processing tool – classifications made by authors CB and FMJ were employed as a 'gold standard' (GS). It should be noted that all processes (GS, citizen science (CS), and computer vision) have false positives and negatives, which we can only estimate without knowledge of the underlying truth. Analysis of the data for 100 images (spanning those published here, and the wider *Penguin Watch*

database), randomly selected from the GS (CB) classifications that stated no animals were present, revealed 96% agreement between the GS and aggregated volunteer clicks (i.e. 96% were marked as containing no animals). Of the remaining 4%, three images had been erroneously marked by volunteers, and one image contained animals which had been missed in the GS (albeit there were also six erroneous volunteer classifications in this image). In all cases where there were false positives, only one volunteer had marked the image. Thus, if a minimum threshold for *num_markings* were employed (> 1 marking or above), each of these would be filtered out during subsequent analyses.

Image classifications from four cameras – DAMOa, HALFc, LOCKb, and PETEc (PETEc2013) – were also analysed. These cameras were selected as they represent a range of 'views' (i.e. the cameras are positioned at different angles and distances from a colony), and monitor each of the three *Pygoscelis* species – *Pygoscelis papua* (Gentoo; DAMOa, LOCKb), *P. antarcticus* (Chinstrap; HALFc), and *P. adeliae* (Adélie; PETEc). GS (FMJ) classifications fulfilling two criteria – 1) the image contained animals, and 2) images were marked as complete – were selected. A random sample of 300 images was obtained for each camera, with the exception of HALFc, where the whole sample of 283 photographs was used (see Table 4 (Appendix 4)).

Number of adults, number of chicks, and number of adults and chicks combined, as calculated using *Penguin Watch* data, were compared to the GS counts. Probability thresholds for adults and chicks were set to P>0.5 for this analysis (see 'Explanation of terms'). Four different threshold levels for *num_markings* were applied to the *Penguin Watch* classifications. For example, a threshold level of greater than two (>2) requires three or more *Penguin Watch* participants to click on a 'penguin' for it to be counted.

Mean average differences, measured in number of penguins, between the GS count and the count produced using each threshold level, are shown in Tables 1 and 2. The proportion of differences that were either zero (i.e. the GS and *Penguin Watch* data showed the same result) or one are also shown for each category.

The results show that accuracy – where higher accuracy is defined as a smaller mean difference between the GS and CS data – varies depending on site, threshold level, and whether adults and chicks are considered together or separately. However, the most accurate results for adults at all four sites were obtained using a threshold level of >3, where the mean average differences ranged from 0.88 (σ = 1.34) to 2.36 (σ = 2.69) penguins (Table 1). At this threshold level, the number of false positives and false negatives is at a minimum. If the threshold level is increased further, the proportion of overestimated values continues to decrease (as more clicks are required for an 'individual' to be counted), but the proportion of underestimated values increases (Table 3). For the sites examined here, a threshold of >3 reflects an appropriate balance between over- and under-estimation. Since the four sites vary in camera angle and species, this threshold level would be a suitable starting point when analysing data from different cameras.

The greatest deviation from the GS (with a mean difference of 2.36 adults) is associated with PETEc. This camera captured the greatest number of individuals on average (a mean of 36.43 (2013 image set) compared to 11.79, 15.03 and 10.99 at DAMOa, HALFc and LOCKb, respectively). Greater error with increasing colony size is expected, and suggests that citizen scientists have more difficulty distinguishing individuals when presented with larger groups. Segmentation of the image to focus on a region of interest may offer a

	Threshold employed							
	Only **adult** classifications				All **adult** and **chick** classifications			
	>1	>2	>3	>4	>1	>2	>3	>4
DAMOa	n = 300				n = 300			
Mean average difference	1.88	1.50	1.18	1.24	1.94	1.58	1.37	1.53
σ	2.15	1.78	1.56	1.40	2.22	1.86	1.69	1.61
Proportion of differences that are 0 or 1	0.56	0.67	0.75	0.69	0.56	0.64	0.69	0.62
HALFc	n = 283				n = 283			
Mean average difference	1.28	0.99	0.88	0.99	1.45	1.27	1.34	1.55
σ	2.06	1.58	1.34	1.28	2.08	1.71	1.64	1.72
Proportion of differences that are 0 or 1	0.74	0.82	0.85	0.82	0.69	0.71	0.68	0.64
LOCKb	n = 300				n = 300			
Mean average difference	1.41	1.25	1.17	1.19	1.40	1.24	1.18	1.29
σ	2.31	2.12	1.98	1.96	2.30	1.89	1.74	1.78
Proportion of differences that are 0 or 1	0.73	0.77	0.77	0.75	0.72	0.73	0.73	0.68
PETEc	n = 300				n = 300			
Mean average difference	3.60	2.48	2.36	3.27	4.29	3.54	4.37	6.06
σ	3.50	2.99	2.69	2.86	4.51	4.07	4.41	5.20
Proportion of differences that are 0 or 1	0.28	0.46	0.46	0.30	0.25	0.37	0.33	0.20

Table 1: Comparison of counts between *Penguin Watch* (CS) and GS data for four cameras: DAMOa, HALFc, LOCKb and PETEc. 'Threshold employed' relates to the filtering threshold applied during analysis. For example, a threshold of '>2' means that at least three people must have marked an area before it is counted as a penguin. 'Average difference' is the mean average of all differences between the GS and *Penguin Watch* counts, measured in number of individuals. 'Proportion of differences that are 0 or 1' is the proportion of images for which the *Penguin Watch* count was either equal to the GS, or different by one individual (either higher or lower), respectively. The results for adults, and adults and chicks combined, are presented here. To be counted as an adult or a chick, *probability_of_adult* or *probability_of_chick* was >0.5 (see 'Explanation of terms').

	Threshold employed							
	Only **chick** classifications (all images)				Only **chick** classifications			
	>1	>2	>3	>4	>1	>2	>3	>4
DAMOa	n = 300				n = 60			
Mean average difference	0.25	0.27	0.29	0.33	1.25	1.33	1.45	1.65
σ	0.68	0.71	0.75	0.85	1.04	1.05	1.06	1.20
Proportion of differences that are 0 or 1	0.94	0.93	0.92	0.91	0.68	0.67	0.60	0.53
HALFc	n = 283				n = 109			
Mean average difference	0.46	0.51	0.58	0.65	1.18	1.31	1.50	1.69
σ	0.89	0.99	1.11	1.24	1.09	1.22	1.35	1.50
Proportion of differences that are 0 or 1	0.90	0.87	0.83	0.82	0.73	0.65	0.57	0.54
LOCKb	n = 300				n = 122			
Mean average difference	0.94	0.90	0.88	0.92	2.32	2.22	2.16	2.25
σ	2.21	2.08	2.07	2.12	2.97	2.79	2.79	2.84
Proportion of differences that are 0 or 1	0.81	0.81	0.83	0.80	0.53	0.53	0.59	0.52
PETEc	n = 300				n = 179			
Mean average difference	1.89	2.07	2.49	3.06	3.17	3.46	4.17	5.12
σ	2.69	2.69	3.12	3.70	2.85	2.71	3.06	3.51
Proportion of differences that are 0 or 1	0.60	0.55	0.52	0.49	0.33	0.25	0.20	0.15

Table 2: Comparison of chick counts between *Penguin Watch* (CS) and GS data for four cameras: DAMOa, HALFc, LOCKb and PETEc; see Table 1 legend for definitions. The first column shows the results for all images in the sample, including those where – according to the GS and CS data – no chicks were present. The second column presents the results for the sample of images where chicks were present, according to the GS and/or CS classifications. When extracting *Penguin Watch* data, *probability_of_chick* was set to >0.5 (see 'Explanation of terms').

partial solution to this problem, although obscuration of individuals by their conspecifics remains an issue.

The accuracy of chick counts (Table 2) – where lowest mean average differences range from 0.25 (σ = 0.68) to 1.89 (σ = 2.69), and the proportion of differences equating to zero or one is as high as 94% for DAMOa (>1) – appears to be greater than that of the adult counts. However, this is owing to the large number of images that do not contain chicks (according to the GS and CS data). If these images are removed, the accuracy decreases, with the lowest values at each site ranging from 1.18 (σ = 1.09) to 3.17 (σ = 2.85).

One explanation for the lower accuracy of chick estimation compared to adults could be misinterpretation of individuals (i.e. labelling a chick as an adult, or vice versa). This could occur when chicks are losing their down feathers, and thus look similar to adult birds. However, if this was the main issue, accuracy levels should be optimised when adult and chick counts are combined, which is not the case (Table 1). Instead, the data suggest that chicks are often missed by volunteers. While overestimation is more common than underestimation at the >1 and >2 threshold level in adults, the proportion of underestimated values consistently outweighs the proportion of overestimates at each threshold level for chick counts, as shown in Fig. 5 and Table 3.

Chicks are often partially concealed by a parent, particularly during the brood-guard stage, which may explain why they are sometimes missed. Increased training on the *Penguin Watch* interface – for example through the addition of imagery showing very young chicks, and species-specific information on expected lay dates – may help to overcome this issue. It should also be noted that different classification thresholds may be appropriate for adults and chicks. For example, chick accuracy was optimised at the >1 level at the DAMOa,

HALFc and PETEc sites. Where chicks are often missed, it is logical that fewer volunteers should be required to tag them in order for them to be recognised.

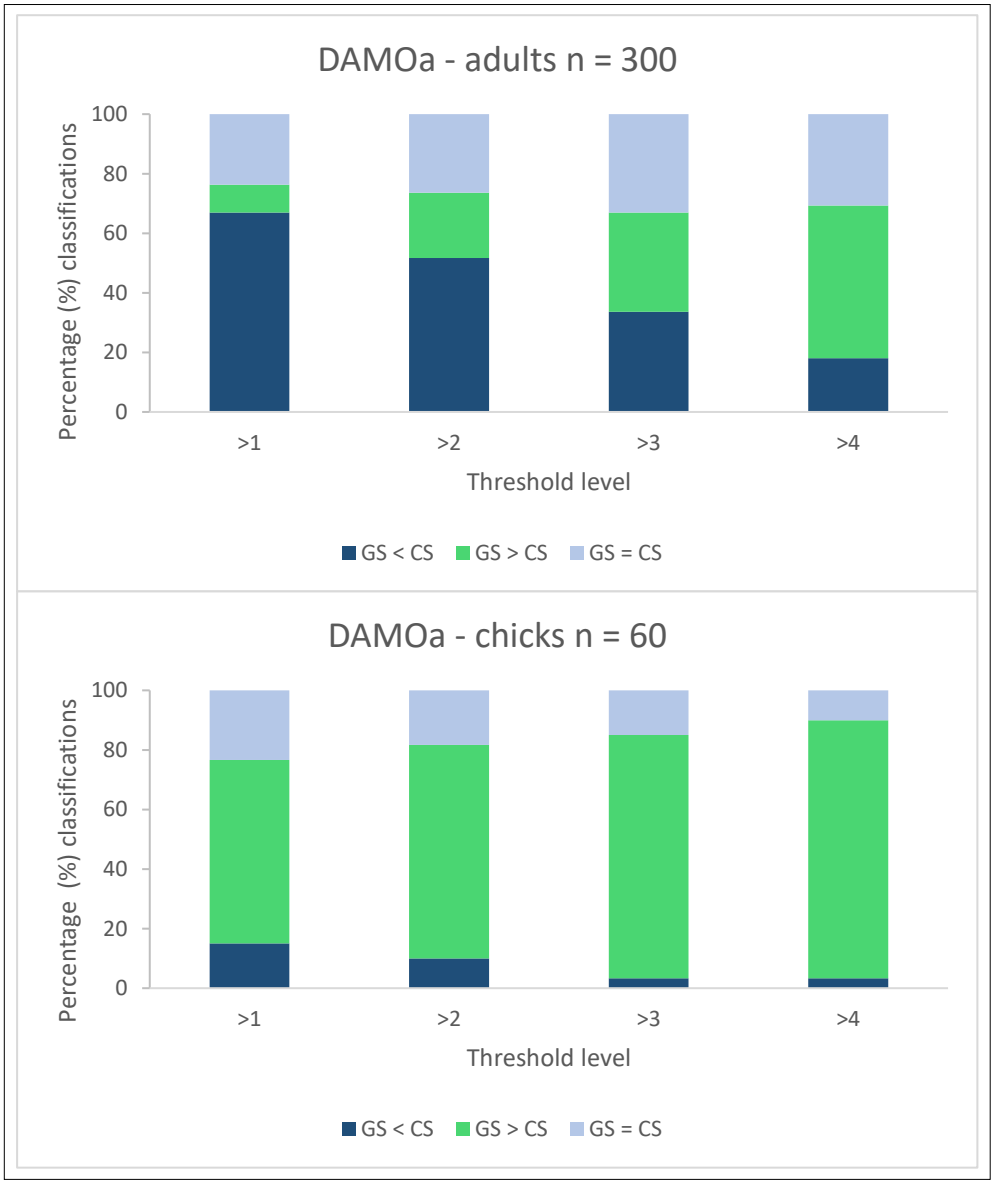


Figure 5: Percentage of *Penguin Watch* (CS) classifications that are greater than (GS < CS), less than (GS > CS) or equal to (GS = CS) gold standard (GS) classifications (i.e. overestimates, underestimates or matches) for DAMOa. Top: Adult classifications only (n=300); bottom: chick classifications only (for images where chicks are present according to GS and/or CS data; n=60).

	Threshold employed							
	Only **adult** classifications				Only **chick** classifications			
	>1	>2	>3	>4	>1	>2	>3	>4
DAMOa	n = 300				n = 60			
GS < CS (%)	67.00	51.67	33.67	18.00	15.00	10.00	3.33	3.33
GS > CS (%)	9.33	22.00	33.33	51.33	61.67	71.67	81.67	86.67
GS = CS (%)	23.67	22.33	33.00	30.67	23.33	18.33	15.00	10.00
HALFc	n = 283				n = 109			
GS < CS (%)	48.06	35.34	12.61	14.49	17.43	11.01	2.75	0.92
GS > CS (%)	12.72	20.14	32.16	48.76	57.80	63.30	73.39	79.82
GS = CS (%)	39.22	44.52	45.23	36.75	24.78	25.69	23.85	19.27
LOCKb	n = 300				n = 122			
GS < CS (%)	48.67	37.67	27.33	22.33	27.05	23.77	17.21	9.84
GS > CS (%)	11.00	18.33	26.67	32.00	51.64	54.92	63.11	66.39
GS = CS (%)	40.33	44.00	46.00	45.67	21.31	21.31	19.67	23.77
PETEc	n = 300				n = 179			
GS < CS (%)	79.33	55.33	26.33	13.67	32.96	20.67	10.06	6.70
GS > CS (%)	13.00	30.67	54.00	77.00	58.66	72.07	83.24	91.60
GS = CS (%)	7.67	14	19.67	9.33	8.38	7.26	6.70	1.68

Table 3: Percentage of *Penguin Watch* (CS) classifications that are greater than (GS < CS), less than (GS > CS) or equal to (GS = CS) gold standard (GS) classifications (i.e. overestimates, underestimates or matches) for DAMOa, HALFc, LOCKb and PETEc. Results are shown for adult classifications (all images) and chick classifications (only images where chicks are present, according to GS and/or CS data).

A second reason for the underestimation of chicks could stem from the *Penguin Watch* work-flow. After a volunteer has classified 30 individuals, they are given the option of

moving on to the next image, to maintain interest. The results suggest that this is not detrimental to the accuracy of adult counts, with the combined efforts of each volunteer covering an entire image (counts were overestimated in 79.33% of cases at PETEc (>1 threshold), where the number of individuals frequently exceeds 30). However if the majority of volunteers choose to tag all the adults first, before classifying any chicks, chicks may be consistently missed in images containing >30 penguins. An alteration to the *Penguin Watch* tutorial, requesting that volunteers classify a variety of penguin 'types' before moving onto a new image, may help to overcome this.

Through careful selection of threshold levels, consideration of associated error, and continued evaluation of data quality, *Penguin Watch* can provide us with a reliable means with which to process large quantities of imagery, which would otherwise be intractable.

Acknowledgements

The authors express their gratitude to the Zooniverse team members, the Zooniverse partners (see https://www.penguinwatch.org/#/team) and the 47,961 registered *Penguin Watch* volunteers. The authors also wish to thank Quark Expeditions, Cheeseman's Ecology Safaris and Oceanwide Expeditions for their assistance with fieldwork logistics, and Brooke Simmons for helpful comments on the manuscript. Zooniverse and *Penguin Watch* were supported by grants from the Alfred P. Sloan Foundation and a Google Global Impact Award. This work was supported by NERC, the Darwin Initiative and the Citizen Science Alliance, as well as individual contributors to *Penguin Watch*, particularly via donations during Quark Expeditions' voyages. Work by Australian Antarctic Division researchers to establish a camera network in East Antarctica was carried out under AAS

projects 2722 and 4088. Fiona Jones and Tom Hart had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Author Contributions

FMJ wrote the paper, with input from TH; TH and AZ initiated the *Penguin Watch* project; Campbell Allen, GM, RS and CL designed *Penguin Watch*, with input from AZ and Carlos Arteta on relevance for computer vision; CB helped design *Penguin Watch* and provided gold standard annotations; GH wrote the clustering algorithm; JA, ZM and HT managed the *Penguin Watch* 'Talk' forum, answering volunteers' questions; CS and LME provided imagery to test the *Penguin Watch* interface. All authors reviewed and approved the manuscript.

Competing Interests

The authors declare no competing financial interests.

References

Arteta, C., Lempitsky, V., & Zisserman, A. (2016). Counting in the Wild. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer Vision - ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9911* (pp. 483–498). https://doi.org/https://doi.org/10.1007/978-3-319-46478-7_30

Beaumont, C. N., Goodman, A. A., Kendrew, S., Williams, J. P., & Simpson, R. (2014). The Milky Way Project: leveraging citizen science and machine learning to detect

interstellar bubbles. *The Astrophysical Journal Supplement Series*, *214*(1), 3. https://doi.org/10.1088/0067-0049/214/1/3

Black, C., Collen, B., Johnston, D., & Hart, T. (2016). Why Huddle? Ecological Drivers of Chick Aggregations in Gentoo Penguins, *Pygoscelis papua*, across Latitudes. *Plos One*, *11*(2), e0145676. https://doi.org/10.1371/journal.pone.0145676

Black, C., Rey, A. R., & Hart, T. (2017). Peeking into the bleak midwinter: investigating nonbreeding strategies of Gentoo Penguins using a camera network. *The Auk*, *134*(3), 520–529. https://doi.org/10.1642/AUK-16-69.1

Bolton, M., Butcher, N., Sharpe, F., Stevens, D., & Fisher, G. (2007). Remote monitoring of nests using digital camera technology. *Journal of Field Ornithology*, *78*(2), 213–220. https://doi.org/10.1111/j.1557-9263.2007.00104.x

Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009). Citizen science: a developing tool for expanding science knowledge and scientific literacy. *BioScience*, *59*(11), 977–984. https://doi.org/10.1525/bio.2009.59.11.9

Hart, T., Jones, F. M., & Freeman, R. (2017). Penguin Watch Image Processing Script (R). figshare. Retrieved from https://doi.org/10.6084/m9.figshare.5432974.v2

Hines, G. (2017). Penguin Watch Aggregation Script (Python). figshare. Retrieved from https://doi.org/10.6084/m9.figshare.5472544.v1

Holden, J., Yanuar, A., & Martyr, D. J. (2003). The Asian Tapir in Kerinci Seblat National Park, Sumatra: evidence collected through photo-trapping. *Oryx*, *37*(01), 34–40. https://doi.org/10.1017/S0030605303000097

Homsy, V. (2014). *Engagement in the Zooniverse*. University of Oxford.

Krishnappa, Y. S., & Turner, W. C. (2014). Software for minimalistic data management in large camera trap studies. *Ecological Informatics*, *24*, 11–16. https://doi.org/10.1016/j.ecoinf.2014.06.004

Lynch, H. J., Naveen, R., Trathan, P. N., & Fagan, W. F. (2012). Spatially integrated assessment reveals widespread changes in penguin populations on the Antarctic

Peninsula. *Ecology*, *93*(6), 1367–1377. https://doi.org/10.1890/11-1588.1

Newbery, K. B., & Southwell, C. (2009). An automated camera system for remote monitoring in polar environments. *Cold Regions Science and Technology*, *55*(1), 47–51. https://doi.org/10.1016/j.coldregions.2008.06.001

Pettorelli, N., Lobora, A. L., Msuha, M. J., Foley, C., & Durant, S. M. (2010). Carnivore biodiversity in Tanzania: revealing the distribution patterns of secretive mammals using camera traps. *Animal Conservation*, *13*(2), 131–139. https://doi.org/10.1111/j.1469-1795.2009.00309.x

R Core Team. (2016). *R: A language and environment for statistical computing.* Retrieved from https://www.r-project.org/

Silver, S. C., Ostro, L. E. T., Marsh, L. K., Maffei, L., Noss, A. J., Kelly, M. J., … Ayala, G. (2004). The use of camera traps for estimating jaguar *Panthera onca* abundance and density using capture / recapture analysis. *Oryx*, *38*(2), 148–154. https://doi.org/10.1017/S0030605303000000

Steenweg, R., Hebblewhite, M., Kays, R., Ahumada, J., Fisher, J. T., Burton, C., … Rich, L. N. (2017). Scaling up camera traps: monitoring the planet's biodiversity with networks of remote sensors. *Frontiers in Ecology and the Environment*, *15*(1), 26–34. https://doi.org/10.1002/FEE.1448

Swanson, A., Kosmala, M., Lintott, C., & Packer, C. (2016). A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. *Conservation Biology*, *30*(3), 520–531. https://doi.org/10.1111/cobi.12695

Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A., & Packer, C. (2015). Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Scientific Data*, *2*, 150026. https://doi.org/10.1038/sdata.2015.26

Turner, W. (2014). Sensing biodiversity. *Science*, *346*(6207), 301–302. https://doi.org/10.1126/science.1256014

Data Citation

1. Jones, F.M., Allen, C., Arteta, C., Arthur, J., Black, C., Emmerson, L.M., Freeman, R., Hines, G., Lintott, C.J., Macháčková, Z., Miller, G., Simpson, R., Southwell, C., Torsey, H.R., Zisserman, A., & Hart, T. *Dryad Digital Repository* doi:10.5061/dryad.vv36g (2018). Revised July 2019.

Chapter 3: Processing citizen science- and machine-annotated time-lapse imagery for biologically meaningful metrics[5]

Authors

Fiona M. Jones[1], Carlos Arteta[2], Andrew Zisserman[2], Victor Lempitsky[3], Chris J. Lintott[4], Tom Hart[1]

Affiliations

1. Department of Zoology, University of Oxford, 11a Mansfield Road, Oxford, OX1 3SZ, UK
2. Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK
3. Samsung AI Center, Butyrskiy Val Ulitsa, 10, Moscow, Russia, 125047 & Skolkovo Institute of Science and Technology (Skoltech), Bolshoy Boulevard 30, bld. 1, Moscow, Russia 121205
4. Zooniverse, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH, UK

Corresponding authors: F. M. Jones (fiona.jones@zoo.ox.ac.uk)

T. Hart (tom.hart@zoo.ox.ac.uk)

[5] A version of this chapter was in review at the time of thesis submission (Jones et al. in review) and is now published as Jones, F.M., Arteta, C., Zisserman, A., Lempitsky, V., Lintott, C.J. & Hart, T. (2020). Processing citizen science- and machine-annotated time-lapse imagery for biologically meaningful metrics. *Scientific Data, 7*(1), 1–15. Copyright with Authors.

Abstract

Time-lapse cameras facilitate remote and high-resolution monitoring of wild animal and plant communities, but the image data produced require further processing to be useful. Here we publish pipelines to process raw time-lapse imagery, resulting in count data (number of penguins per image) and 'nearest neighbour distance' measurements. The latter provide useful summaries of colony spatial structure (which can indicate phenological stage) and can be used to detect movement – metrics which could be valuable for a number of different monitoring scenarios, including image capture during aerial surveys. We present two alternative pathways for producing counts: 1) via the Zooniverse citizen science project *Penguin Watch* and 2) via a computer vision algorithm (*Pengbot*), and share a comparison of citizen science-, machine learning-, and expert-derived counts. We provide example files for 14 *Penguin Watch* cameras, generated from 63,070 raw images annotated by 50,445 volunteers. We encourage the use of this large open-source dataset, and the associated processing methodologies, for both ecological studies and continued machine learning and computer vision development.

Background and Summary:

Seabird population changes are considered a reflection of changes within the marine environment, making seabird species key indicators of ecosystem health (Croxall et al., 2012; Paleczny et al., 2015). The penguins (Family: Spheniscidae) are amongst the most threatened, facing environmental stressors such as invasive species, overfishing, pollution and climate change (Croxall et al., 2012; Paleczny et al., 2015; Trathan et al., 2015). In the Southern Ocean, two key threats – sea ice change (Ducklow et al., 2013; Stammerjohn et al., 2012) and an expanding krill fishery (Nicol et al., 2012) – have the potential to act synergistically, with a longer ice-free season facilitating a longer harvest (Kawaguchi et al., 2009). In order to implement effective mitigation strategies, rigorous monitoring of penguin colonies is required. However, the challenges of on-ground monitoring in Antarctica (logistical, practical and financial) have left gaps in data surveys, and the majority of monitoring has occurred on a small scale (Lynch et al., 2012).

Time-lapse cameras offer a solution to these challenges, and Newbery and Southwell (2009) pioneered the use of automated cameras for penguin monitoring. A large, collaborative network of remote time-lapse cameras now exists in Antarctica, and in Jones et al. (2018) we discuss the *Penguin Watch* project, which currently operates over 90 cameras in the region (Jones et al., 2018). These cameras are located in the Falkland Islands, South Georgia, the Antarctic Continent and the South Sandwich Islands, and at a minimum usually capture images once per hour year-round between the hours of 0700 and 2000 (Jones et al., 2018). The camera units (Reconyx HC500 Hyperfire Trail Cameras) are serviced once per year if conditions allow, and are powered either by 12xAA lithium

ion batteries or an external rechargeable battery connected to a solar cell (Jones et al., 2018).

Owing to the wealth of image data collected, expert manual annotation is unfeasible. Instead, two parallel methods have been employed to efficiently identify penguins in photographs, using – 1) citizen science (Jones et al., 2018), and 2) computer vision (Arteta et al., 2016). The *Penguin Watch* citizen science project (www.penguinwatch.org), hosted by the Zooniverse online platform (www.zooniverse.org), was launched in September 2014. The project tasks volunteers with identifying individual penguins in randomly presented time-lapse images, and categorising them as either an 'adult', 'chick' or 'egg' (Jones et al., 2018). There is also the option to assign an individual or object to 'other', which allows classification of other fauna, humans, or ships (Jones et al., 2018). To date, over 6.5 million classifications have been made by roughly 50,000 registered volunteers and numerous anonymous (unregistered) participants.

In Jones et al. (2018) we describe a dataset of 63,070 raw *Penguin Watch* time-lapse images from 14 different cameras, which is stored in the Dryad Digital Repository (Data Citation 1; please also see Jones et al. (2019) – an erratum correcting for a duplicate dataset). Also in the repository are metadata (such as date/time and temperature information), raw anonymised citizen science classifications (i.e. xy coordinate data for each 'click' on a photograph), and 'consensus click' data, for each image. The latter are produced via a clustering algorithm (Hines, 2017; Jones et al., 2018) – multiple volunteers (a total of ten if the photograph contains animals) annotate each image, meaning a 'consensus click' (average xy coordinate) must be generated for each penguin.

This publically available data collection (Data Citation 1) serves as a resource for ecological researchers, comprising images and citizen science annotations for multiple penguin colonies between March 2012 and January 2014 (Jones et al., 2018). In addition, *Penguin Watch* volunteer clicks have been employed to train the machine-learning algorithm *Pengbot* – the second annotation method – which detects penguins in raw images using pixel densities (Arteta et al., 2016).

Whichever approach is adopted, the data must be further processed in order to extract biologically meaningful metrics such as population counts and measures of movement. Here we present a second open-source database in the Dryad Digital Repository (Data Citation 2), which contains the following files for each of the 14 cameras described in Jones et al. (2018):

1) using citizen science: *Kraken Files* (filtered 'consensus click' data combined with metadata) and *Narwhal Files* (count data, 'nearest neighbour distance' measurements, and metadata);

2) using computer vision: *Pengbot Count Files* (count data). Also included in the repository are the *Penguin Watch Manifest* (image metadata), summary graphs, animations, *Pengbot Out Files* and *Pengbot Density Maps*. The pipelines used to produce these files are presented (in the form of scripts written in the R programming language (currently v3.6.0 (R Core Team, 2019)); see Code Availability). These scripts have the potential to be applied to a broad range of data types. For example, a metric of movement, calculated using 'nearest neighbour distances', could be useful for analysis of images captured during aerial surveys in subsequent studies.

Methods:

In Jones et al. (2018) we outline the methodology associated with the *Penguin Watch* remote time-lapse camera network. Here we present the 'next steps' in the data processing pipeline, and offer two alternative workflows for extracting count data (i.e. the number of penguins per image): 1) via citizen science (Fig. 1 – left) and 2) via computer vision (Fig. 1 – right). The citizen science workflow is also used to calculate 'nearest neighbour distances', a metric which can be used to examine colony spatial structure and movement.

Access to two alternative processing methods is advantageous for a number of reasons. Firstly, certain images – for example photographs partially obscured by snow – may be more suitable for analysis using citizen science, while others – such as those containing high numbers of individuals – may be more appropriate for computer vision annotation. Furthermore, citizen science is a useful public engagement and educational tool (Homsy, 2014), while computer vision is useful for batch processing large quantities of data which may be considered 'less interesting' for human participants (e.g. winter images which contain few animals). Finally, citizen science classifications can be used to train future machine learning algorithms, and the counts produced by each can be cross-validated, ensuring the analyses remain reliable. In fact, Wright et al. (2017) demonstrate that a combination of methods (i.e. citizen science and machine learning) can outperform either method used alone.

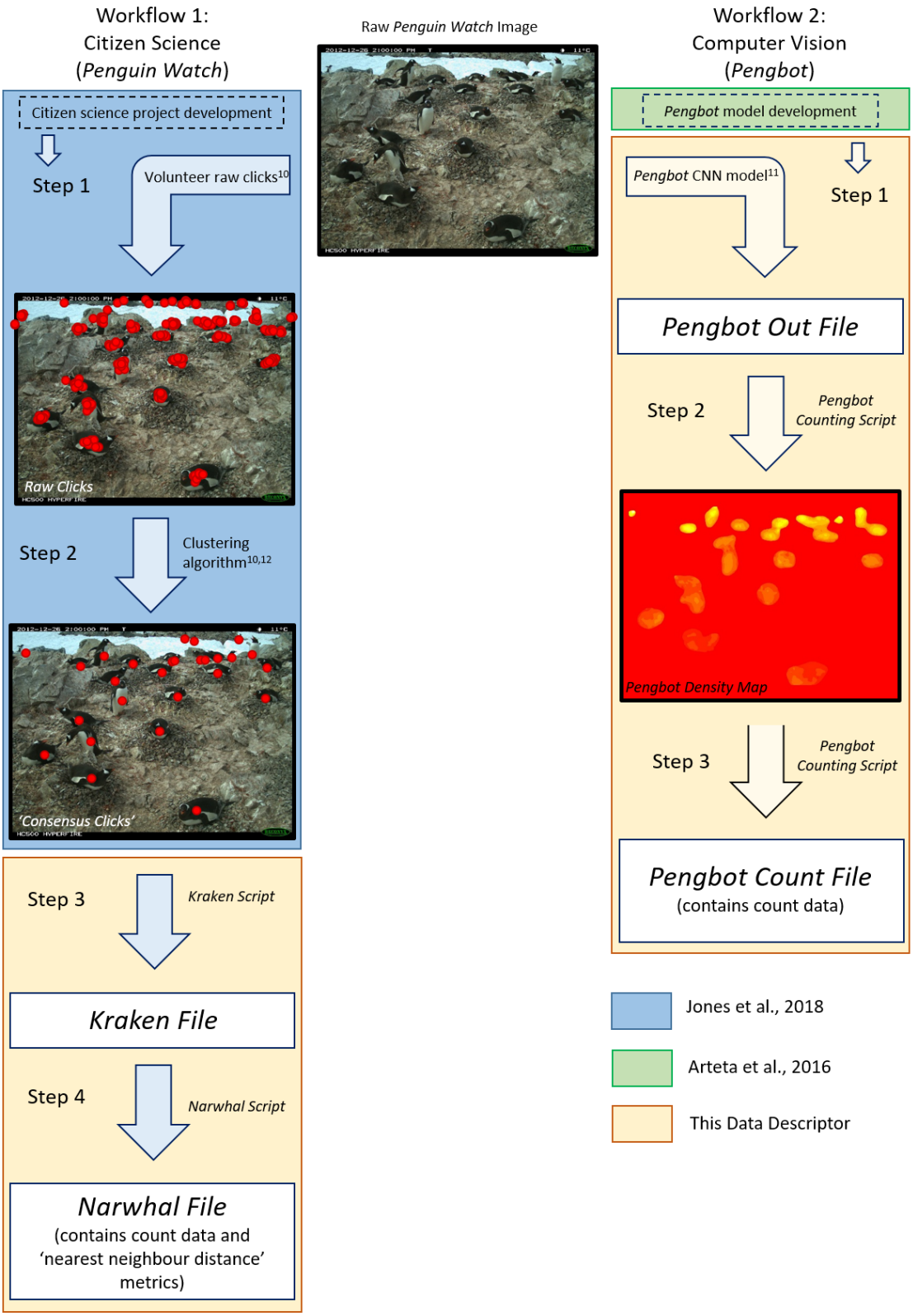


Figure 1: Flow diagram to show the two data processing pipelines using citizen science (the *Penguin Watch* project (Jones et al., 2018)) and computer vision (*Pengbot* (Arteta et al., 2016)).

The Citizen Science Pipeline – *Penguin Watch* (Fig. 1 – Left)

Steps 1 & 2: Raw images are annotated by volunteers; annotations are clustered (Jones et al., 2018).

The methodology behind the *Penguin Watch* citizen science project is discussed in Jones et al. (2018). One of the fundamental paradigms of all such projects is that multiple volunteers process each subject (e.g. image or audio clip), meaning average annotations can be extracted, and errors minimised, through filtering. In Jones et al. (2018) we therefore include discussion of a clustering algorithm (Hines, 2017) which uses agglomerative hierarchical clustering to take the xy coordinates of multiple volunteer clicks and produce a single 'consensus click', representing a single individual (penguin or 'other').

Step 3: Clustered annotations are filtered, and stored in *Kraken Files*

Here we present the next stage in the data processing pipeline (Fig. 1, left, 'Step 3'), taking these 'consensus clicks' and filtering them, to improve data reliability (Jones et al., 2018). The *Kraken Script* (see Code Availability) is used to achieve this, and the results (filtered 'consensus clicks' and associated metadata) are stored in *Kraken Files* (Data Citation 2). The *Kraken Script* extracts the 'consensus clicks' (Data Citation 1) and metadata (*Penguin Watch Manifest*; Data Citation 2) associated with the camera of interest, and merges them into a single file. In order to be extracted however, an adult 'consensus click' must have been generated from four or more raw volunteer clicks. The threshold is lower for chicks and eggs: two or more volunteer clicks must have formed the 'consensus click'. These rules can be modified within the *Kraken Script* to allow for different filtering thresholds, but we employ these levels as they were found to produce counts most similar to the gold

standard (Jones et al., 2018). The lower threshold level of *'num_markings > 1'* has also been used for eggs, because we believe they are often missed by volunteers, but a baseline level of filtering must be implemented to eradicate erroneous clicks.

Step 4: Penguin counts and 'nearest neighbour distances' are calculated.

Kraken Files are the input file type for the *Narwhal Script* (see Code Availability), which produces *Narwhal Files*. These files (Data Citation 2) include penguin counts, listed separately for 'adult', 'chick' and 'egg', and 'nearest neighbour distance' metrics. Looking at how mean average 'nearest neighbour distances' change over the breeding season is a useful indicator of movement, which in turn can signal the start of a new phenological stage, such as chick crèche. We also attempt to calculate the movement of each individual between images, again using 'nearest neighbour distances'. The metrics calculated are:

- Number of adults, chicks and eggs
- Average adult 'nearest neighbour distance' (the mean distance between each adult in image *i* and its nearest adult neighbour)
- Average chick 'nearest neighbour distance' (the mean distance between each chick in image *i* and its nearest chick neighbour)
- Average chick second 'nearest neighbour distance' (the mean distance between each chick in image *i* and its second nearest chick neighbour). Since *Pygoscelis* species lay two eggs, a chick's first nearest neighbour is likely to be its sibling in the nest. Therefore, looking at second nearest neighbour is a more useful indication of whole-colony movement.
- The standard deviations of each 'nearest neighbour distance' metric listed above.

- Movement of adults (see Fig. 2; the mean distance between each adult [*j*] in image [*i*] and its nearest neighbour in image [*i-1*] – i.e. the previous image. This means the coordinates of adult [*j*] are appended to all the adult coordinates in the previous image (including itself, if it was present), to make a new coordinate set ([*i-1*] with one additional data point – adult [*j*] in image [*i*]). The same adult will thus be represented by two data points in this new set of coordinates, if it was present in both images. 'Nearest neighbour distance' is then calculated for the adult of interest [*j*]. It is likely that adult [*j*]'s nearest neighbour will be itself (see Fig. 2, red arrow), in its previous position, meaning this value is equivalent to the distance adult [*j*] has moved between images. The mean is calculated to give the average distance moved by all adults in the image.).
- Movement of chicks (as above, for chicks).

This approach to quantifying movement is only robust when movement is limited; i.e. if an individual moves too far (or another individual comes very close), then its nearest neighbour in image *i-1* will not be itself, but a different individual (see Fig. 2, blue arrow). Therefore, this metric is most reliable in the period between egg lay and the end of the guard phase, when individuals are generally found at the nest. However, since individual penguins cannot be tracked using coordinate clicks – or indeed even in time-lapse photographs unless there are distinctive markings (e.g. in African penguins) – this remains a useful and novel indicator of movement. Furthermore, detection of 'change-points' is often more important than movement per se. By taking the mean average of all the movement in an image, these can be identified, and can provide a starting point for determining phenological dates.

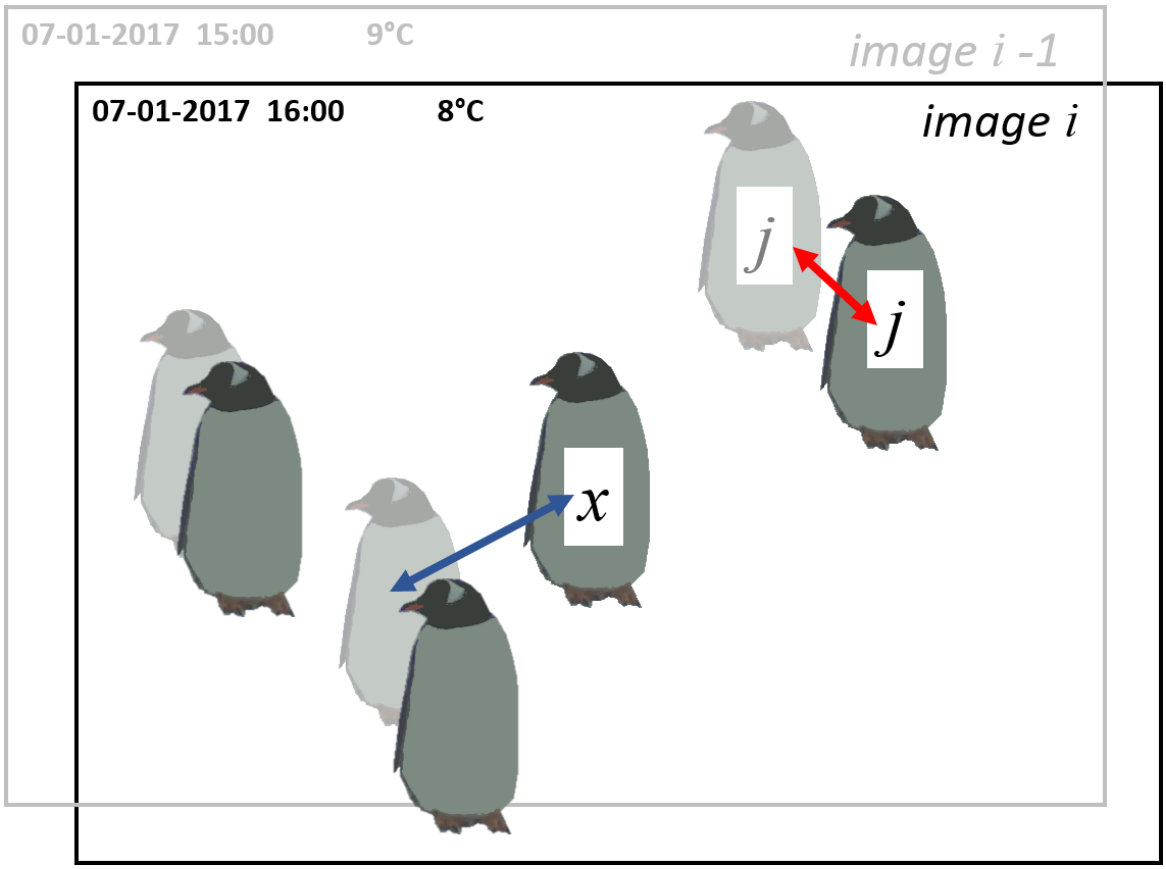


Figure 2: Diagram to illustrate how movement between images is calculated using 'nearest neighbour distances'. This figure shows two images, image *i* (black text and bold penguins) and image *i-1* (i.e. the previous image; grey text and faded penguins). If the coordinates of adult *j* (image *i*) are appended to the coordinates of the penguins in image *i-1*, adult *j* has a new nearest neighbour (in this case, adult *j* – itself – in its previous position). This 'nearest neighbour distance' (shown by the red arrow) therefore represents the distance adult *j* has moved between the two images. However, if an individual moves significantly between photographs (see adult x, which appears only in the later image) the 'nearest neighbour distance' (blue arrow) will not represent the individual's movement. Therefore, this metric is best used when movement is limited, for example during the incubation and guard phase.

An explanation of the different variables and commands found within the *Narwhal Script* is provided in Table 1.

Variable in *Narwhal File*	R code (in *Narwhal Script*) to calculate (for the *i*th value)	Explanation of variable
nadults	length(currentadult[,1])	Number of adults. The length (number of rows) of the 'adult' 'consensus click' data (clustered xy coordinates) associated with the current image.
nchicks	length(currentchick[,1])	Number of chicks. The length (number of rows) of the 'chick' 'consensus click' data (clustered xy coordinates) associated with the current image.
neggs	length(currenteggs[,1])	Number of eggs. The length (number of rows) of the 'egg' 'consensus click' data (clustered xy coordinates) associated with the current image.
adultndout	mean(nndist(currentadultxy)	Mean 'nearest neighbour distance' between each adult in image [*i*] and its nearest neighbour.

adultsdndout	sd(nndist(currentadultxy))	Standard deviation of the adult 'nearest neighbour distances'.
chickndout	mean(nndist(currentchickxy))	Mean 'nearest neighbour distance' between each chick in image [*i*] and its nearest neighbour.
chicksdndout	sd(nndist(currentchickxy))	Standard deviation of the chick 'nearest neighbour distances'.
chick2ndout	mean(nndist(currentchickxy, k=2))	The mean distance between each chick and its second nearest neighbour within each image. The second nearest neighbour is calculated to account for the fact that most nests have two chicks, so the first nearest neighbour is likely to be a sibling (and will therefore be less informative regarding movement in the colony).
chicksd2ndout	sd(nndist(currentchickxy, k=2))	Standard deviation of the chick 'second nearest neighbour distances'.
meanchangeadult	Create a data frame of xy coordinates for each adult [*j*] for	The mean distance between each adult [*j*] in image [*i*] and its nearest neighbour in image

	each image [*i*], and the previous image [*i-1*]: currentadultxy<- data.frame(currentadult$x_centre, currentadult$y_centre) prevadultxy<- data.frame(prevadult$x_centre, prevadult$y_centre) Add adult [*j*] in image [*i*] to the data frame for image [*i-1*]: prevadultx<-c(currentadultx, prevadult$x_centre) prevadulty<-c(currentadulty, prevadult$y_centre) prevadultxy<- data.frame(prevadultx, prevadulty) Calculate nnd for just the adult of interest [*j*]: temp<-nndist(prevadultxy)) prevadultndout[*j*]<-temp[1]	[*i-1*]. This means the coordinates of adult [*j*] must be added to all the adult coordinates in the previous image, to make a new data frame ([*i-1*] with one additional data point – adult [*j*] in image [*i*]). The same adult will thus be represented by two data points in this new data frame. 'Nearest neighbour distance' is then calculated as before, but only for the adult of interest [*j*]. It is likely that adult [*j*]'s nearest neighbour will be itself, in its previous position, meaning this value is equivalent to the distance adult [*j*] has moved between images. The mean is calculated to give the average distance moved by all adults in the image.

	Calculate the mean of these nnds: meanchangeadult[*i*]<- mean(prevadultndout)	
meanchangechick	Create a data frame of xy coordinates for each chick [*k*] for each image [*i*], and the previous image [*i-1*]: currentchickxy<- data.frame(currentchick$x_centre, currentchick$y_centre) prevchickxy<- data.frame(prevchick$x_centre, prevchick$y_centre) Add chick [*k*] in image [*i*] to the data frame for image [*i-1*]: prevchickx<-c(currentchickx, prevchick$x_centre) prevchicky<-c(currentchicky, prevchick$y_centre) prevchickxy<- data.frame(prevchickx, prevchicky)	The mean distance between each chick [*k*] in image [*i*] and its nearest neighbour in image [*i-1*]. This means the coordinates of chick [*k*] must be added to all the chick coordinates in the previous image, to create a new data frame ([*i-1*] with one additional data point – chick [*k*] in image [*i*]). The same chick will thus be represented by two data points in this new data frame. 'Nearest neighbour distance' is then calculated as before, but only for the chick of interest [*k*]. It is likely that the nearest neighbour to chick [*k*] will be itself in its previous position, meaning this value is equivalent to the distance chick [*k*] has moved between images. The mean is calculated to give the average distance

	Calculate nnd for just the chick of interest [*k*]: temp2<-nndist(prevchickxy)) prevchickndout[*k*]<-temp2[1] Calculate the mean of these nnds: meanchangechick[*i*]<- mean(prevchickndout)	moved by all chicks in the image.

Table 1: The variables contained in a *Narwhal File*, alongside the sections of R code (taken from the *Narwhal Script*) required to calculate them.

Step 5: *Narwhal Files* are used to create plots and animations

The *Narwhal Plotting Script* (see Code Availability) takes the *Narwhal Files* and creates plots of cumulative summary statistics for each time-lapse image: abundance (for adults and chicks), chick 'second nearest neighbour distance', and mean adult 'nearest neighbour distance' between the *i*th and (*i-1*)th image (all moving mean averages). We have grouped these graphs with their associated raw time-lapse image to complete each plot, and created a *Penguin Watch Narwhal Animation* for each camera (Data Citation 2) using the open source software GIMP (v 2.8; *GNU Image Manipulation Program (GIMP)*, 2014) and VirtualDub (v 1.10.4; Lee, 2013), to allow easy visualisation of population trends.

The Computer Vision Pipeline – *Pengbot* (Fig. 1 – Right)

Step 0: The *Pengbot* computer vision algorithm – development (Arteta et al., 2016).

Within *Penguin Watch* time-lapse images, penguins are often occluded and there can be a high degree of scale variation. This presents a challenge to computer vision developers, and makes simple counting-density estimation algorithms and binary object detectors insufficient for the task of producing reliable counts (Arteta et al., 2016). Furthermore, variation in camera angle and distance from a colony, obscuration owing to snow or ice, and shape similarity between the objects of interest (i.e. penguins) and background objects (e.g. rocks), all combine to make automatic detection a complex problem (Arteta et al., 2016).

The dot annotations (raw xy coordinate 'clicks'; Data Citation 1) of *Penguin Watch* volunteers offer a large training dataset for machine-learning algorithms (Jones et al., 2018), yet bring their own complications. For example, there is substantial variation amongst the annotations of different volunteers, meaning any model must be robust to noisy labels (Arteta et al., 2016). Arteta et al. (2016) present the first solution to the problem of counting from crowd-sourced dot annotations – the *Pengbot* model. The learning architecture of the *Pengbot* model is a deep multi-task convolutional neural network (CNN) (Arteta et al., 2016). It makes use of the variability in raw markings by using it to 1) estimate the 'difficulty' of an image, and 2) make inferences about local object scale (an individual penguin can measure fewer than 15 pixels or in excess of 700 pixels) (Arteta et al., 2016).

When the neural network is learning from the images, it tries to learn, among other things, which visual patterns belong to penguins. The optimal way to train it would therefore be to provide complete information regarding every pixel – i.e. which image pixels are part of an object (in this case a penguin) and which are not. This is extremely expensive – both in terms of experts' time and (if professional annotators are paid) financially – therefore the algorithm relies instead on the assumption that volunteers are very likely to click on penguins. Thus, the more clicks that are used, the more complete the information regarding what 'penguin pixels' look like. For this reason, the algorithm is trained on volunteer raw clicks, as opposed to 'consensus clicks'. There is a step of the learning process that uses a form of count consensus derived from the raw clicks. However, such consensus is defined over the continuously changing definition of what a penguin is (i.e. as the learning progresses), meaning raw clicks must be available all the time.

Step 1: The *Pengbot* model is used to generate *Pengbot Out Files*

The *Pengbot* CNN and associated code are publically available (see Code Availability). Running the algorithm on raw *Penguin Watch* images produces *Pengbot Out Files* (Data Citation 2). These are MATLAB files which contain matrices comprising 'penguin densities' for each pixel in the photograph. These matrices are visually represented in *Pengbot Density Maps*.

Step 2: *Pengbot Out Files* are used to produce *Pengbot Density Maps*

The *Pengbot Counting Script* (see Code Availability) is used to produce a *Pengbot Density Map* (see Fig. 3, right) from each *Pengbot Out File*, for each raw photograph (Data Citation 2). The lighter (yellow) parts of the image represent pixels that the algorithm interprets as

'penguin', as opposed to background (e.g. snow, ice and rocks). The brightness of these pixels corresponds to a scale of density, such that summing the pixels belonging to a single penguin should give approximately one. In general, pixels towards the top of the image (i.e. in the distance) are brighter – of a higher 'penguin density' – than those in the foreground. This is because penguins appear smaller towards the top/back of the image,

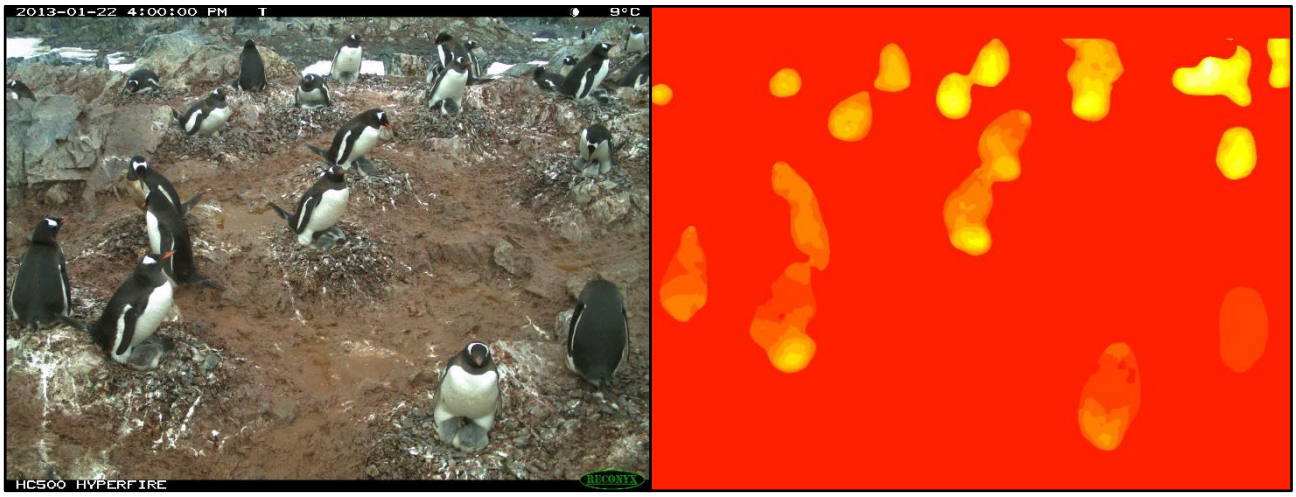


Figure 3: *Penguin Watch* time-lapse image GEORa2013b_000153 (left; captured by camera GEORa (64.67°S, 62.67°W) on the 22nd January 2013 at 4pm), alongside its *Pengbot Density Map* (right), generated by the *Pengbot* CNN.

so fewer pixels are required to comprise a whole individual. Conversely, duller pixels in the foreground have a lower 'penguin density', and more are required to count one individual. This method allows the oblique angle and vantage point of the camera to be taken into account when counts are generated from the density maps.

Step 3: *Pengbot Count Files* are generated using the density data

The *Pengbot Counting Script* (see Code Availability) is also used to calculate the number of penguins in each photograph, by summing pixel densities. Note that these counts are for the total number of all individuals, as the model cannot currently distinguish between adults and chicks. The data are stored in *Pengbot Count Files* (Data Citation 2).

Code Availability

Code used to produce files within Data Citation 1:

- Raw *Penguin Watch* time-lapse photographs are renamed and resized using an R (currently v3.6.0; (R Core Team, 2019)) script. The code is publically available via GitHub at https://github.com/zooniverse/Data-digging/blob/master/example_scripts/Penguin_Watch/Penguin_Watch_ImageProcessingScript.R. A static version (written using v3.4.1) is archived on Figshare (Hart et al., 2017).
- Raw *Penguin Watch* volunteer classifications (xy coordinate clicks) were clustered into 'consensus clicks' using agglomerative hierarchical clustering (Fig. 1, left, 'Step 2') (Hines, 2017; Jones et al., 2018). The aggregation algorithm is written in Python (v 2.7; (Python Software Foundation, 2010)) and can be found on GitHub at https://github.com/zooniverse/aggregation/blob/master/penguins/aggregate.py. A static version of this script is also archived on Figshare (Hines, 2017).

Code used to produce files within Data Citation 2:

- The *Kraken Script* (Fig. 1, left, 'Step 3'; output = *Kraken Files*) is written in R (v3.6.0; R Core Team, 2019); it can be accessed through GitHub at https://github.com/zooniverse/Data-digging/blob/master/example_scripts/Penguin_Watch/Kraken_Script.R, and a static version is archived on Zenodo at DOI:10.5281/zenodo.3238554.
- The *Narwhal Script* (Fig. 1, left, 'Step 4'; output = *Narwhal Files*) is written in R (v3.6.0); it can be accessed through GitHub at https://github.com/zooniverse/Data-digging/blob/master/example_scripts/Penguin_Watch/Narwhal_Script.R and a static version is archived on Zenodo at DOI: 10.5281/zenodo.3238573.
- The *Narwhal Plotting Script* (output = *Narwhal Plots* – graphs displaying *Narwhal* summary statistics) is written in R (v3.6.0); it can be accessed through GitHub at https://github.com/zooniverse/Data-digging/blob/master/example_scripts/Penguin_Watch/Narwhal_Plotting_Script.R and a static version is archived on Zenodo at DOI: 10.5281/zenodo.3238583.
- The *Pengbot* model (Arteta et al., 2016), associated code and instructions, and the dataset used to train the neural network can be found at the following address: https://www.robots.ox.ac.uk/~vgg/data/penguins/.
- The *Pengbot Counting Script* (Fig. 1, right, 'Step 2' and 'Step 3') is written in R (v3.6.0); it can be accessed through GitHub at https://github.com/zooniverse/Data-digging/blob/master/example_scripts/Penguin_Watch/Pengbot_Counting_Script.R and a static version is archived on Zenodo at DOI: 10.5281/zenodo.3238590.

Data Records

In Jones et al. (2018) we describe a dataset uploaded to the Dryad Digital Repository (Data Citation 1) comprising 63,070 unique raw time-lapse images captured by the *Penguin Watch* remote camera network, metadata for these photographs (e.g. date/time and temperature information), raw anonymised classifications from the *Penguin Watch* citizen science project, and 'consensus clicks' derived from these classifications. Please refer to Jones et al. (2018 and 2019) for full details and an explanation of these data types.

Here we present a second dataset made available through the Dryad Digital Repository (Data Citation 2). This contains two main file types. Firstly, files containing data directly derived from *Penguin Watch* citizen science classifications: *Kraken Files* and *Narwhal Files* (Appendix 5 -Table 2; produced using the *Kraken Script* and *Narwhal Script*, respectively), *Narwhal Plots* (Fig. 4 and Appendix 5 - Table 3; generated using the *Narwhal Plotting Script*), and *Penguin Watch Narwhal Animations* (Appendix 5 - Table 3, produced using GIMP (v2.8; *GNU Image Manipulation Program (GIMP)*, 2014) and VirtualDub (v1.10.4; Lee, 2013). Secondly, files produced using computer vision – i.e. the *Pengbot* CNN (albeit initially trained on citizen science dot annotations): *Pengbot Out Files*, *Pengbot Density Maps*, and *Pengbot Count Files* (Appendix 5 - Table 4). All of these files directly relate to those provided in Data Citation 1; i.e. an example within every file type is provided for each of the 63,070 raw time-lapse images captured by 14 cameras, meaning the processing pipeline can be traced for every photograph.

Also included in the repository (Data Citation 2) are a *Penguin Watch Manifest* file, containing metadata for all of the images, and a *Method Comparison File*, which contains

analysis of gold standard, citizen science and *Pengbot* counts for 1183 *Penguin Watch* images (see Technical Validation).

Explanation of Terms

Citizen Science Workflow (Fig. 1 – Left)

'Penguin Watch Manifest'

name: Unique image reference for identification, in the format: SITExYEARx_imagenumber; e.g. DAMOa2014a_000001.

datetime: Time and date information for the image, in the format YYYY:MM:DD HH:MM:SS.

zooniverse_id: Unique identification code assigned to each image within Zooniverse (the online platform that hosts *Penguin Watch* – see www.zooniverse.org).

path: Folder pathway, which includes the image name (e.g. DAMOa/DAMOa2014a_000025).

classification_count: Number of volunteers who classified the image before it was retired.

state: State of completion – either complete (the image has been shown to the required number of volunteers) or incomplete (the image requires classification by further volunteers).

temperature_f: Temperature (in degrees Fahrenheit, as recorded by the camera) at the time the photograph was taken.

lunar_phase: Moon phase when the image was captured (one of eight options: “full” (full), “new” (new), “newcres” (new crescent), “firstq” (first quarter), “waxinggib” (waxing gibbous), “waninggib” (waning gibbous), “lastq” (last quarter) or “oldcres” (old crescent)).

URL: Link to an online thumbnail image of the time-lapse photograph (lower resolution than the raw time-lapse imagery included in the repository, but useful for reference).

‘Kraken Files’ (Appendix 5 - Table 2)

Kraken Files comprise filtered ‘consensus clicks’ and metadata. The filtering threshold levels are ‘*num_markings > 3’* for adults and ‘*num_markings > 1’* for chicks and eggs (see Methods). Each row contains the following information (note that a row of NAs (for probability values etc.) shows no penguins were marked for that image):

name: Unique image reference for identification, in the format: SITExYEARx_imagenumber; e.g. DAMOa2014a_000001.

probability_of_adult: Estimated probability that the corresponding individual is an adult – based on the number of volunteers classifying it as such, as a proportion of the total number of clicks on that individual.

probability_of_chick: Estimated probability that the corresponding individual is a chick – based on the number of volunteers classifying it as such, as a proportion of the total number of clicks on that individual.

probability_of_egg: Estimated probability that the corresponding marking indicates an egg – based on the number of volunteers classifying it as such, as a proportion of the total number of clicks on that area.

num_markings: The number of volunteer clicks that were aggregated to produce the 'consensus click' coordinate values (i.e. the number of individual clicks on a specific area of the image). Owing to the filtering process, *num_markings* is '*>3*' for adults and '*>1*' for chicks and eggs in the files provided alongside this Data Descriptor. These threshold levels can be changed in the *Kraken Script*, to create *Kraken Files* with different degrees of filtering.

x_centre: x coordinate value (in pixels) for the 'consensus click' (i.e. the coordinate calculated by the clustering algorithm) (Hines, 2017; Jones et al., 2018). The origin (point 0, 0) is located in the top left-hand corner of the image, meaning it may be necessary to reverse the y-axis of a plot in order to overlay the 'consensus clicks' correctly. One coordinate value denotes one individual penguin/'other'.

y_centre: y coordinate value (in pixels) for the 'consensus click' (i.e. the coordinate calculated by the clustering algorithm) (Hines, 2017; Jones et al., 2018). The origin (point 0, 0) is located in the top left-hand corner of the image, meaning it may be necessary to reverse the y-axis of a plot in order to overlay the 'consensus clicks' correctly. One coordinate value denotes one individual penguin/'other'.

For *datetime*, *temperature_f*, *lunar_phase* and *URL* please see *'Penguin Watch Manifest'*.

'Narwhal Files' (Appendix 5 - Table 2)

Narwhal Files contain penguin count data, 'nearest neighbour distance' metrics, and metadata. Each row contains the following information:

imageid: Unique image reference for identification, in the format: SITExYEARx_imagenumber; e.g. DAMOa2014a_000001.

nadults: The total number of adults counted in the corresponding image, based on filtered (here, *num_markings > 3*) 'consensus click' data (see *Kraken Files*).

nchicks: The total number of chicks counted in the corresponding image, based on filtered (here, *num_markings > 1*) 'consensus click' data (see *Kraken Files*).

neggs: The total number of eggs counted in the corresponding image, based on filtered (here, *num_markings > 1*) 'consensus click' data (see *Kraken Files*).

adultndout: The mean adult 'nearest neighbour distance'. Calculated by taking the distance between each adult and its nearest (adult) neighbour, and finding the mean average.

adultsdndout: Standard deviation of the adult 'nearest neighbour distances'.

chickndout: The mean chick 'nearest neighbour distance'. Calculated by taking the distance between each chick and its nearest (chick) neighbour, and finding the mean average.

chicksdndout: Standard deviation of the chick 'nearest neighbour distances'.

chick2ndout: The mean distance between each chick and its second nearest (chick) neighbour within each image. The second nearest neighbour is calculated because the *Pygoscelis* penguins (the primary focus of the *Penguin Watch* camera network) usually lay two eggs. Since a chick's nearest neighbour is therefore likely to be its sibling in the nest, the distance to the second nearest neighbour is calculated, to provide more information about the spatial distribution of chicks within the colony.

chicksd2ndout: Standard deviation of the chick 'second nearest neighbour distances'.

meanchangeadult: Movement of each adult [*j*] between image [*i-1*] and image [*i*] (see Fig. 2). This is calculated by appending the xy coordinate of adult [*j*] in image [*i*] to a data frame of the adult xy coordinates in image [*i-1*], and calculating the 'nearest neighbour distance' for adult [j]. The nearest adult to [*j*] is likely itself in the previous image, therefore this distance represents movement between the two images. It is possible that the nearest neighbour is a different individual, so a mean average is calculated to provide an indication of movement within the field of view. This metric is most appropriate between incubation and the end of the guard phase, when adults are often at the nest (see Methods).

meanchangechick: Movement of each chick [*k*] between image [*i-1*] and image [*i*] (see Fig. 2). This is calculated by appending the xy coordinate of chick [*k*] in image [*i*] to a data frame of the chick xy coordinates in image [*i-1*], and calculating the 'nearest neighbour distance' for chick [*k*]. The nearest chick to [*k*] is likely itself in the previous image, therefore this distance represents movement between the two images. It is possible that the nearest neighbour is a different individual, so a mean average is calculated to provide an indication of movement within the field of view. This metric is most appropriate between incubation and the end of the guard phase, when chicks are at the nest (see Methods).

tempf: Temperature (in degrees Fahrenheit, as recorded by the camera) at the time the photograph was taken.

tempc: Temperature (in degrees Celsius) at the time the photograph was taken.

For *datetime*, *lunar_phase* and *URL*, please see *'Penguin Watch Manifest'*

Narwhal Plots and Penguin Watch Narwhal Animations (Appendix 5 - Table 3)

Narwhal Plots (see Fig. 4), generated using the *Narwhal Plotting Script*, provide a visualisation of summary statistics: graph 1: abundance of adults and chicks; graph 2: chick 'second nearest neighbour distances'; and graph 3: mean adult 'nearest neighbour distance' between the *i*th and (*i-1*)th image.

A plot is produced for each time-lapse image, showing the moving mean average trends up to, and including, that image. Therefore, to see the complete trend, a plot should be created for the final image in the data series. The complete sets of graphs have been used to create *Penguin Watch Narwhal Animations*, which show the trends developing through time, alongside the time-lapse images. The folders of *Narwhal Plots* and *Penguin Watch Narwhal Animations* correspond to those described in Table 4 of Jones et al. (2018; Appendix 4).

Computer vision (*Pengbot*) Workflow (Fig. 1 – Right)

Pengbot Out Files (Appendix 5 - Table 4)

Pengbot Out Files are MATLAB files containing a matrix of 'penguin densities', calculated using the *Pengbot* model. A density value is provided for each pixel in the corresponding raw time-lapse photograph. A file is provided (Data Citation 2) for each of the 63,070 images presented in Data Citation 1; they are separated into folders according to camera (e.g. DAMOa_pengbot_out).

Pengbot Density Maps (Appendix 5 - Table 4; Fig. 3, right)

A *Pengbot Density Map* (Data Citation 2), generated using the *Pengbot Counting Script* (see Code Availability), is provided for each of the 63,070 raw time-lapse photographs provided

in Data Citation 1. The maps are a visual representation of the pixel densities presented in the *Pengbot Out Files*. Maps are stored in the database according to camera (e.g. DAMOa_density_map).

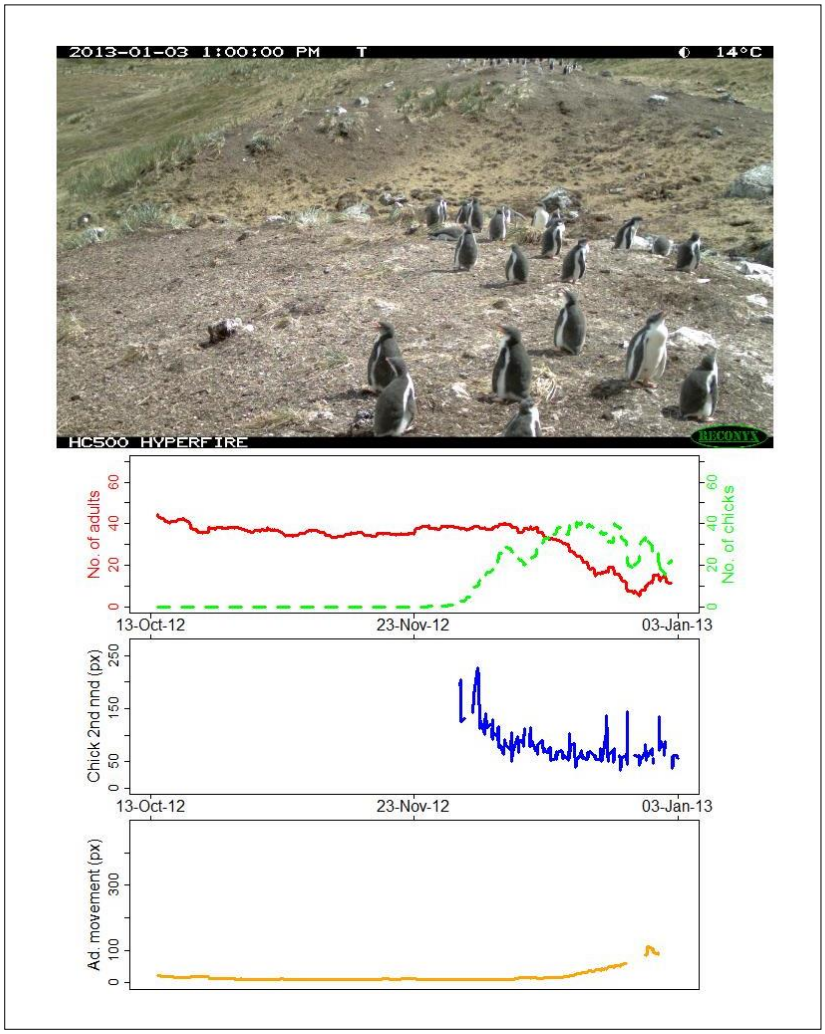


Figure 4: The *Narwhal Plot* for MAIVb2012a_000661 (the last image in the MAIVb2012a folder (Data Citation 1), therefore showing the complete trends for the data series). The plot comprises, from top down: the original time-lapse image, graph 1: number of adults and chicks (moving mean average, n=20), graph 2: mean chick 'second nearest neighbour distances' (moving mean average, n=2), and graph 3: mean adult 'nearest neighbour distance' between the *i*th and (*i-1*)th image (moving mean average, n=20).

Pengbot Count Files (Appendix 5 - Table 4)

These are generated using the *Pengbot Counting Script* (see Code Availability). They provide a penguin count (adults and chicks combined) for each image by summing the 'penguin density' of each pixel in the *Pengbot Out Files*.

imageid: Unique image reference for identification, in the format: SITExYEARx_imagenumber; e.g. DAMOa2014a_000001.

raw_count: The number of penguins in the image, as calculated by the *Pengbot* model (Arteta et al., 2016). Since the counts are generated by summing pixel densities, they are unlikely to be an integer value.

count: The number of penguins in the image as calculated using *Pengbot,* rounded to the nearest integer.

Method Comparison File

This file contains penguin counts for 1183 images, from four different cameras – DAMOa (Damoy Point; n = 300), HALFc (Half Moon Island; n = 283), LOCKb (Port Lockroy; n = 300) and PETEc (Petermann Island; n = 300) – as calculated by an expert ('Gold Standard'), through *Penguin Watch* ('Citizen Science'), and by *Pengbot* ('Computer Vision').

imageid: Unique image reference for identification, in the format: SITExYEARx_imagenumber; e.g. DAMOa2014a_000001.

datetime: Time and date information for the image, in the format YYYY:MM:DD HH:MM:SS.

GS_adults: The number of penguin adults in the corresponding image, as identified and calculated by an expert.

GS_combined: The number of penguin adults and chicks in the corresponding image, as identified and calculated by an expert.

CS_adults: The number of penguin adults in the corresponding image, generated using clustered citizen science data from the *Penguin Watch* project (threshold = *num_markings > 3*).

CS_combined: The number of penguin adults and chicks, as generated using clustered citizen science data from the *Penguin Watch* project (threshold = *num_markings > 3* for adults and *num_markings > 1* for chicks).

CV_rounded: The rounded number of penguin individuals (adults and chicks combined) in the corresponding image, calculated using the *Pengbot* model (Arteta et al., 2016).

The file also includes columns for six comparisons:

GS_combined vs. CV_rounded; GS_adults vs. CV_rounded; CS_combined vs. CV_rounded; CS_adults vs. CV_rounded; GS_combined vs. CS_combined; GS_adults vs. CS_adults.

Two columns are provided for each comparison. The first column shows the difference (in raw number of penguins) between the counts generated via the two methods. For example, if the gold standard combined count was 15 and the computer vision count was 12, then the 'GS_combined vs. CV_rounded' column would contain the value 3. The second column shows the same values, with any negative signs removed (whether the counts are under- or over-estimates is irrelevant when calculating average differences).

The mean average difference in counts, the standard deviation of this difference, the proportion of counts that are equal or differ by only one penguin, and the number of over- and under-estimates are also provided in the spreadsheet (and summarised in Table 5). Overestimates and underestimates relate to the second variable, i.e. for 'GS_combined vs. CV_rounded', an 'underestimate' would mean an underestimate by computer vision, as compared to the gold standard.

Technical Validation

In Jones et al. (2018) we provide a technical validation of *Penguin Watch* citizen science data, comparing clustered counts (at four different threshold levels) to counts derived from expert annotations (the 'gold standard'). Here we extend this analysis by comparing counts generated by the *Pengbot* computer vision algorithm to both gold standard and citizen science count data (see Table 5).

As in Jones et al. (2018), images from cameras at four sites (Damoy Point (DAMOa), Half Moon Island (HALFc), Port Lockroy (LOCKb) and Petermann Island (PETEc)) were employed in the analysis, to ensure that images showing different camera angles and colony proximity, and capturing all three *Pygoscelis* species, were included. Gold standard classifications (annotations by author FMJ) and citizen science counts were taken from Jones et al. (2018), where a sample of 300 images was randomly selected for each site from the photographs that were 1) marked as containing animals and 2) were marked as complete in *Penguin Watch*. (Note that an exception to this is HALFc, where the whole sample of 283 images was used, to give a total of 1183.)

	GS vs. CV (combined)	GS vs. CV (adults)	CS vs. CV (combined)	CS vs. CV (adults)	GS vs. CS (combined)	GS vs. CS (adults)
DAMOa	n = 300					
Mean difference	1.95	2.01	2.21	2.25	1.36	1.18
σ	1.84	2.15	2.16	2.38	1.66	1.56
Proportion 0 or 1	0.49	0.49	0.47	0.47	0.70	0.75
Overestimate	121	143	116	145	102	101
Underestimate	115	90	136	102	113	100
HALFc	n = 283					
Mean difference	1.50	1.21	1.46	1.45	1.20	0.88
σ	1.36	1.08	1.70	1.56	1.53	1.34
Proportion 0 or 1	0.61	0.69	0.65	0.66	0.72	0.85
Overestimate	60	107	70	109	58	64
Underestimate	161	101	134	99	116	91
LOCKb	n = 300					
Mean difference	1.83	3.71	2.13	3.50	1.23	1.17
σ	1.70	4.27	2.50	4.10	2.03	1.98
Proportion 0 or 1	0.55	0.45	0.52	0.46	0.72	0.77
Overestimate	68	157	80	159	61	82
Underestimate	168	96	151	81	92	80
PETEc	n = 300					
Mean difference	12.13	7.17	11.31	7.09	3.70	2.36
σ	5.07	3.39	7.19	4.17	4.12	2.69
Proportion 0 or 1	0	0.06	0.02	0.08	0.36	0.46
Overestimate	0	81	1	84	85	79
Underestimate	300	212	297	207	170	162

Table 5: Comparison between gold standard (GS; expert classifications), citizen science (CS*; Penguin Watch*) and computer vision (CV; *Pengbot*) counts (n = 1183). Mean average differences (in raw count) are provided, alongside the standard deviation of these differences, the proportion of counts that were equal or differed by only one penguin, and the number of over- and underestimates. For GS and CS, combined counts (i.e. adults and

chicks) and adult-only counts are included. The filtering threshold levels used for the CS counts are *num_markings > 3* for adults and *num_markings > 1* for chicks. CV cannot distinguish between adults and chicks, so a single value is provided. Overestimates and underestimates relate to the second variable, i.e. for DAMOa 'GS vs. CV (combined)', computer vision overestimated the penguin count in 121 cases, and underestimated it in 115 cases, as compared to the gold standard. GS vs CS counts are included for completeness – please see Jones et al. (2018) for a full discussion.

Here, citizen science counts generated using a clustering threshold level of >3 for adults and >1 for chicks (i.e. four or more volunteer clicks, or two or more volunteer clicks, were required to form a 'consensus click', respectively) are used, since these filtering levels were previously found to produce counts most similar to the gold standard, in most cases (see Jones et al., 2018). These are also the threshold levels used to create the *Kraken* and *Narwhal Files* discussed in this Data Descriptor. Adult counts and combined counts (i.e. adults and chicks) are provided for the gold standard and citizen science methods, to examine the impact of this on the agreement with computer vision, which cannot yet distinguish between these groups.

The mean difference between the gold standard (combined) and computer vision derived counts was less than two for DAMOa, HALFc and LOCKb (1.95 (σ=1.84), 1.50 (σ=1.36) and 1.83 (σ=1.70), respectively) – similar to the differences between gold standard (combined) and citizen science (combined) counts (see Fig. 5). The degree of error was greater for PETEc, where the count differed by 12.13 penguins on average. The greatest discrepancies between the gold standard and citizen science counts were also associated with this site (likely owing to the higher mean number of adults and chicks at this site – see Jones et al.,

2018). With the exception of DAMOa, the number of underestimates outweighed the number of overestimates for each site, showing that the discrepancies were mainly owing to 'missed' individuals. As discussed in Jones et al. (2018), chicks are often missed by citizen science volunteers – perhaps owing to obscuration by an adult, misidentification, or – since volunteers can move onto another image after marking 30 individuals – simply going unannotated. While the latter is irrelevant for computer vision, it is possible that chicks go undetected by the algorithm owing to their small size, and position underneath (or directly in front of) a parent.

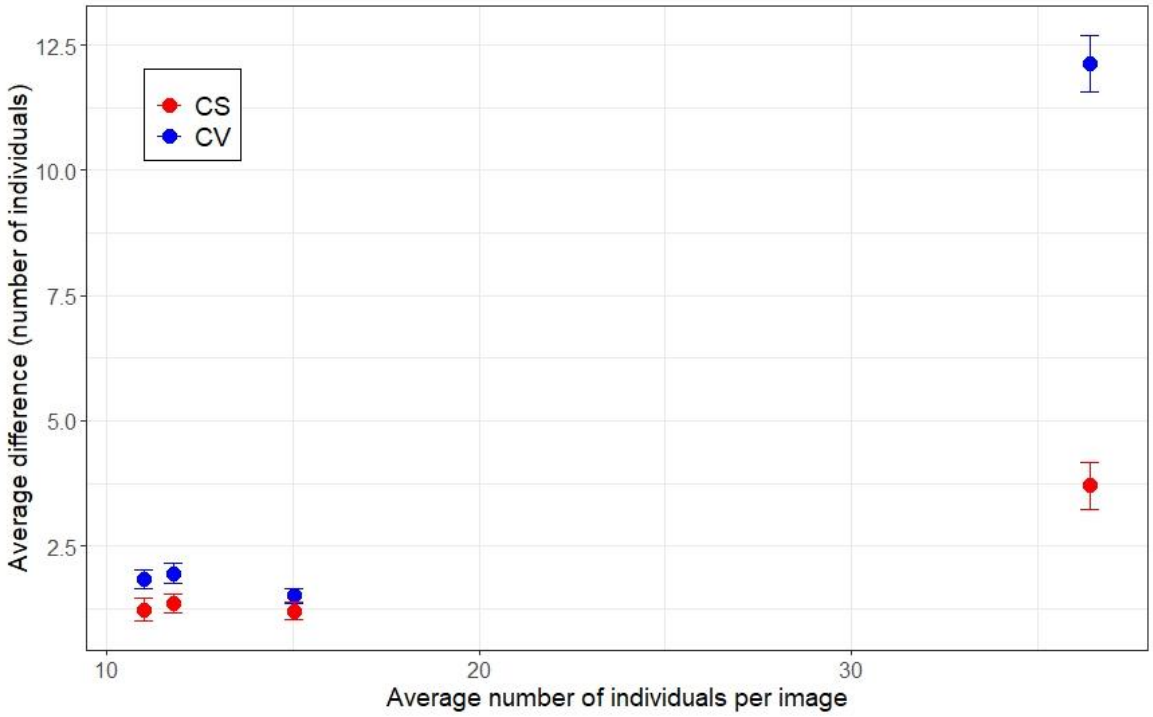


Figure 5: The mean average difference between GS (combined) and CV counts (blue) and GS (combined) and CS (combined) counts (red), shown against the mean average number of individuals per image. A = DAMOa, B = HALFc, C = LOCKb, D = PETEc. Error bars show the 95% confidence interval of the means.

Since the proportion of underestimates generally increases with an increasing mean number of chicks per image, it is possible that undetected chicks are a notable contributor to the error rate. However, while the proportion of overestimates is increased when chick counts are excluded (as expected, since the computer vision counts remain constant), the overall agreement is not improved in all cases (i.e. for DAMOa and LOCKb). Therefore, there are other factors contributing to the high proportion of underestimated counts. As stated, the largest discrepancies between the computer vision derived counts and gold standard counts are associated with PETEc, where 100% of counts were underestimates when compared to the gold standard (combined). This is also the site which has the highest number of individuals per image on average (36.43 adults compared to 11.79, 15.03 and 10.99 at DAMOa, HALFc and LOCKb, respectively). A greater number of individuals leads to a higher probability of crowding, with some individuals occluded by others (see Fig. 6). While the human eye can detect a partially-obscured penguin (in an extreme case, being able to mark an individual when only its beak is visible), the computer vision algorithm – which counts penguins by summing pixel densities – cannot achieve this. Furthermore, a problem which has the potential to affect all images is an "edge-effect". Again, if only half a penguin is visible on the perimeter of an image, a human annotator will record this as 'one' penguin. However, the computer vision algorithm can only record it as half an individual, lowering the overall estimation.

These three computer vision issues – missed chicks, occlusion of individuals, and the edge-effect – likely act in combination to produce underestimates of counts, particularly when images contain high numbers of individuals, such as at PETEc. However, as shown by Table 5 and Fig. 5, computer vision offers a valid alternative to expert annotation or citizen



Figure 6: Raw time-lapse image (top) and density map (bottom) for PETEc2013b_000023, showing how crowding and occlusion can lead to underestimation by *Pengbot*. While the computer vision algorithm is correctly identifying 'penguin pixels', the degree of occlusion – particularly in the top left hand corner of the image – makes counting a challenge.

science when examining images with fewer individuals. Moreover, when investigating colony phenology, overall population trends are of greater interest than raw population counts. As shown in Fig. 7, the main population trends observed at PETEc appear to be preserved irrespective of the annotation method used. This is supported by the Pearson's Correlation Coefficent values (see Table 6), which are greater than $r(298) = 0.90$ ($p < 0.001$) for both citizen science and computer vision, as compared to the gold standard population

trend. Interestingly, although there is a greater agreement between the gold standard and citizen science data than the gold standard and computer vision data in terms of raw count (see Table 5 and Fig. 5), the strongest Pearson's Correlation Coefficient is found between the gold standard and computer vision trends at the PETEc site: $r(298) = 0.94$, $p < 0.001$. This further justifies the use of computer vision analysis for comparision of population trends between colonies and seasons.

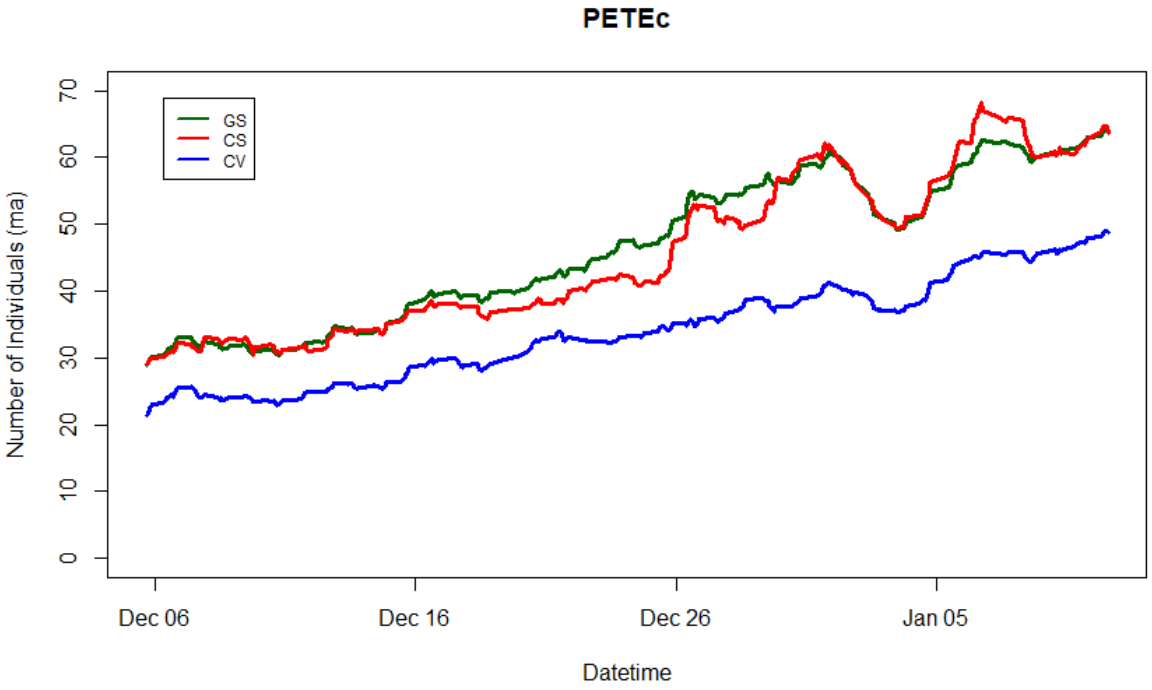


Figure 7: The penguin population (adults and chicks) trend at Petermann Island (PETEc) between the 3rd December 2012 and the 11th January 2013, as determined using expert-annotations (gold standard (GS); green), *Penguin Watch* (citizen science (CS); red), and the *Pengbot* algorithm (computer vision (CV); blue). Trends are shown as a moving mean average (n=15). Data are from a random sample of 300 images taken during the stated time period.

Annotation Method	GS	CS	CV
df = 298	Pearson's Correlation Coefficient (r); p-value		
GS	1.00	0.91 (p<0.001)	0.94 (p<0.001)
CS		1.00	0.88 (p<0.001)
CV			1.00

Table 6: Pearson's Correlation Coefficient (r) values, with corresponding p-values, for the relationship between the number of individuals (adults and chicks) at Petermann Island (PETEc) between the 3rd December 2012 and the 11th January 2013, as determined using expert-annotations (gold standard (GS)), *Penguin Watch* (citizen science (CS)), and the *Pengbot* algorithm (computer vision (CV)). Data are from a random sample of 300 images taken during the stated time period. Degrees of freedom (df) = 298. All three correlations are significant at the $p < 0.001$ level, with the strength of the relationship ranging from $r = 0.88$ (CS vs CV) to 0.94 (GS vs CV).

Remote camera technologies have the potential to monitor animal and plant populations in locations which may otherwise be impossible to effectively survey. Here we show that citizen science and computer vision offer alternative, albeit complementary, approaches to image analysis, meaning that the vast quantity of data produced via camera networks should not be a barrier to their use. We demonstrate how meaningful biological metrics can be extracted from imagery, and hope that the *Penguin Watch* data presented here may form a case study for those wishing to carry out similar analyses, and may be useful to both ecologists and computer vision developers.

Acknowledgements

The authors wish to thank the Zooniverse team[1], *Penguin Watch* partners and collaborators[2], and citizen science volunteers. We would also like to thank J. Arthur, H.R. Torsey and Z. Macháčková for moderating the *Penguin Watch* 'Talk' forum. Zooniverse and *Penguin Watch* were supported by grants from the Alfred P. Sloan Foundation and a Google Global Impact Award. This work has been supported by NERC and individual donors to the *Penguin Watch* project. Finally, we express our gratitude to Quark Expeditions, Cheesemans' Ecology Safaris and Oceanwide Expeditions for their collaboration during fieldwork. Fiona Jones and Tom Hart had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

1. https://www.zooniverse.org/about/team
2. https://www.zooniverse.org/projects/penguintom79/penguin-watch/about/team

Author Contributions

FMJ wrote the paper with input from TH; FMJ and TH developed processing code; TH, AZ and CJL developed the *Penguin Watch* citizen science project; CA, VL and AZ developed the *Pengbot* computer vision algorithm; FMJ provided gold standard classifications and performed the Technical Validation analysis. All authors reviewed and approved the manuscript.

Competing Interests

The authors declare no competing financial interests.

References

Arteta, C., Lempitsky, V., & Zisserman, A. (2016). Counting in the Wild. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer Vision - ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9911* (pp. 483–498). https://doi.org/https://doi.org/10.1007/978-3-319-46478-7_30

Croxall, J. P., Butchart, S. H. M., Lascelles, B., Stattersfield, A. J., Sullivan, B., Symes, A., & Taylor, P. (2012). Seabird conservation status, threats and priority actions: a global assessment. *Bird Conservation International*, *22*(1), 1–34.

Ducklow, H. W., Fraser, W. R., Meredith, M. P., Stammerjohn, S. E., Doney, S. C., Martinson, D. G., … Amsler, C. D. (2013). West Antarctic Peninsula: An ice-dependent coastal marine ecosystem in transition. *Oceanography*, *26*(3), 190–203. https://doi.org/http://dx.doi.org/10.5670/oceanog.2013.62

GNU Image Manipulation Program (GIMP). (2014). Retrieved from https://www.gimp.org/

Hart, T., Jones, F. M., & Freeman, R. (2017). Penguin Watch Image Processing Script (R). figshare. Retrieved from https://doi.org/10.6084/m9.figshare.5432974.v2

Hines, G. (2017). Penguin Watch Aggregation Script (Python). figshare. Retrieved from https://doi.org/10.6084/m9.figshare.5472544.v1

Homsy, V. (2014). *Engagement in the Zooniverse*. University of Oxford.

Jones, F. M., Allen, C., Arteta, C., Arthur, J., Black, C., Emmerson, L. M., … Hart, T. (2018). Time-lapse imagery and volunteer classifications from the Zooniverse Penguin Watch project. *Scientific Data*, *5*, 180124. https://doi.org/10.1038/sdata.2018.124

Jones, F. M., Allen, C., Arteta, C., Arthur, J., Black, C., Emmerson, L. M., … Hart, T. (2019). Author Correction: Time-lapse imagery and volunteer classifications from the Zooniverse Penguin Watch Project. *Scientific Data*, *6*(1), 1–2.

Kawaguchi, S., Nicol, S., & Press, A. J. (2009). Direct effects of climate change on the Antarctic krill fishery. *Fisheries Management and Ecology*, *16*(5), 424–427. https://doi.org/10.1111/j.1365-2400.2009.00686.x

Lee, A. (2013). *VirtualDub*. Retrieved from http://www.virtualdub.org/

Lynch, H. J., Naveen, R., Trathan, P. N., & Fagan, W. F. (2012). Spatially integrated assessment reveals widespread changes in penguin populations on the Antarctic Peninsula. *Ecology*, *93*(6), 1367–1377. https://doi.org/10.1890/11-1588.1

Newbery, K. B., & Southwell, C. (2009). An automated camera system for remote monitoring in polar environments. *Cold Regions Science and Technology*, *55*(1), 47–51. https://doi.org/10.1016/j.coldregions.2008.06.001

Nicol, S., Foster, J., & Kawaguchi, S. (2012). The fishery for Antarctic krill - recent developments. *Fish and Fisheries*, *13*(1), 30–40. https://doi.org/10.1111/j.1467-2979.2011.00406.x

Paleczny, M., Hammill, E., Karpouzi, V., & Pauly, D. (2015). Population Trend of the World's Monitored Seabirds, 1950-2010. *PLoS ONE*, *10*(6), 1–11. https://doi.org/10.1371/journal.pone.0129342

Python Software Foundation. (2019). *Python Language Reference*. Retrieved from http://www.python.org

R Core Team. (2019). *R: A language and environment for statistical computing.* Retrieved from https://www.r-project.org/

Stammerjohn, S., Massom, R., Rind, D., & Martinson, D. (2012). Regions of rapid sea ice change: An inter-hemispheric seasonal comparison. *Geophysical Research Letters*, *39*(6), L06501. https://doi.org/10.1029/2012GL050874

Trathan, P. N., Garcia-Borboroglu, P., Boersma, D., Bost, C.-A., Crawford, R. J. M., Crossin, G. T., … Wienecke, B. (2015). Pollution, habitat loss, fishing, and climate change as critical threats to penguins. *Conservation Biology*, *29*(1), 31–41. https://doi.org/10.1111/cobi.12349

Wright, D. E., Lintott, C. J., Smartt, S. J., Smith, K. W., Fortson, L., Trouille, L., … Young, D. R. (2017). A transient search using combined human and machine classifications. *Monthly Notices of the Royal Astronomical Society*, *472*(2), 1315–1323. https://doi.org/10.1093/MNRAS/STX1812

Data Citations

Data Citation 1: Jones, F.M., Allen, C., Arteta, C., Arthur, J., Black, C., Emmerson, L.M., Freeman, R., Hines, G., Lintott, C.J., Macháčková, Z., Miller, G., Simpson, R., Southwell, C. Torsey, H.R., Zisserman, A. & Hart, T. (2018). Data from: Time-lapse imagery and volunteer classifications from the Zooniverse Penguin Watch project. Dryad Digital Repository. DOI:10.5061/dryad.vv36g

Data Citation 2: Jones, F.M., Arteta, C., Lempitsky, V., Zisserman, A., Lintott, C.J. & Hart, T. Data from: Processing citizen science- and machine- annotated time-lapse imagery for biologically meaningful metrics. Dryad Digital Repository. DOI:10.5061/dryad.94sp17b

Chapter 4: Using citizen science to examine phenology and behaviour: new insights from the *Penguin Watch* project

Authors

Fiona M. Jones[1], Ignacio Juarez Martinez[1], Coleman Krawczyk[2], Victoria Hope[3], Chris J. Lintott[4] & Tom Hart[1]

Affiliations

1. Department of Zoology, University of Oxford, 11a Mansfield Road, Oxford, OX1 3SZ, UK

2. Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Burnaby Road, Portsmouth, PO1 3FX, UK

3. School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK

4. Zooniverse, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH, UK

Corresponding author: F. M. Jones (fiona.jones@zoo.ox.ac.uk)

Abstract

Phenology, the study of recurring stages – and crucially, the timing of these stages – within the life-history of an organism, can shed light on the nature and rate of key environmental changes. However, phenological dates are poorly documented for many species, including the Chinstrap penguin (*Pygoscelis antarcticus*), particularly over large spatial scales. Remote time-lapse cameras offer novel and invaluable insight into colony phenology and behaviour, but the volume of imagery they produce can be overwhelming to process. Here we show that volunteer annotations from the *Penguin Watch* online citizen science project can be used to examine the timing of key breeding parameters across multiple sites and years, and assess the reliability of date estimates against gold standard data. We find that adult arrival is an extremely reliable parameter to extract using citizen science (differing from an expert estimate by 0.9 days (mean average; $\sigma = 1.1$), while adult departure is the least reliable (with a mean difference of 4.4 days; $\sigma = 6.4$). We show that trends in chick 'nearest neighbour distances' can be used to detect chick aggregation (i.e. huddling) behaviour, and find that days with the tightest aggregations are colder on average than the surrounding two week period. We suggest a new diagnostic parameter for 'peak crèche' based on chick to adult ratios, which show similar patterns amongst samples. These findings highlight the potential for citizen science-aided 'big data' studies in the field of ecology, and suggest that new key breakthroughs will be made when they are combined with more traditional hypothesis-driven methods.

Introduction

The study of recurring stages within the life-history of an organism, termed phenology, is a burgeoning field, with changes in the timing of key annual stages offering insight into adaptation (Phillimore et al., 2010) and changing environmental conditions (Black, 2016; Edwards & Richardson, 2004). When predicting the future health of a population or species, monitoring phenological shifts is essential, not least because changes may occur at different rates across functional groups and trophic levels, leading to a phenomenon termed "trophic mismatch" (Both et al. 2009; Edwards & Richardson, 2004; Post & Forchhammer, 2008). In avian species, avoiding mismatch between the timing of breeding and peak food availability is critical to adult and offspring fitness (Campbell et al., 2019; Cresswell & McCleery, 2003; Thomas et al., 2001; Van Noordwijk et al., 1995), yet a colonial breeding strategy – as adopted by many seabird species – may make it challenging to shift in alignment with environmental conditions (Youngflesh et al., 2017). Furthermore, differing degrees of adaptation amongst sympatric species may lead to greater overlap in breeding cycles, intensifying competitive interactions (Lynch et al., 2012a).

Lynch et al. (2012a) recorded greater plasticity in clutch initiation date (CID) in Gentoo penguins (*Pygoscelis papua*) than in the congeneric Adélie (*P. adeliae*) or Chinstrap (*P. antarcticus*) species, being able to shift CID in response to warming temperatures by 3.2 days per °C – almost twice as much as its sister taxa (see also Youngflesh et al., 2017). These differing responses narrowed the time between peak Gentoo and Adélie breeding activity in warmer years, with implications for interspecies competition in sympatric colonies. These findings, combined with climate-driven changes in the relative distribution of *Pygoscelis* species in the Antarctic Peninsula region (Lynch et al., 2012b), make it

important to test whether their phenology varies across space and time if we are to assess the resilience of these species to future changes. Of the three *Pygoscelis* species, Chinstraps have been least studied in terms of their phenology (Black, 2016), with known dates mainly restricted to the South Shetland Islands (e.g. Moreno et al., 1998; Trivelpiece et al., 1984; Trivelpiece & Volkman, 1979) and South Orkney Islands (e.g. Carlini et al., 2005; Lishman, 1985; Rootes, 1988).

Traditionally, behavioural ecology has involved intensive observational (e.g. Charmantier et al., 2008; Croxall et al., 1990; Coulson et al., 2001) or manipulative (e.g. Cox et al., 2019; Neff, 2003; Rebstock & Boersma, 2011) studies at a single location, with subsequent inference to a wider population. While these studies are vital in revealing temporal trends, inferring mechanisms and examining key behaviours, they are – by their nature – subsamples of a larger population. As the environment, predators and conspecific conditions can vary over the extent of a population, there is no implicit justification for phenological dates, population trends or behavioural mechanisms inferred at one location to be extrapolated across a whole species' range. Survival rates are known to vary in some species between the centre and edge of their range (Anderson & Boutin, 2002; Kolbe & Janzen, 2002), and phenology is influenced by a number of factors including photoperiod, temperature and precipitation (Crick & Sparks, 1999; Körner & Basler, 2010; Zhou et al., 1995). Furthermore, collective animal behaviours like schooling or huddling may depend on external drivers such as temperature or predation risk (Le Bohec et al., 2005; Weetman et al., 1999). For example, Black et al. (2016) found that aggregation behaviour in Gentoo penguins during the crèche phase varied across latitudes, and emphasised the importance of sampling multiple colonies across a species' range.

Studies across wide geographical and temporal ranges can be practically and financially challenging, but remote-sensing networks can help to alleviate logistical obstacles. Southwell and Emmerson (2015) have demonstrated that arrival and lay dates of Adélie penguins can be successfully extracted from time-lapse imagery, and using the *Penguin Watch* remote camera network Black et al. (2018) substantially reduced the Chinstrap knowledge gap, determining the timing of key phenological stages at five locations in the Scotia Arc region, including two on the Antarctic Peninsula. However, manual annotation of time-lapse imagery can be a laborious and time-consuming process, and may prove a challenge to small research teams hoping to employ similar camera technologies in future studies. Here we use the same *Penguin Watch* network to identify parameters of interest and to assess whether citizen science can be used to reliably extract penguin phenological dates from time-lapse imagery, presenting data from eight sites in the Scotia Arc region. This includes data from additional years, and two new sites on the Antarctic Peninsula – the Gaston and Tetrad Islands. We also investigate whether citizen science annotations of time-lapse imagery can allow us to detect crèching behaviour through analysis of spatial distribution patterns, and briefly assess the relationship between huddling and temperature, highlighting the potential for 'big data' approaches to ecological questions. This draws similarities with the analysis of optic flow patterns through machine learning, which has been used to shed light on animal behaviour (Shen et al., 2019; Steen et al., 2013) and welfare (Dawkins et al., 2009; Dawkins et al., 2012), and can detect effects not hypothesised by the researcher. It is likely key breakthroughs in ethological and ecological research will stem from an increase in the use of this methodology, with results feeding into more traditional hypothesis-driven approaches to consolidate and explain findings.

Materials and Methods

Monitoring Chinstrap colonies via a remote time-lapse camera network

Remote time-lapse cameras (Reconyx HC500 Hyperfire Trail Cameras; commercially available from Reconyx Inc., Holmen, WI, USA) were used to monitor Chinstrap penguin colonies at eight locations on the South Shetland Islands (n = 3), Antarctic Peninsula (n = 4) and South Sandwich Islands (n = 1) over various time periods between the 2010-2011 and 2017-2018 breeding seasons (Table 1; Fig. 1). Devices were deployed as part of the *Penguin Watch* project, which currently operates a remote camera network comprising approximately 80 units in the Southern Ocean region (Jones et al., 2018).

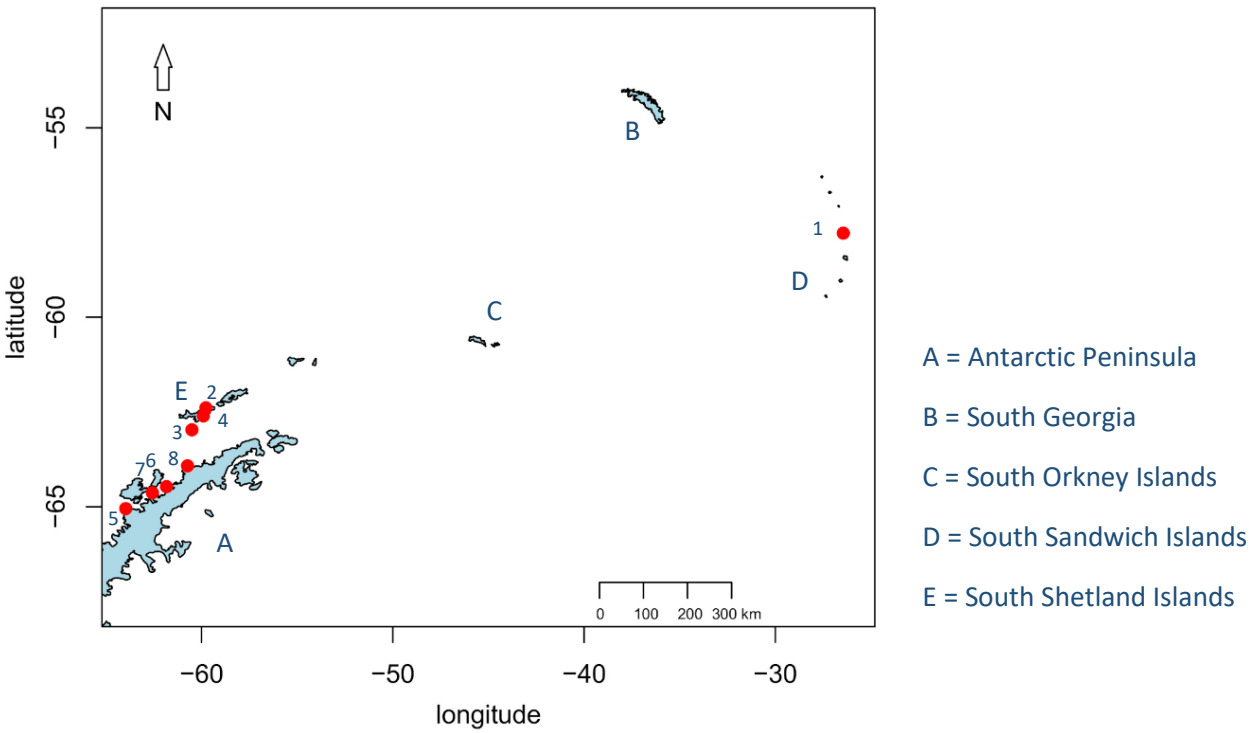


Figure 1: Map of the Scotia Arc region of the Southern Ocean, showing Chinstrap penguin (*Pygoscelis antarcticus*) study sites (red dots; numbers correspond to those in Table 1).

Following on from the work of Newbery and Southwell (2009), each camera was fixed to an aluminium scaffold pole, which was secured using either a tripod or wire rock basket (see Black et al., 2018 and Jones et al., 2018 for details). Each unit was constructed during the austral summer (October to February), and serviced the following summer if weather and ice conditions allowed. Cameras were programmed to capture images once per hour over a fixed period, and Secure Digital (SD) cards were retrieved and replaced during annual servicing.

Region	Site	Location	Camera	Breeding season
South Sandwich Islands	1. Saunders Island	57.78°S, 26.45°W	SSISAa	2013-2014
				2014-2015
South Shetland Islands	2. Aitcho Islands	62.39°S, 59.77°W	AITCd	2016-2017
	3. Baily Head, Deception Island	62.97°S, 60.5°W	BAILa	2012-2013
				2013-2014
			BAILb	2016-2017
				2017-2018
	4. Half Moon Island	62.60°S, 59.90°W	HALFa	2010-2011
				2011-2012
			HALFb	2013-2014
				2014-2015
			HALFc	2012-2013
Antarctic Peninsula	5. Booth Island	65.05°S, 63.96°W	BOOTc	2016-2017
	6. Gaston Islands	64.47°S, 61.83°W	GASTa	2016-2017
	7. Orne Harbour	64.63°S, 62.57°W	ORNEa	2013-2014
				2014-2015
				2016-2017
			ORNEb	2012-2013
	8. Tetrad Islands	63.92°S, 60.73°W	TETRa	2016-2017
				2017-2018

Table 1: List of Chinstrap penguin (*Pygoscelis antarcticus*) colony sites/breeding seasons included in this study, grouped by region. Site numbers correspond to those on Figure 1.

Image annotation using Penguin Watch

Images were uploaded to the *Penguin Watch* citizen science project (www.penguinwatch.org), on the Zooniverse online platform (www.zooniverse.org) (Jones et al., 2018). Volunteers were tasked with annotating each individual animal or object in a randomly presented time-lapse photograph, and labelling it as an 'adult', 'chick' or 'egg' (for penguins), or 'other' (for different fauna or ships); note that species identification by volunteers is not required for this project, because images contain a view of part of a single pre-identified species' colony (Jones et al., 2018). Each image was seen by four volunteers, and was presented to a further six (total ten) if any of the first four volunteers indicated that there were animals present.

Each annotation by a volunteer is represented by an xy coordinate 'click', where x and y values correspond to pixel numbers (starting from the top left-hand corner of the image). In order to derive 'consensus clicks' – i.e. single xy coordinate points generated from multiple volunteers' input (see Fig. 2 and Jones et al., 2018) – we clustered the raw clicks for each image using the open source Panoptes Aggregation software (DOI:10.5281/zenodo.2593399), programmed in Python 3.0 (Python Software Foundation, 2019). This software was specifically designed for use by Zooniverse project researchers, and has wide functionality, with tools falling into two main categories: 'extraction' (where data pertaining to individual tasks within multi-task projects are formatted and extracted) and 'reduction' (where data are streamlined or – in the case of point data – clustered). We chose to use this software owing to its specificity to our data type and requirements, but also for study repeatability. Using the Panoptes Aggregation Graphical User Interface (GUI) other researchers may aggregate their data using the same

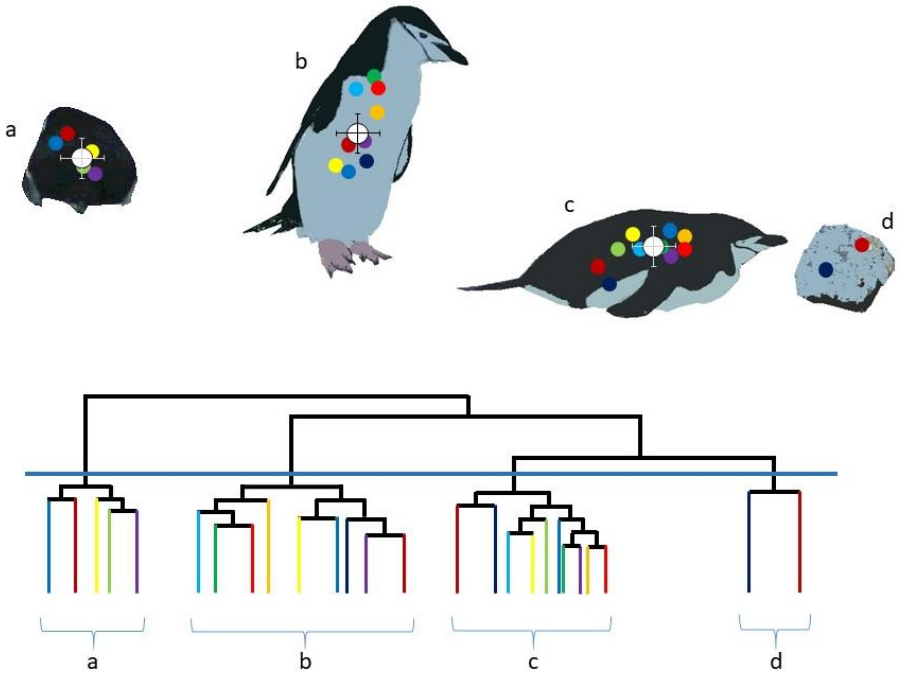


Figure 2: Images containing animals are viewed and annotated by ten different *Penguin Watch* (www.penguinwatch.org) volunteers; clicks are represented here by dots, where each colour corresponds to a different volunteer. Adult annotations, shown here, are aggregated using hierarchical clustering (HDBSCAN) to produce 'consensus clicks' (large white circles). In this example, a rock (object 'd') has been erroneously annotated. However, it has only been marked by two volunteers, and so would not be resolved as a cluster using the threshold parameter values applied here (min_cluster_size = 4). Object 'a' has been marked fewer times than object 'b' or 'c', as it is more difficult to distinguish as a penguin, but is still counted as it passes the minimum cluster threshold.

clustering algorithm, without knowledge of the Python programming language (documentation for the GUI can be found at: https://aggregation-caesar.zooniverse.org/GUI.html).

Volunteer classifications were exported as a list of raw xy coordinates, alongside associated metadata, such as volunteer username and Zooniverse image identification number. These raw clicks were formatted using the Panoptes Aggregation 'point extractor' tool, which returns a list of x and y coordinate positions for every volunteer click on each photograph. Clicks labelled as 'chick' were clustered using the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) 'point reducer', while clicks labelled as 'adult' were clustered using the HDBSCAN (Hierarchical DBSCAN) 'point reducer'. The HDBSCAN algorithm is better able to deal with obliquity in images (a common occurrence in *Penguin Watch* photographs), but, unlike DBSCAN, is unable to assign all points to a single cluster with no outliers (note that the algorithm can be adapted to allow this, but doing so will lead to a bias towards the creation of single rather than multiple clusters). There is thus a trade-off between coping with obliquity and efficiently processing single clusters; since an image is more likely to contain a single chick than a single adult, different algorithms were selected for the different groups.

The parameter values used for DBSCAN were *eps* (ε) = 5 and *min_samples* = 2, where *eps* is the maximum Euclidean distance between two points where the two points will be grouped into the same cluster, and *min_samples* can be interpreted here as the number of points required to form a cluster. The parameter values used for HDBSCAN were *min_cluster_size* = 4 and *min_samples* = 3. For HDBSCAN, *min_cluster_size* relates to the smallest number of points from which a cluster may be formed – effectively equivalent to the DBSCAN parameter *min_samples*. Increasing the value of *min_samples* for HDBSCAN coerces more points to be treated as noise, with only denser groups of points being resolved as clusters. To determine the most appropriate parameter values, we ran the

DBSCAN and HDBSCAN 'point reducers' on citizen science classifications for Damoy Point (2013-2014; n = 300) using a range of parameter values. The process was repeated with a narrower range of parameter values on data from Half Moon Island (2012-2013; n = 283)), and those which gave the most accurate 'adult' and 'chick' population counts when compared to gold standard (expert-annotated) data were chosen. When comparing the citizen science and gold standard counts, we calculated *i)* the mean average count difference in number of penguins, *ii)* the proportion of differences that were zero or one and *iii)* the proportion of differences that were overestimates/underestimates, and assessed reliability based on these three factors. Since different factors sometimes supported the use of different threshold values, 'mean average count difference' was considered the most important metric when making decisions. The choice of parameter values was also supported by our previous findings using a different clustering algorithm. In Jones et al. (2018) we report that a threshold value of *num_markings* > 1 is appropriate for chick clusters, and *num_markings* > 3 is most reliable for adult clusters, which is equivalent to the *min_samples* = 2 and *min_clust_size* = 4 used here for chicks and adults, respectively.

Using citizen science to estimate arrival and departure

To generate population time-series data (i.e. the number of adults and chicks in each image over time) from *Penguin Watch* 'consensus clicks', *Narwhal Files* were created for each site and year. *Narwhal Files* contain the following information, and are produced using the methodology described in Chapter 3 (Jones et al., in review):

- Number of adults and chicks (and eggs, if included)
- Adult and chick nearest neighbour distances (nnd)

- Chick second nnd (since a chick's first nearest neighbour is likely its sibling in the nest)
- Mean adult and mean chick movement between image *i-1* and image *i*

These metrics were visualised using *Narwhal Plots* (e.g. Fig. 3; see Chapter 3). Arrival and departure dates of adults and chicks were estimated using the population counts from noon images only (or the image closest to noon), to allow a fair comparison between sites. First arrival/chick hatch was defined as the first image in which an adult or chick was recorded (i.e. population count > 0), respectively, between two predetermined dates. Dates were constrained so that 'false starts' – owing to incorrect classifications or autumn/winter attendance by individuals – were discarded, and were determined using the trends observed in each *Narwhal Plot*. Adult departure and chick fledge were estimated by detecting the first image in the time-series which was followed by three or five 'empty' images, respectively, between two predetermined dates (as above). This was achieved by calculating the moving average population trend for each group (adults: n = 3; chicks: n = 5), and using R (v 3.6.0) to find the first occurrence of a zero mean average. The date preceding the zero mean average was extracted as the departure date. Any adult or chick classifications recorded after this point (i.e. after the series of 'zero' images) were checked manually to ensure departure date had been correctly assigned.

In order to assess the reliability of citizen science derived estimates, we also recorded phenological dates manually, to provide gold standard (expert-derived) data for comparison. To do this, we viewed the imagery from each camera in turn, examining the photographs in chronological order and noting down a) first adult appearance, b) first chick appearance (i.e. hatch) and c) adult departure. Chick fledge was not recorded, as this

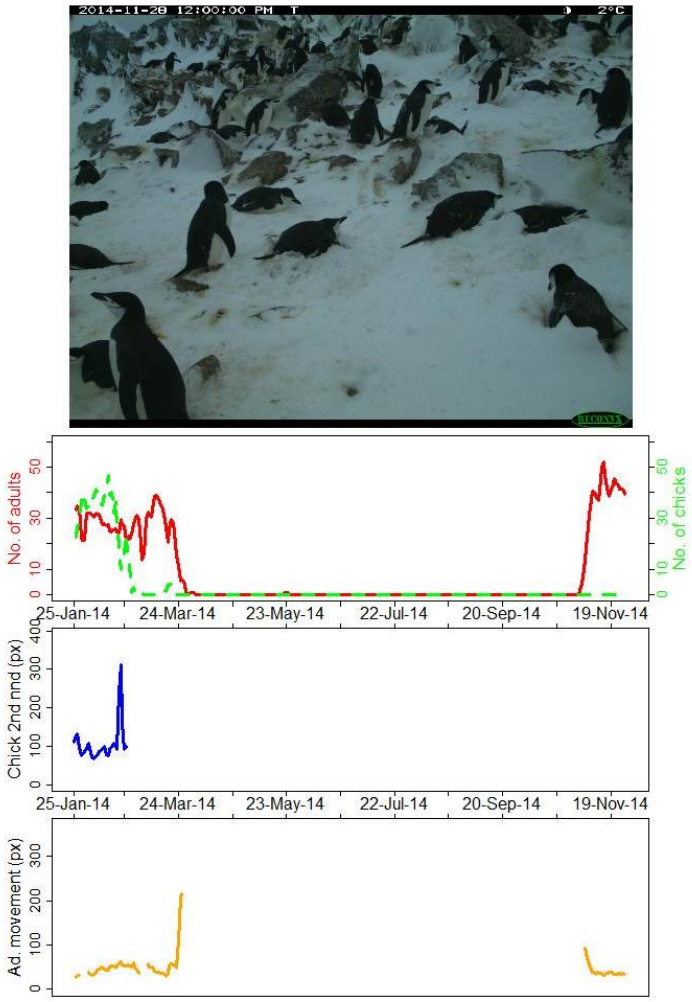


Figure 3: Example of a *Narwhal Plot*, here used to visualise population data for Orne Harbour between 25/01/2014 and 28/11/2014. From top to bottom: 1) the final time-lapse image in the dataset; 2) moving mean average (n = 3) adult (red) and chick (green) population counts; 3) moving mean average (n = 2) mean chick second nearest neighbour distance; 4) moving mean average (n = 3) adult movement between image *i-1* and image *i*. All metrics shown here were calculated using noon images only (or the image closest to noon), to allow comparison between sites (which vary in number of images captured per day).

parameter is difficult to constrain to a single day (see Discussion). Gold standard dates were recorded using both all available imagery and noon images only, to examine the influence of image frequency on the dates obtained. Note that while phenological dates for some sites and years are published in Black et al. (2018), all images were reanalysed to ensure a consistent detection method was used throughout.

Using citizen science to detect crèching behaviour

As discussed by Young (1994), the term 'crèche' has been used to mean both the period when a chick is left unguarded by a parent, and the tight aggregations formed during this stage. Here we use 'crèche' synonymously with 'post-guard phase', and the term 'huddle' to describe aggregations. To find a potential citizen science-derived parameter representative of chick crèche, we calculated the following variables across the breeding season, using all available citizen science data (i.e. all available classified images, rather than noon images only): 1) distance between each nest and its nearest adult; 2) ratio of chicks to adults per day; 3) mean chick second and third nearest neighbour distance (nnd). Code can be found at DOI: 10.5281/zenodo.3476672 (R v 3.6.0; R Core Team, 2019).

1) Nest to adult distance. The centre of each nest zone was annotated by an expert using *Penguin Watch 3.0* (a version of www.penguinwatch.org used for testing purposes), generating a set of xy coordinates for the nests in each site/year dataset. The mean average (and standard deviation) distance between each nest centre and the nearest annotated adult was calculated for each image. Images containing fewer than five annotated adults (as identified in the *Narwhal Files*) were not incorporated, as this

suggests the image was partially obscured, and would lead to calculation of erroneously high nest-adult distances.

2) Ratio of adults to chicks. The number of chicks per adult per day was calculated by dividing the daily mean number of chicks by the daily mean number of adults. Daily mean ratios were used to ensure that trends were not masked by hourly fluctuations in adult/chick count (e.g. owing to snow on the camera lens or chicks sheltered by adults during high winds, and therefore out of view).

3) Chick second and third nnd (code to calculate first and second nnd is also provided in the *Narwhal Script* (R v 3.6.0; R Core Team, 2019; see Chapter 3), which can easily be adapted to other nnds by altering the value of k in "mean(nndist(currentchickxy, k=))"). Distances were only calculated when there were more than five chicks annotated in an image, as identified in the *Narwhal Files*. Counts under five suggest that many chicks are obscured, and distances between 'neighbours' will therefore be artificially high. Assuming there are usually two chicks per nest, the second nnd metric gives a more informative measure of chick spatial distribution than first nnd, which will return the (small) distance between siblings during the guard phase. Assuming chicks are clearly visible and correctly annotated, the mean chick second nnd will decrease when chicks begin to huddle, providing a possible indicator of the crèche phase. Sometimes chicks can be 'double-marked' owing to the parameters used for the DBSCAN clustering algorithm (changing the parameters to avoid this issue can increase the occurrence of 'missed' chicks, so there is a trade-off between under- and over- estimation). Chick third nnds were therefore also calculated, to investigate and mitigate this effect.

To minimise noise all metrics were smoothed using a loess-smoother (function 'loess'; package 'stats') or a moving average (function 'rollmean'; package 'zoo') (using R v 3.6.0 (R Core Team, 2019)). Different smoothing parameters were selected for each of these metrics in order to achieve the most meaningful graphs, and are outlined in the legends of their respective graphs.

We also recorded an expert-derived crèche date for each site and season. Since an exact start date can be difficult to discern, we provided a subjective 'certainty' score alongside each estimate, on a scale of one (very uncertain) to ten (very certain). In addition, we indicate the amount of huddling behaviour observed on that date via a 'Huddle Score', ranging from one (no evidence of huddling) to ten (large distinctive chick aggregation(s)); see Appendix 6a for an example image of each Huddle Score.

Assessing temperature as a driver of huddling behaviour

Air temperature data (°C) were recorded by each camera device, and were extracted from raw time-lapse images using the R package 'exiftoolr' (R v 3.6.0; O'Brien, 2019), which employs an ExifTool (https://sno.phy.queensu.ca/~phil/exiftool/) to read Exchangeable Image File Format (EXIF) data. To examine the relationship between huddling behaviour and temperature, the mean average, minimum and maximum daytime (here defined as 0900 – 1700) temperatures over a two week period – one week before and one week after the 'maximum huddle' date (the day where chick third nnds were lowest) – were calculated for each site (see Appendix 6b – Table 1). These were then compared to the mean average, minimum and maximum daytime (0900-1700) temperatures for the 'maximum huddle' date (Appendix 6b – Tables 1 and 2).

Results

Using citizen science to estimate arrival and departure

Estimates of adult arrival, chick hatch, chick fledge and adult departure are shown in Table 2; dates are given as Julian Days, where 31st December = 365 and 1st January = 366 to ensure they are numbered continuously over the breeding season. Alongside the dates derived via citizen science, expert estimates – based on a) one image per day (usually taken at noon) and b) all available image data – are provided in parentheses for adult arrival/departure and chick hatch. For example, a result for adult arrival of '300 (-2,-3)' denotes an estimate of Day 300 (27th October) using citizen science (noon images), two days previous (i.e. Day 298 (25th October)) using expert analysis (noon images) and three days previous (i.e. Day 297 (24th October)) using expert analysis (all available image data). Expert estimates are not provided for chick fledge, as this parameter is difficult to define – see Discussion.

The greatest error – defined as the difference between citizen science and expert-derived estimates – is associated with adult departure, where the deviation between groups is 4.4 days on average ($\sigma = 6.4$) when using noon images only, and 9.8 days on average ($\sigma = 9.1$) when the expert has access to all available time-lapse imagery (n = 11 site/years; note that 'average' here refers to the mean average). The most accurate estimates are obtained for adult arrival, where the difference between citizen science and expert-derived estimates is 0.9 days on average ($\sigma = 1.1$) for noon images, and 2.3 days on average ($\sigma = 1.1$) for all available images (n = 7 site/years). The error associated with chick hatch is 4.1 days on

Region	Site	Breeding season	Adult arrival	Chick hatch	Chick fledge	Adult departure
South Sandwich Islands	SSISAa	2013-2014		370 (-4, -6)	433	493 (0, +13)
		2014-2015	309 (-1, -1)			
South Shetland Islands	AITCd	2016-2017			405	471 (0, +1)
	BAILa	2012-2013		366 (-7, -7)	418	452 (+10, +10)
		2013-2014	294 (-1, -1)	362 (-7, -7)		
	BAILb	2016-2017		358 (-4, -6)	402	406 (+22, +34)
		2017-2018	300 (-2, -3)			
	HALFa	2010-2011				446 (+1, +2)
		2011-2012	309 (0, -1)			
	HALFc	2012-2013		365 (-1, -1)		
	HALFb	2013-2014			420	469 (0, +2)
		2014-2015	300 (0, -3)			
Antarctic Peninsula	BOOTc	2016-2017			419	450 (0, +5)
	GASTa	2016-2017		371 (-6, -8)	417	
	ORNEb	2012-2013			410	467 (+5, +10)
	ORNEa/b	2013-2014	309 (-3, -4)	372 (0, -5)	428	457 (0, +18)
	ORNEa	2014-2015	307 (0, -3)			
		2016-2017			407	470 (+4, +7)
(Grey cells = image data unavailable)	TETRa	2016-2017			408	476 (+6, +6)
		2017-2018	302 (0, -2)			

Table 2: Adult arrival, chick hatch, chick fledge and adult departure dates (as Julian dates, where 31st December = 365 and 1st January = 366) as estimated by citizen science, using noon images. Numbers in parentheses represent the difference in date when calculated by 1) an expert using noon images only, and 2) an expert using all available time-lapse data. For example, a recording for adult arrival of '300 (-2,-3)' denotes an estimate of Day 300 (27th October) using citizen science (noon images), Day 298 (25th October) using expert analysis (noon images) and Day 297 (24th October) using expert analysis (all available data).

average (σ = 2.6) and 5.7 days on average (σ = 2.1) for noon images and all available data, respectively (n = 7 site/years). As such, the error associated with estimation from noon images is similar to that associated with adult departure, but the discrepancy between citizen science and expert-derived dates is ~4 days fewer on average (as compared to adult departure) when the expert has access to all available image data. Fig. 4 shows the variation in phenological dates across sites and years, based on expert-derived (adult arrival/departure, chick hatch) and citizen science-derived (chick fledge) estimates.

Using citizen science to detect crèching behaviour

Graphs showing trends of a) nest to adult distance, b) chick to adult ratio and c) chick third nnd are presented in Fig. 5 (see Appendix 6c for all colonies/years over the chick-rearing period).

Nest-adult distance

As seen in Fig. 5a and Appendix 6c, there is generally a trend of increasing nest to adult distance while chicks are present at the breeding site. Since a chick is left alone at the nest for the first time during the post-guard phase, we hypothesised that a sudden increase in nest to adult distance would correspond with the start of this period. Instead, movement of adults away from the nest appears to be a gradual process, and – with the exception of BAILb2017 and ORNEa2017 (see Appendix 6c) – peak adult to nest distances are loosely aligned with fledge date, at which point crèche must have already taken place. Since there are no consistent identifiable inflection points, nest to adult distance has not been used to extract potential crèche dates from the citizen science data.

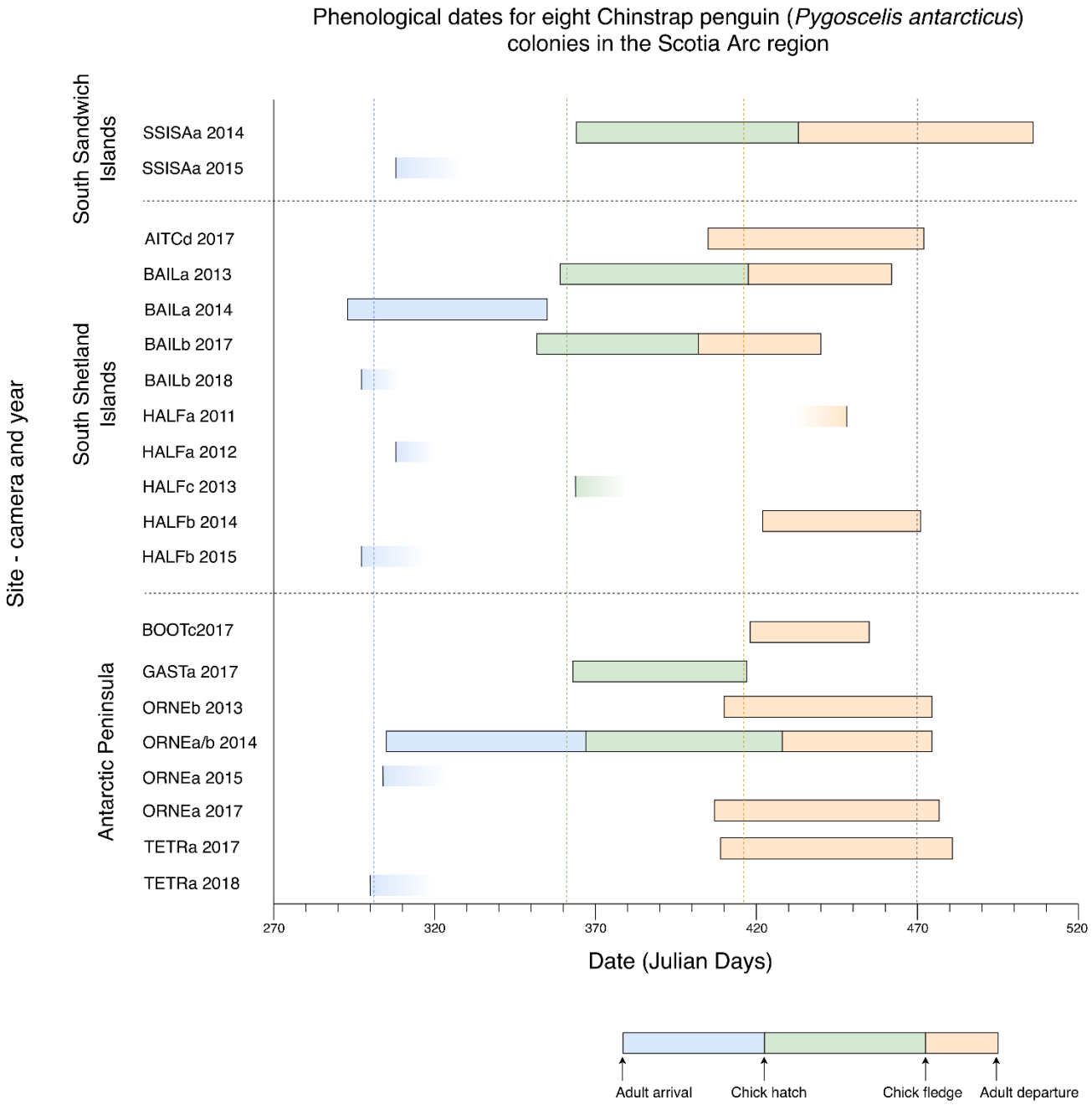


Figure 4: Phenological dates in Julian days (adult arrival (beginning of blue bar), chick hatch (beginning of green bar), chick fledge (beginning of peach bar) and adult departure (end of peach bar)) for the eight Chinstrap penguin (*Pygoscelis antarcticus*) sites (20 site/year samples) included in the study. Dates are derived from expert estimations, with the exception of chick fledge, which is based on citizen science classifications. Bars with fading colour show that images are not available to determine when the next phenological stage begins. Vertical dotted lines are plotted for the mean date of each parameter across sites.

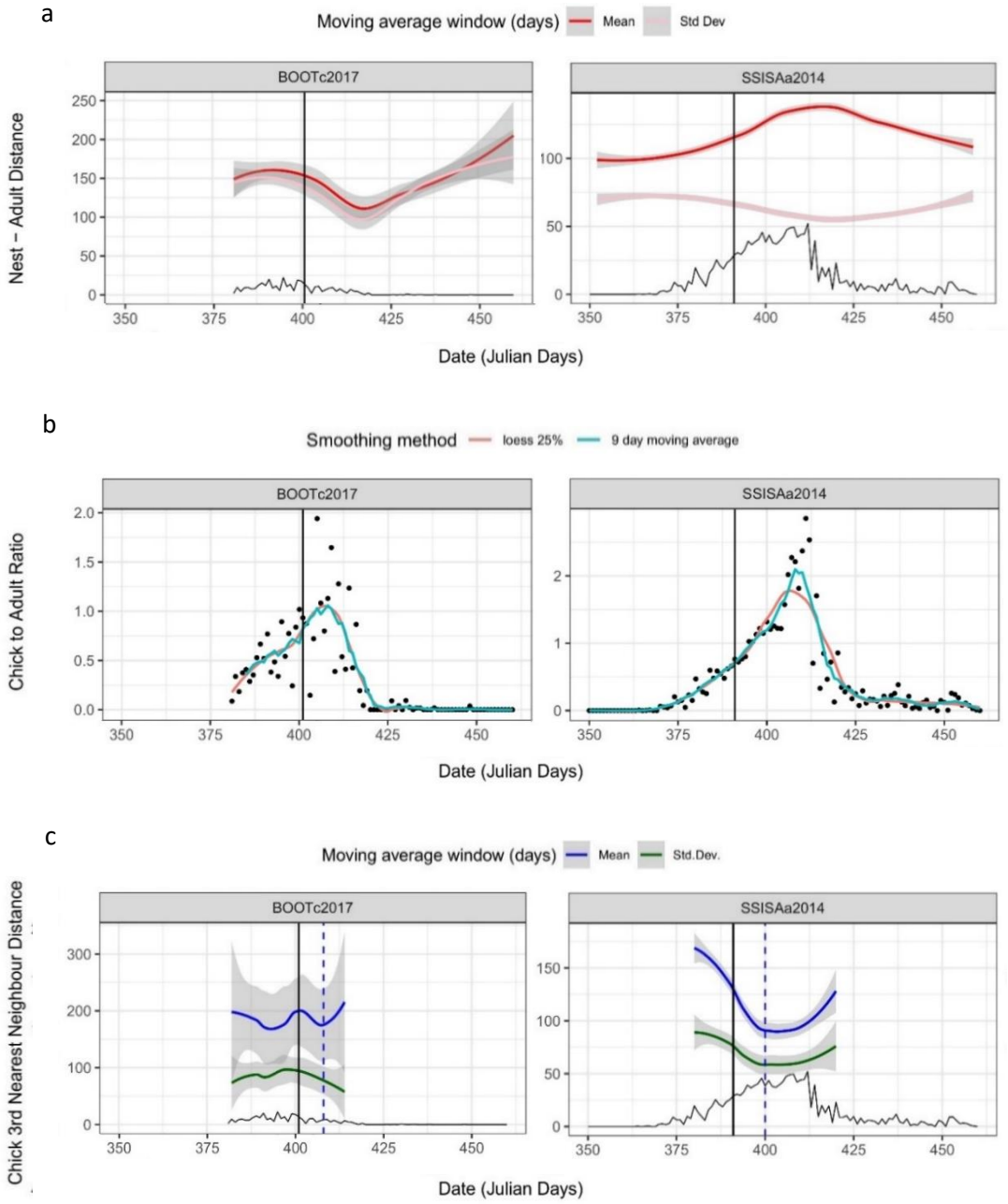


Figure 5: a) 50% loess-smoothed mean (red line) and standard deviation (pink line) for nest-adult distances; grey shaded zones show standard error; b) ratio of chicks to adults (daily mean chicks divided by daily mean adults). A 25% loess smoothing curve (pink line) and 9 day moving average (green line) are fitted, and days corresponding to their maxima are displayed in Table 3; c) 50% loess-smoothed chick third nearest neighbour distances (nnd) – mean (blue line) and standard deviation (green line). Grey shaded zones show standard error. The date corresponding to the minimum daily mean chick third nnd is shown (blue

hashed line) and given in Table 3. Graphs are shown for colonies at Booth Island (2016-2017) and Saunders Island (2013-2014) across the chick-rearing period. Vertical black lines represent crèche date as estimated by an expert; where present, horizontal black lines show number of chicks per day. Graphs for the other sites can be found in Appendix 6c.

Chick to adult ratio

Every site and season shows a similar pattern with a clear peak in chick to adult ratio when fitted with a 25% loess smoothing curve (Fig. 5b and Appendix 6c; pink line) or a nine day moving average (Fig. 5b and Appendix 6c; green line). While the position of the peak varies with respect to the expert estimation of crèche date, the maximum chick to adult ratio – i.e. when there are most chicks in relation to the number of adults – could be a useful parameter of 'peak crèche'. Dates (in Julian days) corresponding to the maximum chick to adult ratios at each site were calculated from the 25% loess smoothing curve (which is less sensitive to local fluctuations than the nine day moving average; see Fig. 5b), and are provided in Table 3.

Chick second and third nearest neighbour distance (nnd)

Mean chick second and third nnds showed near-identical trends. Since third nnd is more robust to cases of 'double-marking' of individuals, the results associated with this metric are presented here (see Fig. 5c). Most sites show a pattern of decreasing third nnd, followed by a reversal of the trend later in the season. The initial decrease is presumably linked to the hatching and growth of chicks – when more individuals are visible and annotated, nnds will by definition decrease. However, if chicks were annotated correctly, the greatest dip in third nnds could represent huddling (a key characteristic of the crèche

stage), and therefore act as a metric for the post-guard phase, albeit – like chick to adult ratio – not necessarily a signal of its beginning. As such, minimum third nnds were extracted for each site, and the dates corresponding to these minima are provided in Table 3. Where multiple 'dips' in third nnd were observed (e.g. BOOTc2017; see Fig. 5c), local minima were extracted, and the dip with the lowest third nnd was used.

Site	Breeding season	Crèche date					
		Citizen science			Expert		
		Adult/chick ratio	Chick 3rd nnd	Huddle score	Estimate	Certainty	Huddle score
AITCd	2016-2017	400	394	9	391	9	8
BAILa	2012-2013	399	395	9	384	9	5
BAILb	2016-2017	381	372	1	386	8	3
BOOTc	2016-2017	407	408	4	401	5	1
GASTa	2016-2017	398	402	3	405	4	3
ORNEb	2012-2013	401	388	1	408	5	1
ORNEa	2013-2014	406	403	3	414	6	3
ORNEa	2016-2017	399	392	6	405	4	2
SSISAa	2013-2014	406	400	9	391	7	5
TETRa	2016-2017	398	395	4	405	4	1

Table 3: Estimates of crèche date (in Julian Days), as calculated using a) variables derived from citizen science data (chick to adult ratio and chick third nearest neighbour distances (nnds)) and b) expert analysis of images. A certainty level (1 = uncertain, 10 = certain) is provided alongside the expert estimates, as a subjective measure of how difficult/easy it was to detect characteristic crèche behaviours (e.g. huddling) in the raw images. Owing to the nature of the parameters used, the citizen science dates relate to 'peak crèche' (for third nnd, 'maximum huddle'), whereas the expert dates are a best estimate of the crèche start date. A Huddle Score (see Appendix 6a) is shown for each site, as a subjective indication of the degree of huddling observed i) on the date with the lowest mean average chick third nnds and ii) the expert estimate of crèche start date.

Assessing temperature as a driver of huddling behaviour

Mean, minimum and maximum daytime (0900 to 1700 hours) temperatures for the 'maximum huddle' date (see Table 3: 'Chick 3rd nnd') and the surrounding two week period are shown in Table 4.

Site	Season	'Maximum huddle' date (one day)			Surrounding two week period (15 days)		
		Air temperature (°C)					
		Mean (std dev)	Minimum	Maximum	Mean (std dev)	Minimum	Maximum
AITCd	2017	4.00 (0.50)	3	5	5.01 (2.10)	1	11
BAILa	2013	1.22 (1.09)	0	3	2.13 (2.08)	-2	9
BAILb	2017	3.67 (0.87)	2	5	5.73 (3.24)	1	19
BOOTc	2017	2.44 (1.13)	1	4	4.95 (3.58)	-2	15
GASTa	2017	5.44 (1.88)	4	10	5.04 (3.11)	0	14
ORNEb	2013	4.89 (1.45)	2	7	4.44 (4.12)	-1	16
ORNEa	2014	-1.44 (0.73)	-2	0	2.61 (4.27)	-4	18
ORNEa	2017	2.67 (0.71)	2	4	5.70 (4.03)	1	20
SSISAa	2014	0.71 (0.76)	0	2	2.92 (2.09)	0	10
TETRa	2017	4.00 (1.22)	2	6	4.56 (3.76)	0	19

Table 4: Mean average (and standard deviation), minimum and maximum daytime (0900 – 1700) air temperatures (°C) for 'maximum huddle' dates (determined using third nearest neighbour distances (nnds), see Table 3), and the two week period surrounding these dates – i.e. one week before, one week after, and the date itself (15 days in total). For the 'maximum huddle' dates, n = 9; for the 15 day periods, n = 135 (with the exception of SSISAa2014, where photographs were captured less frequently, and n = 7 and n = 105, respectively).

The mean average temperature on the 'maximum huddle' date (μ = 2.76, σ = 2.11) was lower than the mean average temperature of the surrounding two week period (μ = 4.31, σ = 1.29), *t*(9) = -3.32, p < 0.01 (one tailed) (see Appendix 6b). The corrected Hedge's g effect size = -0.85, using formulae 1-3.

$$g^* = \frac{\overline{x_1} - \overline{x_2}}{s^*} \times correction\ factor \quad (1)$$

$$s^* = \sqrt{\frac{{s_1}^2(n_1 - 1) + {s_2}^2(n_2 - 1)}{n_1 + n_2 - 2}} \quad (2)$$

$$correction\ factor = 1 - \frac{3}{4(n_1 + n_2) - 9} \quad (3)$$

Where, given Group 1 comprises the mean temperatures on the 'maximum huddle' date and Group 2 comprises the mean temperatures over the surrounding two week period:

$\overline{x_1}$ is the mean of Group 1 (2.760),

$\overline{x_2}$ is the mean of Group 2 (4.309),

s_1 is the standard deviation of Group 1 (2.107),

s_2 is the standard deviation of Group 2 (1.292),

n_1 is the sample size of Group 1 (10), and

n_2 is the sample size of Group 2 (10).

In 80% of cases, the mean average temperature on the 'maximum huddle' date was lower than for the surrounding two week period, with an average difference in means of -1.55°C (range = 0.45°C warmer to 4.05°C colder). Mean temperatures (with 95% confidence intervals) are shown in Fig. 6.

The minimum temperature for the 'maximum huddle' date always exceeded (or in one case was equal to) the minimum temperature for the surrounding two week period, but the maximum temperature was always lower on the 'maximum huddle' date.

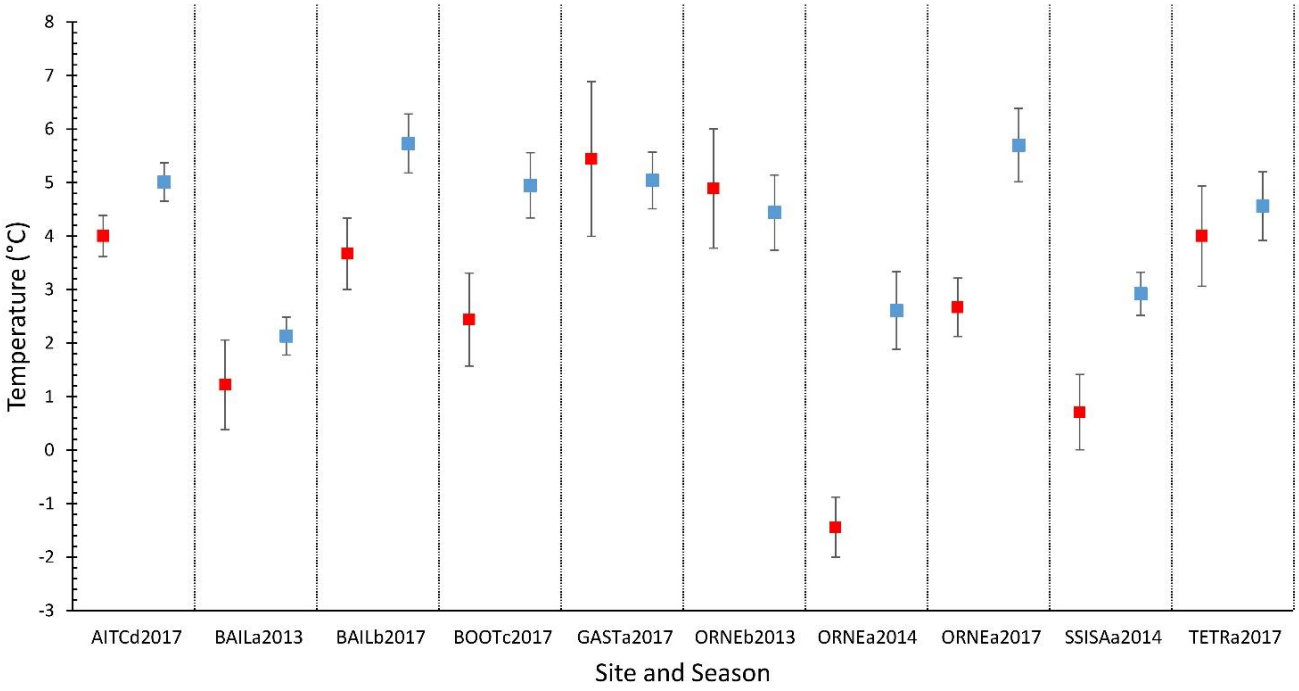


Figure 6: Mean average daytime (0900 – 1700 hours) temperatures (°C) for the 'maximum huddle' date (see Table 3: 'Chick 3rd nnd') (red) and the surrounding two week period (blue), with 95% confidence intervals, for each Chinstrap penguin (*Pygoscelis antarcticus*) colony studied.

A simple linear regression, and visual analysis of the data points, indicated no strong linear relationship between the citizen science 'Huddle Score' (see Table 3) and the difference in mean temperature between the two samples (R^2 = 0.01), or the average mean temperature on the 'maximum huddle' date (R^2 = 0.11); see Fig. 7.

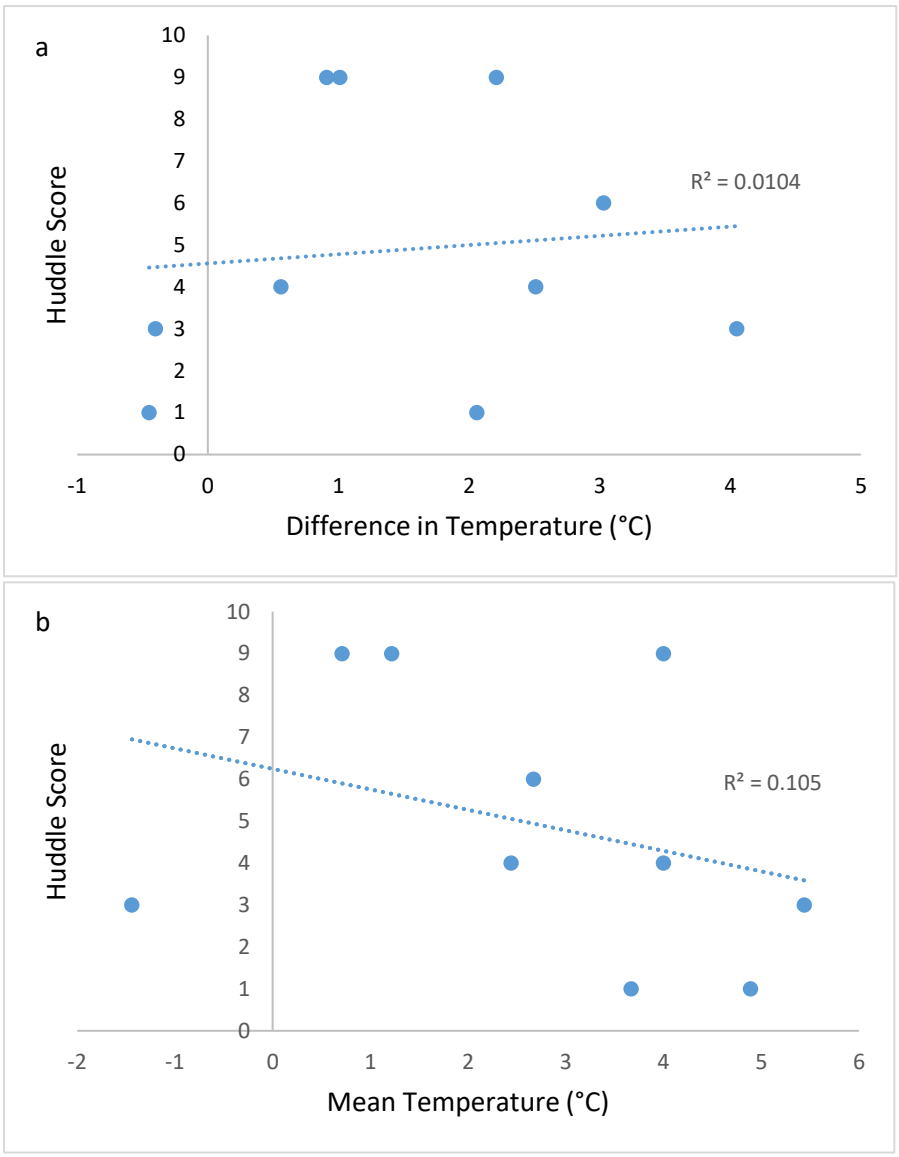


Figure 7: a) The relationship between the citizen science 'Huddle Score' (see Table 3) and the difference in mean daytime (0900 – 1700 hours) temperature (°C) between the 'maximum huddle' date (see Table 3: Chick 3rd nnd) and the surrounding two week period; b) the relationship between the citizen science 'Huddle Score' and the mean daytime temperature on the maximum huddle date. Linear regression lines have been plotted on both graphs, and the R^2 values are displayed.

Discussion

Using citizen science to estimate arrival and departure

The use of citizen science to detect adult arrival, adult departure, and chick hatch has potential, but currently requires manual fine-tuning by an expert. In this respect, it is already a tool for researchers, as it reduces the time needed to pinpoint key dates by providing a starting point (image or date). It could be argued that as long as error is consistent across datasets, citizen science-derived dates could be used to make comparisons between colonies. However, while relative dates (e.g. knowing whether one colony departs before another) are important, robust calendar dates are essential for comparisons with pre-existing and future datasets, and for analysis against temporally changing environmental parameters.

Here we define error as the difference between an estimate and the gold standard (i.e. expert-derived data), in number of days; of course, it is important to remember that any estimate, including the gold standard, may differ from the underlying truth. It is difficult to parse citizen science-derived error from error owing to a lack of data, but here we attempt to crudely do so by calculating the number of days' difference between the citizen science-derived estimates and a) gold standard estimates derived from noon images only ('error 1'), and b) gold standard estimates derived using all available data ('error 2'). We then divide 'error 1' by 'error 2' to calculate the proportion of error that can be accounted for solely by the citizen science process (since the same number of images were used for both methods). If 'error 2' is greater than 'error 1', the additional discrepancy must be the result of the expert having access to additional information, and thus is attributed to a lack of data.

The greatest error rates were associated with adult departure and chick hatch, although the primary cause of this error seems to differ between the two variables. Indeed, 38.3% of the error associated with estimation of adult departure was owing to missed annotations (i.e. citizen science), while 61.7% of the error was accounted for by only having access to one image per day. Contrastingly, 72.6% of the error associated with estimation of chick hatch was linked to citizen science, with the remaining 27.4% a consequence of the number of images sampled. This is in alignment with our previous findings (see Jones et al., 2018) that volunteers annotate adults more reliably than chicks. In Jones et al. (2018) we suggest that obscuration of chicks by adults is one possible cause of this, and this will be especially true just after hatch, when chicks are smallest and most concealed by their parents (Southwell & Emmerson, 2015). Regardless of the citizen science error, however, it is clear that a sample size of one image per day is insufficient for accurate estimation of these phenological parameters. We suggest that a comparison of accuracy is performed using different sample sizes, to find the most appropriate trade-off between volunteer effort and reliability. Since Black et al. (2018) found no significant difference in dates derived using minute-interval photographs and six images per day, we suggest a comparison between six and five images per day as a starting point.

Interestingly, estimation of adult arrival was substantially more reliable than estimation of adult departure. Here, 42.7% of error was associated with the citizen science method, while 57.3% was accounted for by the sample size of images used. We suggest this can be explained by the differing nature of adult arrival and departure. When adults arrive at a site, they generally remain there to establish a nest, meaning they will be detected in the first noon image following their arrival. In contrast, the leaving process takes a longer

period of time (i.e. it is less constrained than arrival date), with some individuals returning to the site sporadically after the majority of the colony has left. As such, the site may be empty at noon, but occupied during other parts of the day. We therefore recommend that the images uploaded to *Penguin Watch* are evenly spaced throughout the day and night, in order to capture these behaviours. Alternatively, a different definition of leaving date – for example, where >80% of penguins have left the breeding site for five consecutive days – could prove more robust (although images with poor visibility would need to be discarded prior to this analysis).

Identifying chick fledge is more challenging than adult departure, as it is difficult to distinguish chicks from adults once they have undergone moult. Without tagging numerous individuals, this would also be a problem in the field, and is not a consequence of time-lapse monitoring or the citizen science method. Instead, we suggest that citizen science provides a novel way of estimating fledge date. Volunteers are presented with time-lapse images at random (i.e. not in chronological order) and are therefore not subject to the same biases as an expert viewing a chronologically-organised dataset. Volunteers will mark juveniles as adults once they have lost their down – in other words, when they no longer recognise them as a chick – which means we can extract a date linked to the immediate post-moult period. Although individual volunteers may tag a chick beyond this point, more than one marking is required for the click to be carried forward (see Methods), and the cut-off point is thus detected via a 'wisdom of the crowds' logic (Galton, 1907). As such, we here define chick fledge as the date on which no chicks are recognised, although they may still be present at the site. This parameter is useful for comparison between colonies/years in this study, and allows a date associated with chick fledge to be recorded,

when it would otherwise be challenging to do so. However, as mentioned above, care must be taken when comparing dates to other studies, which may employ a different definition of fledge.

Finally, It should be noted that the thresholds used to determine departure dates (three and five consecutive 'zero' days for adults and chicks, respectively) were appropriate for use. After quality checking, the original adult departure dates were retained in 72.7% of cases (eight out of 11 sites). Of the remaining three sites, two had correct classifications after the threshold, and one image dataset contained Gentoo penguins (*P. papua*) before the threshold, meaning classifications were not necessarily incorrect, but of the wrong species.

For the eleven sites where chick fledge dates were extracted, there were 54 cases where images after the series of 'zero' days were classified as containing chicks. Of these, 46 (85.2%) were indeed incorrect, usually because moulting adults had been mistaken for chicks. The remaining eight were false negatives. For two sites, false positives were detected; one moulting adult was tagged prior to the threshold for BAILa2013, and moulting adults were mistaken for chicks prior to the SSISA2014 threshold in 26 images. Reducing the threshold to four empty days would not have resolved this issue, because there was never a string of more than one zero for these 26 images. Furthermore, reducing the threshold may have led to a greater frequency of false negatives. It should be noted that chicks were also erroneously annotated in the HALFa2011 dataset, but no fledge date was extracted for the site, as the image dataset started after this point.

It is important to remember that attendance (and indeed behaviours such as huddling) may occur outside the field-of-view of a camera. However, Southwell and Emmerson (2015) have shown that adult arrival dates of Adélie penguins can be extracted from imagery with minimal error (camera-based estimates were later than direct field-based observations by just 0.9 days on average). This error should of course be taken into account when comparing estimates generated via different means, but we argue that the benefits of a camera network – particularly the temporal and spatial breadth of analysis it facilitates – outweigh the associated error.

Fig. 4 shows our best estimates of phenology for the sites and seasons included in the study, using a combination of expert- (adult/chick arrival; adult departure) and citizen science-derived (chick fledge) dates. In future, the dates provided here could be combined with the findings of other studies (e.g. Bagshawe, 1938; Black et al., 2018; Lishman, 1985; Rootes, 1988) to assess the relative importance of location in determining key phenological dates. For example, it appears that longitude may have an effect on departure date, with the most easterly colony (Saunders Island, South Sandwich Islands, 57.78°S, 26.45°W) departing 24 days after the second latest colony. Bost and Jouventin (1991) suggest that the winter breeding observed in Gentoo penguins on the Crozet Islands (the northernmost edge of the species' range) is linked to the timing of peak food availability in the region. It is possible that the late departure of Chinstraps from the South Sandwich Islands similarly coincides with a peak in resource abundance. Temperature may also play a role, although colonies arriving earlier or departing later are generally subject to cooler conditions, meaning care must be taken not to confuse cause and effect.

Using citizen science to detect crèching behaviour

Huddling is often considered a diagnostic behaviour of the *Pygoscelid* crèche phase (Black et al., 2016), and as such, the ability to automatically detect an aggregation of chicks from citizen science data would be a useful indicator of crèche date. Our results support the use of chick third nnds to identify huddling; for example Fig. 8 shows a clear chick aggregation within the colony at Baily Head on 30/01/13 (Julian day 395) – the date with lowest daily mean average chick third nnd. While obvious huddling behaviour was not always observed on the day when mean chick third nnd was lowest (see Table 3: Citizen science – Huddle Score), we suggest that this is not necessarily a failure of the method, but that less huddling behaviour took place at these sites in these years. This is supported by the Huddle Scores given for the expert-derived crèche dates, which show (through visual analysis) that strong chick aggregations did not always occur on these dates either.

Our work supports the hypothesis that the post-guard phase is not always associated with large chick aggregations, and rather that a spectrum of huddling behaviour exists, and in some cases, huddling may not occur at all. Black et al. (2016), observed differences in aggregation behaviour in Gentoo penguins (*P. papua*) across latitudes, mainly attributing these variations to differential exposure to cold temperatures and precipitation. The extent of huddling may also depend on the amount of adult aggression (Le Bohec et al., 2005; Penteriani et al., 2003), degree of colony synchronicity (Black et al., 2016; Viñuela et al., 1996; Williams & Croxall, 1991) and/or predation risk (Wilson, 2009). As such, estimation of crèche dates based solely on huddling behaviour would be unwise, and a bias towards searching for chick aggregations likely led to the very late expert crèche dates estimated for GASTa2017, ORNEa/b 2013/14/17 and TETRa2017 (which had a mean

Huddle Score of 2). Future studies could look at the degree of huddling across a larger sample of sites and years, to investigate whether the variations in behaviour are primarily determined by external factors such as latitude, temperature, precipitation and predation risk, or internal drivers such as adult aggression or colony synchronicity.



Figure 8: An image from dataset BAILa2013, on the date corresponding to the minimum mean average third nearest neighbour distance – 30/01/13 (Julian day 395). A huddle of chicks can be seen in the top right of the image (circled in red).

Our results also support the idea that crèche has no well-defined start date, but is instead a gradual process, presumably owing to asynchronicity in hatch date (Moreno et al., 1997; Penteriani et al., 2003; Viñuela et al., 1996). This is based on our calculations of adult-nest distance, where instead of a sudden increase in mean distance between each nest and its nearest adult – as would be expected if all adults left their chick unguarded on the same day or similar short period – we see a gradual increase in the distances over time (Fig. 5a).

As such, rather than using a crèche start date, which is difficult to pinpoint, as a parameter, we require an alternative metric if the timing of an element of the post-guard phase is to be compared between colonies and across years.

We suggest 'chick to adult ratio' as a candidate for this metric for a number of reasons. Firstly, as shown in Fig. 2 in Appendix 6c, the same trend is identified for this measurement for each site and season in our study, meaning it is robust to variations in colony size, camera angle and geographic location. Secondly, use of this metric makes sense in a biological context. When the ratio is highest, the number of adults per chick is at a minimum, corresponding to the period where many adults have gone to sea to forage and have left their chick(s) unguarded – in other words, the post-guard phase. Lastly, employment of this metric utilises the citizen science process itself, taking advantage of the fact that volunteers are more likely to observe, recognise and therefore tag chicks when they are away from the nest and no longer hidden by a parent. When the ratio is at a maximum, there may not genuinely be more chicks than in the previous images, but more will be identified as a result of a change in behaviour.

Assessing temperature as a driver of huddling behaviour

Huddling behaviour is thought to be a mechanism for thermoregulation in chicks (Black et al., 2016; Chapman et al., 2011; Lawless et al., 2001), with aggregations also known to provide chicks with protection – from both unrelated adults and predator species (Wilson, 2009). The results of this study provide some support for the hypothesis that huddling is driven by its thermoregulatory benefits, with mean daytime temperatures on the 'maximum huddle' date lower than for the surrounding two week period ($t(9) = -3.32$, $p <$

0.01, with a large Hedge's g effect size of -0.85). In eight out of ten cases, the average mean temperature was lower on the 'maximum huddle' date than for the surrounding two week period; in six of these cases, the 95% confidence intervals are non-overlapping (see Fig. 6). However, the small sample size means statistically robust conclusions cannot be drawn, and – as mentioned above – multi-year data from more colonies are required to assess the relative contributions of different variables in promoting huddling behaviour, with key drivers likely changing across time and space, even within a single species. It should be noted that the lower minimum and higher maximum temperatures observed during the two week period surrounding the 'maximum huddle' date are presumably owing to the greater sample size – and variance – of the two week periods' datasets.

It could be expected that a greater disparity between the temperature on the 'maximum huddle' date and the surrounding period would lead to stronger aggregation behaviour. However, plotting the citizen science 'Huddle Score' (a subjective assessment of the degree of huddling – see Appendix 6a) against the temperature difference between the 'maximum huddle' date and the surrounding two week period, showed no obvious relationship (see Fig. 7). Similarly, no obvious linear relationship was observed between the Huddle Scores and the mean temperature on the 'maximum huddle' date (Fig. 7). The removal of one data point (ORNEa2014: mean temperature = -1.44 °C; Huddle Score = 3), increases the R^2 value to 0.46, and suggests the possibility of a negative relationship between temperature and Huddle Score, should more data be collected. However, there is no justification for removing this data point, and no additional statistical tests were carried out owing to the small sample size (n = 10).

Conclusions

Remote time-lapse cameras offer an effective way of monitoring penguin phenology across multiple sites and years, with minimal requirements for on-ground observations. In turn, citizen science provides a way of efficiently processing large quantities of imagery, facilitating 'big data' studies in the field of ecology. *Penguin Watch* citizen science classifications deliver a starting point for determining arrival and departure dates of penguin colonies, though expert manual fine-tuning is currently still required. Moreover, they provide a novel method of estimating chick fledge date and peak crèche – two parameters which can be challenging to pinpoint, even when observing time-lapse imagery – or even a colony – directly. By looking at nest to adult distances, we suggest that crèche is a gradual rather than binary process, which is not necessarily defined by easy-to-label behaviours such as large chick aggregations. This, in addition to determination of peak crèche using maximum chick to adult ratios and investigation of the relationship between temperature and huddling behaviour, demonstrates how 'big data' studies can offer insights into ecological phenomena, in a way which does not necessarily align with, but complements, classical hypothesis-driven methods.

Acknowledgements

The authors would like to express their gratitude to the Zooniverse team[1] and partners[2], and the *Penguin Watch* volunteer base. We also thank J. Arthur, Z. Ročkaiová (nee Macháčková) and H. Torsey for moderating the *Penguin Watch* online 'Talk' forum. Zooniverse and *Penguin Watch* were supported by grants from the Alfred P. Sloan

Foundation and a Google Global Impact Award. FMJ and IJM are supported by NERC DTP studentships. This work was also funded by the Darwin Initiative, the Citizen Science Alliance and the John Ellerman Foundation, as well as by individual contributions to *Penguin Watch* (particularly aboard Quark Expeditions' ships). Finally, we wish to thank Quark Expeditions, Cheeseman's Ecology Safaris and Oceanwide Expeditions for fieldwork support.

1. https://www.zooniverse.org/about/team
2. https://www.zooniverse.org/projects/penguintom79/penguin-watch/about/team

Author Contributions

FMJ and TH conceived the paper; FMJ wrote the manuscript. FMJ extracted and processed *Penguin Watch* data, provided 'gold standard' dates and performed reliability analyses. CK developed the Panoptes Aggregation tool. VH and FMJ devised methods to extract arrival and departure dates. FMJ and IJM investigated proxies for crèche, and IJM wrote the code to extract crèche metrics. CJL leads the Zooniverse and helped to design *Penguin Watch.* TH founded the *Penguin Watch* project. All authors reviewed and approved the manuscript.

Competing Interests

The authors declare no competing financial interests.

References

Anderson, E. M., & Boutin, S. (2002). Edge effects on survival and behaviour of juvenile red squirrels (*Tamiasciurus hudsonicus*). *Canadian Journal of Zoology*, *80*(6), 1038–1046.

Bagshawe, T. W. (1938). Notes on the habits of the gentoo and ringed or Antarctic penguins. *Transactions of the Zoological Society of London*, *24*(3), 185–307.

Black, C., Collen, B., Johnston, D., & Hart, T. (2016). Why Huddle? Ecological Drivers of Chick Aggregations in Gentoo Penguins, *Pygoscelis papua*, across Latitudes. *Plos One*, *11*(2), e0145676. https://doi.org/10.1371/journal.pone.0145676

Black, C., Collen, B., Lunn, D., Filby, D., Winnard, S., & Hart, T. (2018). Time-lapse cameras reveal latitude and season influence breeding phenology durations in penguins. *Ecology and Evolution*, *8*(16), 8286–8296. https://doi.org/10.1002/ece3.4160

Black, C. E. (2016). A comprehensive review of the phenology of *Pygoscelis* penguins. *Polar Biology*, *39*(3), 405–432. https://doi.org/10.1007/s00300-015-1807-8

Bost, C. A., & Jouventin, P. (1991). The breeding performance of the Gentoo Penguin *Pygoscelis papua* at the northern edge of its range. *Ibis*, *133*(1), 14–25.

Both, C., Van Asch, M., Bijlsma, R. G., van den Burg, A. B., & Visser, M. E. (2009). Climate change and unequal phenological changes across four trophic levels: constraints or adaptations? *Journal of Animal Ecology*, *78*(1), 73–83. https://doi.org/10.1111/j.1365-2656.2008.01458.x

Campbell, K. J., Steinfurth, A., Underhill, L. G., Coetzee, J. C., Dyer, B. M., Ludynia, K., … Sherley, R. B. (2019). Local forage fish abundance influences foraging effort and offspring condition in an endangered marine predator. *Journal of Applied Ecology*, *56*(7), 1751–1760. https://doi.org/10.1111/1365-2664.13409

Carlini, A. R., Coria, N. R., Santos, M. M., & Buján, S. M. (2005). The effect of chinstrap penguins on the breeding performance of Adèlie penguins. *Folia Zoologica*, *54*(1–2), 147–158.

Chapman, E. W., Hofmann, E. E., Patterson, D. L., Ribic, C. A., & Fraser, W. R. (2011).

Marine and terrestrial factors affecting Adélie penguin *Pygoscelis adeliae* chick growth and recruitment off the western Antarctic Peninsula. *Marine Ecology Progress Series*, *436*, 273–289. https://doi.org/10.3354/meps09242

Charmantier, A., McCleery, R. H., Cole, L. R., Perrins, C., Kruuk, L. E. B., & Sheldon, B. C. (2008). Adaptive Phenotypic Plasticity in Response to Climate Change in a Wild Bird Population. *Science*, *320*(5877), 800–803. https://doi.org/10.1126/science.1157174

Coulson, T., Catchpole, E. A., Albon, S. D., Morgan, B. J. T., Pemberton, J. M., Clutton-Brock, T. H., … Grenfell, B. T. (2001). Age, Sex, Density, Winter Weather, and Population Crashes in Soay sheep. *Science*, *292*(5521), 1528–1531. https://doi.org/10.1126/science.292.5521.1528

Cox, J. A., Cusick, J. A., & DuVal, E. H. (2019). Manipulated sex ratios alter group structure and cooperation in the brown-headed nuthatch. *Behavioral Ecology*, *30*(4), 883–893. https://doi.org/10.1093/beheco/arz030

Cresswell, W., & McCleery, R. (2003). How great tits maintain synchronization of their hatch date with food supply in response to long-term variability in temperature. *Journal of Animal Ecology*, *72*(2), 356–366. https://doi.org/10.1046/j.1365-2656.2003.00701.x

Crick, H. Q. P., & Sparks, T. H. (1999). Climate change related to egg-laying trends. *Nature*, *399*(6735), 423–424. https://doi.org/10.1038/20839

Croxall, J. P., Rothery, P., Pickering, S. P. C., & Prince, P. A. (1990). Reproductive Performance, Recruitment and Survival of Wandering Albatrosses *Diomedea exulans* at Bird Island, South Georgia. *Journal of Animal Ecology*, *59*(2), 775–796.

Dawkins, M. S., Cain, R., & Roberts, S. J. (2012). Optical flow, flock behaviour and chicken welfare. *Animal Behaviour*, *84*(1), 219–223. https://doi.org/10.1016/j.anbehav.2012.04.036

Dawkins, M. S., Lee, H., Waitt, C. D., & Roberts, S. J. (2009). Optical flow patterns in broiler chicken flocks as automated measures of behaviour and gait. *Applied Animal Behaviour Science*, *119*(3–4), 203–209. https://doi.org/10.1016/j.applanim.2009.04.009

Edwards, M., & Richardson, A. J. (2004). Impact of climate change on marine pelagic phenology and trophic mismatch. *Nature, 430*(7002), 881–884.

Galton, F. (1907). Vox Populi (The Wisdom of Crowds). *Nature*, *75*(7), 450–451. https://doi.org/https://dx.doi.org/10.1002/clc.22367

Jones, F. M., Allen, C., Arteta, C., Arthur, J., Black, C., Emmerson, L. M., … Hart, T. (2018). Time-lapse imagery and volunteer classifications from the Zooniverse Penguin Watch project. *Scientific Data*, *5*(1), 180124. https://doi.org/10.1038/sdata.2018.124

Jones, F. M., Arteta, C., Zisserman, A., Lempitsky, V., Lintott, C. J., & Hart, T. (in review). Processing citizen science- and machine- annotated time-lapse imagery for biologically meaningful metrics. *Scientific Data*.

Kolbe, J. J., & Janzen, F. J. (2002). Spatial and temporal dynamics of turtle nest predation: edge effects. *Oikos*, *99*(3), 538–544. https://doi.org/10.1034/j.1600-0706.2002.11853.x

Körner, C., & Basler, D. (2010). Phenology Under Global Warming. *Science*, *327*(5972), 1461–1462. https://doi.org/10.1126/science.1186473

Lawless, R. M., Buttemer, W. A., Astheimer, L. B., & Kerry, K. R. (2001). The influence of thermoregulatory demand on contact crèching behaviour in Adélie Penguin chicks. *Journal of Thermal Biology*, *26*(6), 555–562. https://doi.org/10.1016/S0306-4565(00)00049-8

Le Bohec, C., Gauthier-Clerc, M., & Le Maho, Y. (2005). The adaptive significance of crèches in the king penguin. *Animal Behaviour*, *70*(3), 527–538. https://doi.org/10.1016/j.anbehav.2004.11.012

Lishman, G. S. (1985). The comparative breeding biology of Adelie and Chinstrap Penguins *Pygoscelis adeliae* and *P. antarctica* at Signy Island, South Orkney Islands. *Ibis*, *127*(1), 84–99. https://doi.org/10.1111/j.1474-919X.1985.tb05039.x

Lynch, H. J., Fagan, W. F., Naveen, R., Trivelpiece, S. G., & Trivelpiece, W. Z. (2012). Differential advancement of breeding phenology in response to climate may alter

staggered breeding among sympatric pygoscelid penguins. *Marine Ecology Progress Series*, *454*, 135–145. https://doi.org/10.3354/meps09252

Lynch, H. J., Naveen, R., Trathan, P. N., & Fagan, W. F. (2012). Spatially integrated assessment reveals widespread changes in penguin populations on the Antarctic Peninsula. *Ecology*, *93*(6), 1367–1377. https://doi.org/10.1890/11-1588.1

Moreno, J., de León, A., Fargallo, J. A., & Moreno, E. (1998). Breeding time, health and immune response in the chinstrap penguin *Pygoscelis antarctica*. *Oecologia*, *115*(3), 312–319. https://doi.org/10.1007/s004420050522

Moreno, J., Barbosa, A., Potti, J., & Merino, S. (1997). The Effects of Hatching Date and Parental Quality on Chick Growth and Creching Age in the Chinstrap Penguin (*Pygoscelis antarctica*): A Field Experiment. *The Auk*, *114*(1), 47–54.

Neff, B. D. (2003). Decisions about parental care in response to perceived paternity. *Nature*, *422*(6933), 716–719. https://doi.org/10.1038/nature01528

Newbery, K. B., & Southwell, C. (2009). An automated camera system for remote monitoring in polar environments. *Cold Regions Science and Technology*, *55*(1), 47–51. https://doi.org/10.1016/j.coldregions.2008.06.001

O'Brien, J. (2019). *exiftoolr: ExifTool Functionality from R.* Retrieved from https://cran.r-project.org/package=exiftoolr

Penteriani, V., Viñuela, J., Belliure, J., Bustamante, J., & Ferrer, M. (2003). Causal and functional correlates of brood amalgamation in the chinstrap penguin *Pygoscelis antarctica*: parental decision and adult aggressiveness. *Polar Biology*, *26*(8), 538–544. https://doi.org/10.1007/s00300-003-0517-9

Phillimore, A., B., Hadfield, Jarrod, D., Jones, O. R., & Smithers, R. J. (2010). Differences in spawning date between populations of common frog reveal local adaptation. *Proceedings of the National Academy of Sciences*, *107*(18), 8292–8297. https://doi.org/10.1073/pnas.1203538109

Post, E., & Forchhammer, M. C. (2008). Climate change reduces reproductive success of an Arctic herbivore through trophic mismatch. *Philosophical Transactions of the*

Royal Society B: Biological Sciences, *363*(1501), 2369–2375. https://doi.org/10.1098/rstb.2007.2207

Python Software Foundation. (2019). *Python Language Reference*. Retrieved from http://www.python.org

R Core Team. (2019). *R: A language and environment for statistical computing.* Retrieved from https://www.r-project.org/

Rebstock, G. A., & Dee Boersma, P. (2011). Parental Behavior Controls Incubation Period and Asynchrony of Hatching in Magellanic Penguins. *The Condor*, *113*(2), 316–325. https://doi.org/10.1525/cond.2011.100162

Rootes, D. M. (1988). The status of birds at Signy Island, South Orkney Islands. *British Antarctic Survey Bulletin*, *80*, 87–119.

Shen, T.-W., Liu, T.-J., Šuta, D., & Lee, C.-C. (2019). Acoustic Perturbation of Breathing: A Newly Discovered Response to Soft Sounds in Rats Using an Approach of Image Analysis. *Journal of Medical and Biological Engineering*, *39*(1), 43–53. https://doi.org/10.1007/s40846-018-0381-2

Southwell, C., & Emmerson, L. (2015). Remotely-operating camera network expands Antarctic seabird observations of key breeding parameters for ecosystem monitoring and management. *Journal for Nature Conservation*, *23*, 1–8. https://doi.org/10.1016/j.jnc.2014.11.002

Steen, K. A., Therkildsen, O. R., Green, O., & Karstoft, H. (2013). Audio-visual recognition of goose flocking behaviour. *International Journal of Pattern Recognition and Artificial Intelligence*, *27*(7), 1350020.

Thomas, D. W., Blondel, J., Perret, P., Lambrechts, M. M., & Speakman, J. R. (2001). Energetic and Fitness Costs of Mismatching Resource Supply and Demand in Seasonally Breeding Birds. *Science*, *291*(5513), 2598–2600. https://doi.org/10.1126/science.1057487

Trivelpiece, W., & Volkman, N. J. (1979). Nest-site competition between Adelie and chinstrap penguins: an ecological interpretation. *The Auk*, *96*(4), 675–681.

Trivelpiece, W. Z., Trivelpiece, S. G., & Volkman, N. J. (1984). Further Insights into Nest-site Competition between Adelie and Chinstrap Penguins. *The Auk*, *101*(4), 882–884.

Van Noordwijk, A. J., McCleery, R. H., & Perrins, C. M. (1995). Selection for the Timing of Great Tit Breeding in Relation to Caterpillar Growth and Temperature. *Journal of Animal Ecology*, *64*(4), 451–458.

Viñuela, J., Moreno, J., Carrascal, L. M., Sanz, J. J., Amat, J. A., Ferrer, M., … Cuervo, J. J. (1996). The effect of hatching date on parental care, chick growth, and chick mortality in the chinstrap penguin *Pygoscelis antarctica*. *Journal of Zoology*, *240*(1), 51–58. https://doi.org/10.1111/j.1469-7998.1996.tb05485.x

Weetman, D., Atkinson, D., & Chubb, J. C. (1999). Water temperature influences the shoaling decisions of guppies, *Poecilia reticulata*, under predation threat. *Animal Behaviour*, *58*(4), 735–741.

Williams, T. D., & Croxall, J. P. (1991). Chick growth and survival in gentoo penguins (*Pygoscelis papua*): effect of hatching asynchrony and variation in food supply. *Polar Biology*, *11*(3), 197–202. https://doi.org/10.1007/BF00240208

Wilson, D. (2009). Causes and Benefits of Chick Aggregations in Penguins. *The Auk*, *126*(3), 688–693. https://doi.org/10.1525/auk.2009.9709

Young, E. (1994). *Skua and penguin: predator and prey*. Cambridge, UK: Cambridge University Press.

Youngflesh, C., Jenouvrier, S., Li, Y., Ji, R., Ainley, D. G., Ballard, G., … Lynch, H. J. (2017). Circumpolar analysis of the Adélie Penguin reveals the importance of environmental variability in phenological mismatch. *Ecology*, *98*(4), 940–951. https://doi.org/10.1002/ecy.1749

Zhou, X., Harrington, R., Woiwod, I. P., Perry, J. N., Bale, J. S., & Clark, S. J. (1995). Effects of temperature on aphid phenology. *Global Change Biology*, *1*(4), 303–313. https://doi.org/10.1111/j.1365-2486.1995.tb00029.x

Chapter 5: Large-scale assessment of intra- and inter-annual nesting success using a remote camera network[6]

Authors

Casey Youngflesh[1*†], Fiona M. Jones[2*], Heather J. Lynch[4], Joan Arthur[3], Zuzana Macháčková[3], Holly R. Torsey[3] & Tom Hart[2]

Affiliations

1. Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, 90095, USA
2. Department of Zoology, University of Oxford, 11a Mansfield Road, Oxford, OX1 3SZ, UK
3. Zooniverse, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH, UK
4. Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794, USA

*Authors contributed equally

†Corresponding author: C. Youngflesh (casey.youngflesh@uconn.edu)

Keywords: capture-recapture model, time-lapse cameras, global change, seabird

[6] A version of this chapter is published as Youngflesh, C., Jones, F.M., Lynch, H.J., Arthur, J., Macháčková, Z., Torsey, H.R. & Hart, T. (2021). Large-scale assessment of intra- and inter-annual breeding success using a remote camera network. *Remote Sensing in Ecology and Conservation*, *7*(1), 97 – 108. Copyright with Authors.

Abstract

Changes in the physical environment along the Antarctic Peninsula have been amongst the most rapid anywhere on the planet. In concert with environmental change, the potential for direct human disturbance resulting from tourism, scientific programs, and commercial fisheries continues to rise in the region. While seabirds, such as the Gentoo penguin (*Pygoscelis papua*), are commonly used to assess the impact of these disturbances on natural systems, research efforts are often hampered by limited spatial coverage and lack of temporal resolution. Using a large-scale remote time-lapse camera network and a modelling framework adapted from capture-recapture studies, we assess drivers of intra- and inter-annual dynamics in Gentoo penguin breeding success across nearly the entire species' range in the Atlantic sector of the Southern Ocean. We quantify the precise timing of egg/chick mortality within each season and examine the role of precipitation events, tourism visitation, and fishing activity for Antarctic krill (*Euphausia superba*, a principal prey resource in the Antarctic) in these processes. We find that nest failure rates are higher in the egg than the chick stage and that neither krill fishing nor tourism visitation had a strong effect on Gentoo penguin breeding success. While precipitation events had, on average, little effect on nest mortality, results suggest that extreme weather events can precipitate sharp increases in nest failure. This study highlights the importance of continuous ecosystem monitoring, facilitated here by remote time-lapse cameras, in understanding ecological responses to environmental stressors, particularly with regard to the timing of events such as extreme weather which may facilitate non-linear responses.

Introduction

The Antarctic Peninsula (AP) region is currently experiencing rapid change resulting from both direct and indirect anthropogenic impacts. Air temperatures have increased significantly over the recent past (Mayewski et al., 2009), coinciding with changes in sea ice dynamics (Stammerjohn et al., 2012) and precipitation (Kirchgäßner, 2011; Turner et al., 2005). Tourism visitation (Bender et al., 2016) and commercial fishing efforts (Nicol et al., 2012) also continue to increase in the region.

Given the complexity of natural systems, assessing the repercussions of these various stressors on Antarctic ecosystem dynamics is a difficult task. Seabirds, such as the Gentoo penguin *Pygoscelis papua,* are commonly used to assess the impacts of environmental and anthropogenic stressors, as these birds are easily monitored and represent key predators in marine ecosystems. While Gentoo penguins are often considered 'climate change winners' in the Antarctic Peninsula region (Clucas et al., 2014; Hinke et al., 2007), it is uncertain how intensifying anthropogenic stressors and abiotic changes are influencing populations. Other penguin species, such as the Adélie penguin (*P. adeliae*), have shown marked declines in the AP region in recent years (Lynch et al., 2008; 2012). Disentangling how demographic processes in the *Pygoscelis* spp. penguins are responding to various threats driving population changes is key to understanding how the Antarctic system is changing over time and forecasting how these species might fare into the future.

The increasing frequency of precipitation events observed in the AP region (Kirchgäßner, 2011; Turner et al., 2005) is a key concern for penguins on the western Antarctic Peninsula, particularly with regard to breeding success and chick survival. Nest flooding, resulting from heavy rainfall or accumulated snowmelt (Youngflesh, 2018*b*), may lead to substantial

heat loss in chicks and provides one mechanism for the negative relationship observed between precipitation and breeding success for *Pygoscelis* spp. and Magellanic (*Spheniscus magellanicus*) penguins (e.g. Boersma & Rebstock, 2014; Lynch et al., 2010; Ropert-Coudert et al., 2015). This can impact survival at both the egg and chick stage (Cimino et al., 2014; Massom et al., 2006; Thyen & Becker, 2006). Model results suggest that wetting of just 10% (total chick surface area) is sufficient to reduce survival probability through decreased fledging mass (Chapman et al., 2011; Konarzewski & Taylor, 1989; Lustick & Adams, 1977). Increased precipitation has been associated with reduced chick growth rate in a number of bird species (e.g. Dunn, 1975; Konarzewski & Taylor, 1989; Thyen & Becker, 2006), an effect owing to associated heat loss due to wetting and/or a continued decrease in provisioning rate owing to adverse weather conditions (Moss, 1979).

Direct anthropogenic threats, such as an expanding Antarctic krill *Euphausia superba* fishery, are also cause for concern (Nicol et al., 2012; Pfeiffer & Peter, 2004). Krill, an important food resource for *Pygoscelis* spp. penguins (Volkman et al., 1980), currently represent the largest (by tonnage) fishery in the Southern Ocean, with a catch limit of 620,000 tons per year (Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR), 2018). It is important to understand if the krill harvest may be impacting predator populations in the Southern Ocean, particularly in the southwest Atlantic sector (20°W - 80°W), which has historically been the primary focus of the industry (Kawaguchi et al., 2009). An increase in demand for krill-derived products, combined with the potential for a spatial expansion of fishing as sea ice cover declines, further heightens

the need for a more complete understanding of the impact of this threat (Kawaguchi et al., 2009; Nicol et al., 2012; Stammerjohn et al., 2012).

In addition to precipitation and the removal of key prey, tourism-related impacts are also a concern for Gentoo penguins. The tourism industry in Antarctica has grown rapidly in the last several decades, with more than 58,000 visitors during the 2017/2018 season (Lynch et al., 2019). More than 70% of all tour ship landings during the 2012/2013 austral summer occurred at Gentoo penguin colonies (Bender et al., 2016). Disturbance of breeding individuals at colonies is the primary concern, as visitation of colonies may have physiological impacts (e.g. stress) or present opportunities for predators to prey upon eggs or young chicks. Prior studies examining the impact of site visitation on Gentoo penguins, however, have drawn inconsistent conclusions (Holmes, 2007; Lynch et al., 2019).

Effectively evaluating the impacts of these potential threats on Gentoo breeding success requires a substantial monitoring effort. Historically, assessing demographic parameters such as egg and chick survival on a large spatial scale with high temporal resolution has been both challenging (logistically and financially) and labor-intensive, resulting in a trade-off between the depth of demographic information and geographic breadth (Lynch et al., 2012). Remote time-lapse cameras, pioneered for use in penguin monitoring studies by Newbery & Southwell (2009), provide a solution to this problem, allowing colonies to be observed at high frequency across large spatial extents (Hinke et al., 2018; Jones et al., 2018). As such, they facilitate the quantification and comparison of breeding success across large regions and provide an ongoing opportunity to quantify the timing of mortality events across the breeding season. Here we use data from the large-scale *Penguin Watch* remote time-lapse camera network (Jones et al., 2018) to examine

breeding success at 14 Gentoo penguin colonies across the AP region and South Georgia (representing the bulk of the species' South Atlantic sector range), providing the most spatially comprehensive study of Gentoo demographics to date. We employ a modelling framework adapted from capture-recapture studies to explicitly account for non-detection (owing to obscuration of chicks by adults, poor visibility, or human error), and assess changes in chick survival throughout the breeding season. Our goals were to assess: 1) large-scale patterns of Gentoo penguin breeding success across a large portion of their breeding range, 2) variation in the timing of chick mortality over the course of the breeding season, and 3) the impact of three hypothesized drivers of breeding success: precipitation, tourism visitation, and commercial krill fishing.

Materials and Methods

Image data collection via the Penguin Watch remote time-lapse camera network

Images were captured by Reconyx HC500 Hyperfire Trail Cameras (commercially available; Reconyx Inc., Holmen, WI, USA), deployed according to the methodology outlined by Jones et al. (2018). These devices form part of the *Penguin Watch* remote time-lapse camera network, which currently comprises approximately 100 units positioned across the Scotia Arc region (Jones et al., 2018). Devices were programmed to capture images once per hour throughout the breeding season, which were stored on Secure Digital (SD) cards and retrieved the following austral summer.

Chick classification from remote time-lapse camera images

Penguin chick survival was assessed using images captured at 14 Gentoo penguin colonies across the AP and South Georgia between the 2012-2013 and 2017-2018 breeding seasons (approximately December – February; total sites/years = 28; full list of data availability given in Table S1-1, Appendix 7a). Cameras captured images every hour during a set daytime period (for example, 0500 to 2100), with a mean average of 16 images taken per day. The locations of individual nests were annotated on a representative image from each site/year, which were then used to demarcate 'nest zones' through Voronoi tessellation using the package 'deldir' (Turner, 2019) in the R statistical environment (R Core Team, 2019; Fig. 1a, Fig. S1-1 (Appendix 7a)). Voronoi tessellation partitions space into polygons with one 'seed' (nest location, in this case) per polygon, where all possible points inside that polygon are closer to that seed than any other seed.

These zones remained constant throughout each site/year image set, as the camera positions were fixed. The R packages 'jpeg' (Urbanek, 2014) and 'stringr' (Wickham, 2019) were used to render and manipulate images obtained from the remote time-lapse camera network and to support data processing, respectively. Images were annotated using a version of the online citizen science project *Penguin Watch* (www.penguinwatch.org) specifically designed for the purposes of this study (i.e. not publicly accessible). Using this interface, we classified the location of penguin chicks in each image (total number of images = 9831). We defined our focal period as the time between the sighting of the first chick and the crèche stage (the period during which penguin chicks begin to form large chick aggregations; Fig. 1a). All identified nests were included in the analysis, irrespective of survival outcome so as not to bias survival estimates (i.e. nests where no chicks hatched

were included). As Gentoo penguins consistently lay two eggs per clutch, each nest could have zero, one, or two chicks present in each image. Image classifications were used to create time series of the observed number of chicks in each nest zone for each site/year (Fig. 1b). These time series were used in a modified capture-recapture framework (Nichols 1992; where each observation represents a 'capture' for a particular nest) to assess how chick mortality changes throughout a season, and to what degree changes in chick mortality coincide with weather events. While traditional capture-recapture studies rely on marking individuals, the locations of the nests serve as distinct markers to reference the identity of each chick/sibling pair. A two-stage analysis – where breeding success estimates were calculated, then incorporated into three separate linear models (fit to the three explanatory variables) – was carried out to ensure data transparency. Average colony-wide breeding success is the metric commonly used in seabird studies, and this method allows these estimates, and their relationships to the three separate explanatory variables, to be compared to previous and future analyses. Uncertainty in breeding success estimates was carried forward, as discussed below.

Modelling chick survival in a modified capture-recapture framework

Given the possibility of detection failure, the observed state of a penguin chick in any given image (present/absent) cannot be assumed to be the true state (truly present/truly absent). A Bayesian state-space framework was implemented to model both the true (z) and observed (y) states of penguin chicks at each time step (t), for each nest (i), year (j), and site (k). Hourly data were aggregated to daily resolution (i.e. the highest number of chicks seen in a given nest over the course of a day) for computational reasons. The

Bayesian approach provides a flexible modelling framework and allows for the propagation and quantification of uncertainty throughout the analysis (Gelman & Hill, 2006). Crèche dates for each site/year were used to determine an approximate lay date (60 days before crèche date; Hinke et al., 2012). This extrapolated lay date was used as the initial time step of the model so that time series for each site/year were 60 time steps (days) in length. Each nest was assumed to have two eggs at the initial time step, as this is highly conserved within the *Pygoscelis* spp. penguins (Borboroglu & Boersma, 2013).

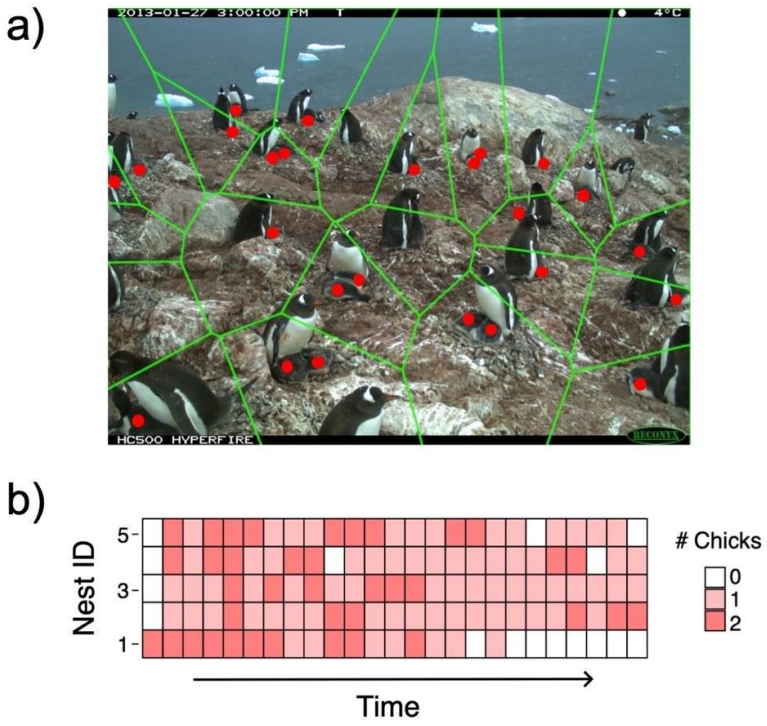


Figure 1: a) Image of a penguin nesting site obtained from the *Penguin Watch* remote time-lapse camera network. Red dots represent penguin chicks (classified by authors using the *Penguin Watch* platform on Zooniverse); green polygons represent nest zones (delineated using a Voronoi tessellation with nest locations as 'seeds'); b) sample time series of the number of chicks observed in each of five hypothetical nests. At each time step, zero, one, or two chicks could be observed.

The true state (truly present/truly absent) of each chick in a given nest depends on its state in the previous time step. This is a binomial process, with daily survival probability ϕ, which was modelled as a function of a global intercept term (μ_ϕ) that varies by site and year. As the two chicks in a given nest cannot be uniquely distinguished in an image, the survival of both chicks was modelled jointly with a binomial distribution (as opposed to the Bernoulli models typically used in capture-recapture studies). This is given as,

$$z_{t,i,j,k} \sim Binom(z_{t-1,i,j,k}, \phi_{j,k}) \quad (1)$$
$$logit(\phi_{j,k}) = \mu_{\phi_{j,k}}$$

where *t* is the time step for nest *i* in year *j* at site *k*. $\mu_{\phi_{j,k}}$ represents the (logit-transformed) survival of each egg in year *j* at site *k*. A logit transform was used to more easily set priors for this parameter. We assumed that site- and year-specific survival ϕ (logit transformed) follows a Normal distribution, which allowed random variation in the site- and year-specific survival parameters under the constraint of a common distribution,

$$\mu_{\phi_{j,k}} \sim N(\theta_\phi, \sigma_{\mu_\phi}) \quad (2)$$

where θ_ϕ and σ_{μ_ϕ} represent the grand mean and standard deviation for this parameter, respectively, with larger variation among sites and years reflected in larger estimates for σ_{μ_ϕ}.

Unavoidably, camera trap images provide information only on what is visible. The contents of a nest may be obscured by the parent or artifacts such as rain on the camera lens (Southwell & Emmerson, 2015). It is therefore necessary to model non-detection of nest contents, which provides a bridge between what actually exists in each nest and what can be observed. The observed state of each nest (*y*) was similarly modelled as a binomial

process, with detection (*p*), that depends on the true state (*z*) of that nest in that time step,

$$y_{t,i,j,k} \sim Binom(z_{t,i,j,k}, p_{t,i,j,k}). \quad (3)$$

Since chicks were the quantity of interest in this study and egg sightings were not quantified, data were only included for *y* over the period of interest for each site/year, from the first sighting of a chick to the start of the crèche period (that is, NA values were used prior to the sighting of the first chick for a given site/year). The logit of the detection probability for time step t, nest i, year j, at site k ($p_{t,i,j,k}$) was modelled as a function of a global intercept (μ_p), a time-varying parameter (β_p) that varied by year and site to account for the change in detection probability of a penguin chick as it grows larger over time (*x*), and an offset for each nest, year, and site (ν_p), as detection is likely to differ across nests, sites, and years depending on the position of the camera and nest,

$$logit(p_{t,i,j,k}) = \mu_p + \nu_{p_{i,j,k}} + \beta_{p_{j,k}} x_t \, . \quad (4)$$

In this way, detection probability can vary across time, nest, site, and year. Note that while β_p represents a linear function of time in logit space, this corresponds to a logistic function in probability space (as illustrated in Fig. S1-5, Appendix 7a). μ_p and ν_p impact where the inflection point of the detection curve is located, while β_p impacts the shape of the detection curve. Both β_p and ν_p were modelled hierarchically, with means μ_{β_p} and 0, and standard deviations σ_{β_p} and σ_{ν_p}, respectively,

$$\beta_{p_{j,k}} \sim N(\mu_{\beta_p}, \sigma_{\beta_p}) \quad (5)$$

$$\nu_{p_{i,j,k}} \sim N(0, \sigma_{\nu_p}) \, .$$

The total number of chicks at each site k in every year j at every time step t ($Z_{t,j,k}$) was calculated as a derived quantity, by summing across the number of chicks alive ($z_{t,i,j,k}$) at each nest i,

$$Z_{t,j,k} = \sum_i z_{t,i,j,k} \, . \tag{6}$$

Uniform priors were used for the standard deviations of the estimated parameters, and Normal priors for the means. We used a Normal (μ = 4, σ = 0.25) prior for θ_ϕ, a Uniform (μ = 0, σ = 3) prior for σ_{μ_ϕ}, a Truncated Normal (μ = 0.1, σ = 10) T(0, 1) prior for μ_{β_p} (set between 0 and 1 as these are the only reasonable values for this parameter), a Uniform (μ = 0, σ = 2) prior for σ_{β_p}, and a Uniform (μ = 0, σ = 3) prior for σ_{ν_p}. The priors are weakly informative, given what we would expect given our knowledge of the biology of the system. Examining the overlap between prior and posterior distributions indicated that the priors did not have a substantial effect on the parameter estimates.

Breeding success (chicks per pair) for each site/year was calculated as a derived quantity, by dividing $Z_{60,j,k}$ (the total number of chicks alive at the start of the crèche period) by the total number of nests at each iteration of the posterior chain. Models were fit using the R package 'rjags' (Plummer, 2016) to interface with JAGS (Plummer, 2003) in the R statistical environment (R Core Team, 2019). Inferences were obtained from 400,000 samples drawn from six chains with a thinning rate of 100, following a 'burn-in' period of 400,000 draws and an adaptation phase of 8000 draws. Convergence of the model was assessed via a visual analysis of the posterior chains, in addition to the use of the Gelman-Rubin convergence diagnostic (Brooks & Gelman, 1998). Models unambiguously converged (Rhat < 1.1) and effective samples sizes were sufficiently large (ess / chain > 100). Posterior

predictive checks (Gelman et al., 2013), whereby data generated by the model is compared to the observed data, did not indicate any model misfit (Appendix 7a). The R packages 'MCMCvis' (Youngflesh, 2018*a*), 'dplyr' (Wickham et al., 2018), and 'ggplot2' (Wickham, 2016) were used to manipulate data and model output, and to create all plots.

Quantifying changes in chick mortality over the breeding season

The estimated number of chicks at each time step ($Z_{t,j,k}$) was used to determine how rates of mortality varied across the breeding season. The modelled breeding season was divided into three segments. The first 30 days were classified as the 'egg period', the next 15 days were classified as the 'young chick period', and the final 15 days were classified as the 'old chick period'. While these periods were chosen somewhat arbitrarily, they represent early, middle, and late stages of the egg/chick portion of the breeding season. To calculate the mean daily mortality over each of these periods, for each site/year the difference between the estimated number of chicks at the beginning and end of each period was divided by the total number of days for that period (i.e. the change in number of chicks per day). Mean average daily mortality for each of these periods was aggregated across sites and used to create density plots (Fig. 3). Differences in mortality rates across these periods were assessed using a Friedman test, used to test for differences in means across repeated measures (measures represented by mortality rates for each site/year) without the assumption of normality (Friedman, 1937). The 95% credible intervals for the posterior means (daily mean mortality rate) were also calculated, to further assess whether mortality rates varied significantly between each period.

The effect of precipitation, fishing, and tourism on Gentoo penguin breeding success

Each time-lapse image was scored according to a precipitation scale by two authors independently, with any discrepancies resolved via discussion. The magnitude of rain events (observed as pooling on the ground) was scored on a scale from R1 (light rain) to R3 (heavy rain). Snow events were similarly classified, with S1 representing light snowfall with patchy coverage between nests, and S3 representing heavy snowfall (Fig. S1-4, Appendix 7a). One score was recorded per precipitation event – i.e. if snow remained for several days before melting, only the initial, causative snowfall event was recorded (although any additional precipitation during this period was scored). The timing of precipitation events was visually compared to demographic dynamics within each season. The total number of precipitation events greater than or equal to magnitude 2 (i.e. R2, R3, S2, S3) was calculated for each site to investigate the effect of the total seasonal precipitation on breeding success across sites/years.

Monthly krill fishery catch data aggregated to CCAMLR small-scale management units (SSMU; areas designed to aid in the management of the Antarctic krill fishery; Appendix 7b) were obtained from the CCAMLR statistical bulletin (CCAMLR, 2019). For a given month, krill catch within a 150km radius of each site was determined by calculating a weighted mean average of SSMU catch (where weights were determined by the proportion of total buffer area covered by a given SSMU; Appendix 7b). The sum of the weighted average krill catch from March (of year $t-1$) through January (year t) was calculated for each site/year (where t is the year in which a given breeding season ends). This metric represents all krill caught at each site prior to and during the breeding season. As Gentoo penguins generally overwinter near breeding sites (Black et al., 2017; Hinke et

al., 2007,), this period would account for any carry-over effects that fisheries may have on breeding success the following season (e.g. reduced adult body condition resulting from increased fishery activity which may lead to lower breeding success). A 150km buffer was chosen based on knowledge of Gentoo penguin winter foraging distances. For example, Tanton et al. (2004) reported that 95.4% of locations transmitted by marked birds (from the colony at Square Pond, South Georgia) between 23rd April and 14th August 2001 were within 30km of the colony, with a maximum distance travelled of 180km. The 150km buffer zone was used to account for the overwinter movement of penguins around the breeding site as well as the spatial advection of krill during the six-month period. The R packages 'rgdal' (Bivand et al., 2019), 'rgeos' (Bivand & Rundel, 2019), 'raster' (Hijmans, 2019), and 'sp' (Bivand et al., 2008) were used to process all spatial data.

The number of tourist visits at each AP and South Georgia site in each year was determined from visitation data provided by the International Association of Antarctica Tour Operators (IAATO) and the Government of South Georgia and the South Sandwich Islands (GSGSSI), respectively. Visitation only over the period from the extrapolated lay date to crèche date was used, as we hypothesize this would have the most direct impact on nesting penguins.

Breeding success at each site/year was modelled as a linear function of the total number of precipitation events, log(total krill catch), and total number of visitors. Krill catch was log transformed, as studies suggest the relationship between seabird breeding success and prey abundance is non-linear, with success decreasing sharply below a certain threshold level of food availability, but plateauing once this threshold level has been surpassed (Cairns, 1988; Cury et al., 2011; Piatt et al., 2007). This robustness to a moderate decrease in prey availability can be explained by behavioural mechanisms, such as prey-switching

(Field et al., 2010), and appears to be applicable to seabirds with a range life-history strategies and population sizes (Cury et al., 2011).

Three separate models were fit to evaluate the effects of the hypothesized drivers on breeding success while accounting for uncertainty in estimates of breeding success. Breeding success was modelled as a linear function of the proposed driver,

$$bs_{obs_i} \sim N(bs_{true_i}, \sigma_{bs_i}) \quad (7)$$

$$bs_{true_i} \sim N(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta x_i$$

where bs_{obs} is the estimated breeding success (posterior mean of the derived quantity from the capture-recapture model) for each site/year, σ_{bs} is the uncertainty about the estimate of breeding success for each site/year (posterior standard deviation of the derived quantity; this is given rather than estimated by the model), bs_{true} is the latent true state of breeding success, *x* is the covariate of interest for site/year *i*, α is the intercept parameter, β is the slope parameter, and σ is the residual error term. Models were fit using the R package 'rjags' (Plummer, 2016) to interface with JAGS (Plummer, 2003) in the R statistical environment (R Core Team, 2019). Inferences were obtained from 10,000 samples drawn from four chains, following a 'burn-in' period of 10,000 draws and an adaptation phase of 5000 draws. Models unambiguously converged (Rhat < 1.1) and effective samples sizes were sufficiently large (ess / chain > 100). Code to reproduce all analyses can be found on GitHub (https://github.com/Penguin-Watch/penguin_watch_model).

Results

Mean breeding success varied across the spatial range of the study (Fig. 2; Appendix 7a). Breeding success ranged from 0.56 (posterior sd < 0.01) to 1.90 (posterior sd < 0.01) chicks per pair and averaged 1.46 (sd = 0.28) chicks per pair across all sites/years (Table S1-1, Appendix 7a).

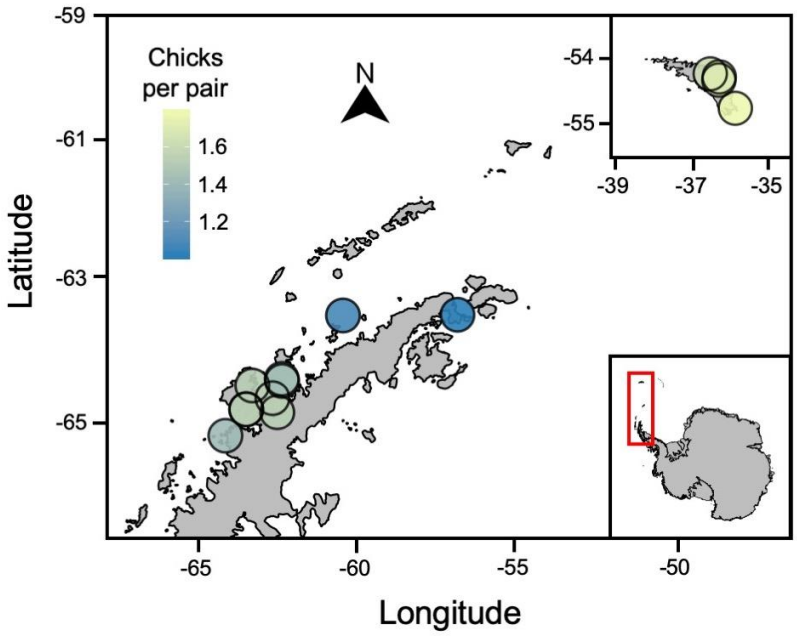


Figure 2: Average breeding success (number of chicks at crèche per nest averaged over all available years) at Gentoo penguin colonies on the AP and South Georgia (inset). Yellow hues represent higher average breeding success, while bluer hues represent lower average breeding success for that site.

Mean daily mortality rates and 95% credible intervals for each period are given in Table S1-2, Appendix 7a. Differences in daily mean mortality rates between each period, and the 95% credible intervals for these differences, are shown in Table S1-3, Appendix 7a. These

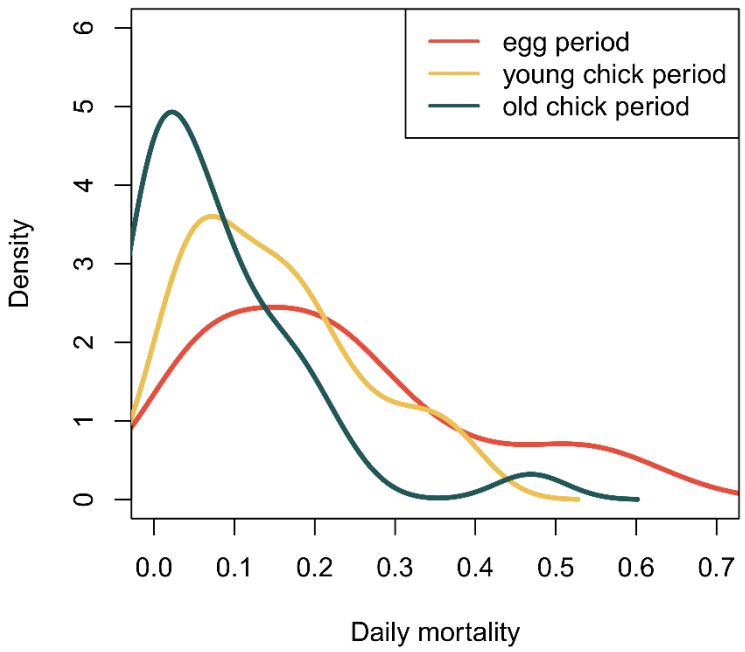


Figure 3: Density plots representing mortality in the 'egg' (red line), 'young chick' (yellow line), and 'old chick' (blue line) stages for each site/year. Mean daily mortality is higher during the earlier portion of the breeding season.

data show that mean daily mortality was generally higher during the 'egg period' than during either the 'young chick' or 'old chick' periods (with greater mortality in 78.6% and 85.7% of cases (i.e. site/years), respectively). Mortality rate was also higher during the 'young chick period' than during the 'old chick period' in 78.6% of cases, suggesting a general trend of decreasing mortality over the breeding season. Mean daily mortality during the 'egg period' and 'young chick period' are not credibly different, with the 95% credible intervals of the differences all bounding zero. Similarly, mean daily mortality rate is only credibly different between the 'young chick' and 'old chick' periods in four cases. However, this is somewhat expected given the uncertainty associated with z (the number of chicks) early in the breeding season, when there are fewer observations to constrain the estimate. Differences in mortality rate between the 'egg period' and 'old chick period'

are credibly different in 17 of the 28 samples (60.7%), and the Friedman test showed mortality to be higher earlier in the season ($p < 0.001$; i.e. during the egg period, rather than the chick period; Fig. 3). We therefore find evidence to suggest greater mortality during the egg period than the older chick period. However, we note that firm conclusions cannot be drawn where the differences in estimates are not credibly different, so only tentatively suggest an overall trend of decreasing mortality over the breeding season.

We found no strong effect of total number of precipitation events (Fig. 4a), no strong effect of krill catch (Fig. 4b), and no strong effect of tourism visitation (Fig. 4c) on penguin breeding success. Credible intervals overlapped zero in all three cases – median

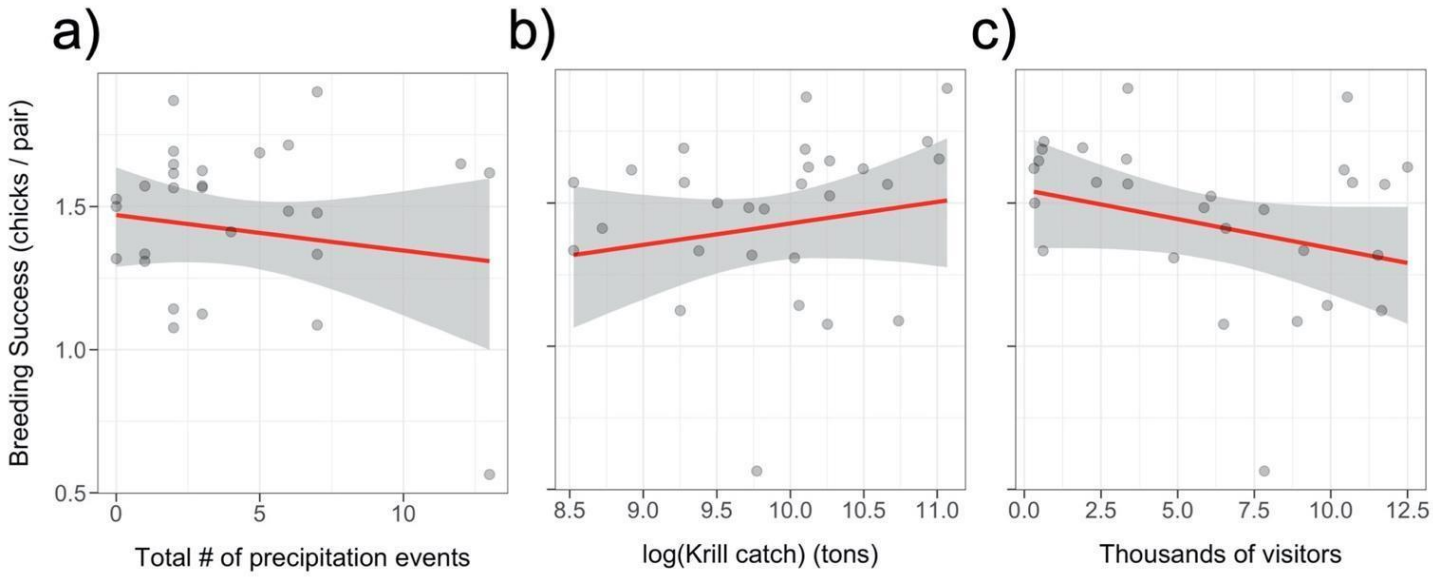


Figure 4: a) Breeding success as a function of the total number of precipitation events, b) breeding success as a function of log(krill catch), and c) breeding success as a function of the total number of tourist visitors over the course of the breeding season. Each point represents a unique site/year, the red lines represent the linear model fit, while the gray ribbons represent the 95% credible intervals.

($\beta_{precipitation}$) = -0.01 [95% CI: -0.04 - 0.02], median (β_{krill}) = 0.08 [95% CI: -0.09 - 0.24], median ($\beta_{tourism}$) = -0.02 [95% CI: -0.05 - 0.01]. In general, the timing of precipitation events were not associated with declines in chick mortality (Appendix 7a). Large snow events, however, coincided with a drop in the numbers of chicks (Fig. 5).

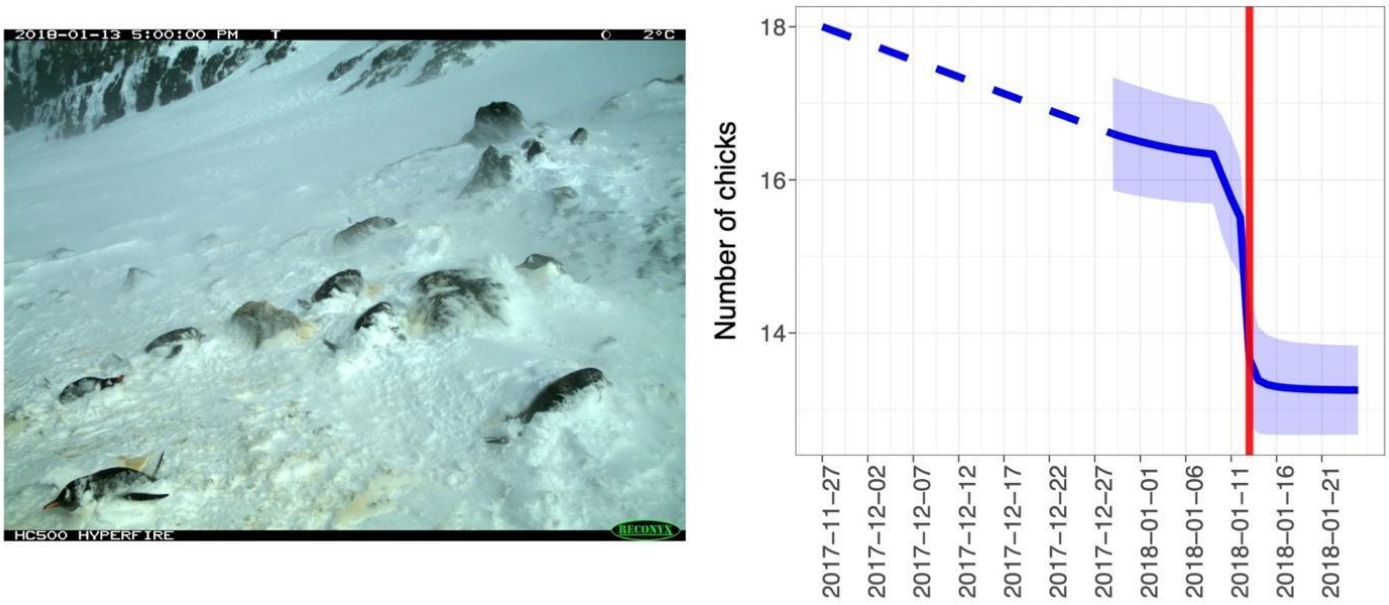


Figure 5: Left - image from the *Penguin Watch* camera network on January 13th, 2018, showing heavy snow covering penguin nests following a large storm at Brown Bluff (63.65° S, 56.89° W). Right - total number of chicks at this site over the course of the 2017-2018 breeding season. The solid blue line shows the total number of penguin chicks at each time step (Z) at this site/year. Uncertainty about that number (one standard deviation) is represented by the blue ribbon. The blue dashed lined represents the transition from the extrapolated egg lay date (60 days before crèche, where the number of chicks present is twice the total number of nests), to the date the first penguin chick was recorded in an image. The red vertical line coincides with the day the image shown at left was recorded.

Discussion

This non-invasive capture-recapture framework, facilitated by remote time-lapse cameras, provides a novel way of quantifying both the intra- and inter-annual dynamics of breeding success. This study provides an important advance in our understanding of environmental and anthropogenic impacts on Gentoo penguin breeding success and illustrates the importance of continuous monitoring in understanding ecological responses to stressors that result from global change. We show that neither moderate precipitation events, krill fishing pressure, nor tourism visitation strongly drive Gentoo penguin reproductive dynamics at these sites, despite prior concerns regarding these proposed drivers. However, discrete extreme weather events had a disproportionately large impact in this regard.

Timing of chick mortality events and the impact of precipitation

There is evidence that mortality is highest during the egg stage (i.e. the first 30 days of the breeding period), in agreement with prior findings (Hinke et al., 2012). This early mortality is, by definition, not linked to chick provisioning rate, though food availability could influence egg attendance (Eikenaar et al., 2003). While the overall number of precipitation events does not seem to have an effect on breeding success (Fig. 4a), there appears to be – from a qualitative review of the data – evidence of a non-linear relationship between these two variables, where extreme weather events may have a disproportionately large effect on breeding success (Fig. 5), though the inherent rarity of these events makes it difficult to apply formal tests in this regard. Since it is only these large precipitation events which appear to have an effect on egg/chick survival, it is likely that nest flooding or chick

wetting and subsequent hypothermia is the primary cause of mortality (Cimino et al., 2014; Lustick & Adams, 1977; Massom et al., 2006; Thyen & Becker, 2006). Rock (AP) and peat substrate (South Georgia) that make up Gentoo penguin nests allow for drainage of precipitation, which presumably prevents flooding during low to moderate precipitation events, and hence may explain why these events do not appear to depress breeding success. During heavy snow events (such as those observed in this study, where snow may fall directly into, or be blown into, nests), the likelihood of egg/chick wetting or drowning is heightened. Whether the increased drainage potential of raised tussock/peat nests used by Gentoo penguins in South Georgia is partially responsible for higher breeding success in this region compared to the AP (Fig. 2) is a subject that merits further study. An increasing frequency of precipitation events on the AP has been documented over recent years (e.g. Kirchgäßner, 2011), and is predicted to increase under climate change projections (Uotila et al., 2007), which may result in lower breeding success of Gentoo and other penguin species (e.g. Boersma & Rebstock, 2014) in the future. Gentoo penguins are known to lay second clutches if the first fails early in the season (Mauget et al., 1995). This was observed at one site during a particularly heavy snow year over the course of this study (Brown Bluff [63.65° S, 56.89° W] 2015-2016), in addition to observations made prior to this study (Yankee Harbor [62.53° S, 59.78° W] 2007-2008). It is possible that instances of re-laying will become more common as precipitation events increase in frequency, particularly if these extreme events occur early in the season.

Effect of direct anthropogenic impacts on Gentoo penguin breeding success

Fisheries closure experiments have had a positive impact on populations of certain predator species, two examples being increased chick survival and condition in African

penguins (*Spheniscus demursus*) (Sherley et al., 2018) and increased breeding productivity in Black-legged Kittiwakes (*Rissa tridactyla*) (Frederiksen et al., 2008), following closures of local sardine (*Sardinops sagax*) and anchovy (*Engraulis encrasicolus*), and sand lance (*Ammodytes marinus*), fisheries, respectively. Such empirical studies justify concerns over the impacts of forage fish fisheries on central-place foraging predators, contradicting the idea that fishing activity has negligible impact on these populations compared to environmental variability (e.g. Hilborn et al., 2017). There has been much concern about the impact of krill extraction on populations of Antarctic predators (e.g. Croxall et al., 1999). Previous reports (CCAMLR Scientific Committee Report, 2003) have highlighted the difficulty of parsing the effects of commercial harvesting and environmental change using the existing CCAMLR Ecosystem Monitoring Program (CEMP) design. Here we present a new and unique dataset, to examine penguin demographics on unprecedented spatial and temporal scales. Future studies taking into account the spatio-temporal scales on which krill and their predators (including fishing vessels) interact are necessary to determine whether catch limits are truly precautionary (Watters et al., 2020), and may be facilitated by the release of higher spatial-resolution krill catch data in the future.

Prior studies have found overlap between penguin foraging areas and krill fishing zones (Hinke et al., 2017). For the Gentoo penguin, however, krill fishing effort does not currently seem to be a major factor in driving breeding success at the sites included in this study (Fig. 4b). We also examined krill catch solely during the breeding season but found a similar lack of relationship (Appendix 7b). This could be due to the dietary flexibility of Gentoo penguins (Lescroel et al., 2004; Miller et al., 2009; Polito et al., 2015) and their near-shore, on-shelf foraging strategy (Ratcliffe et al., 2018), which is in contrast to the Chinstrap

penguin (*P. antarcticus*) and Adélie penguin, though krill availability likely impacts other Gentoo prey species (Nemoto et al., 1988). Krill fishing may have more of an impact during the overwinter period, particularly since harvesting rates are highest during this time (Appendix 7b). This may influence the survival rates of juveniles, though these dynamics are outside the scope of this study. Finally, it is also possible that periods of higher krill catch simply reflect periods of higher krill availability (suggested by the weakly positive relationship between penguin breeding success and krill catch; Fig 4b), for fishing vessels and penguins (Conn et al., 2014; Sherley et al., 2013) making it difficult to infer the true impact of krill fishing on penguin breeding success.

We also found no evidence that tourism visitation had an impact on Gentoo penguin breeding success (Fig. 4c), consistent with recent findings that tourism had little relationship to hormonal stress markers in Gentoo penguins on the AP (Lynch et al., 2019). This is in contrast to findings by Lynch et al. (2010) who documented an association between decreased Gentoo breeding success and 'frequent' site visitation at Petermann Island. Trathan et al. (2008) reported a decrease in the number of nesting individuals in highly trafficked portions of a colony at Goudier Island (Port Lockroy), positing that younger prospecting birds may preferentially choose unvisited sites. This supports the notion that penguins may habituate to human visitation over time (Holmes et al., 2006; Lynch et al., 2019). More extensive recent work at Goudier Island showed little difference in nest productivity between visited and unvisited portions of the breeding colony (Dunn et al., 2019). Our results, in concert with other recent work on tourism impacts, speak to the efficacy of regulatory measures currently in place for the Antarctic tourism industry and the importance of strict visitation guidelines from organizations such as IAATO. A

complete compilation of breeding success estimates, as well as estimates from prior studies is presented in Appendix 7a.

Conclusions

We find no evidence that direct anthropogenic stressors are substantially limiting Gentoo penguin breeding success. It may be that climatic events – i.e. indirect rather than direct human impacts – are more important determinants of these dynamics. In particular, it seems that extreme snow events are having a disproportionately large impact on chick mortality (though no formal test for non-linearity has been undertaken at this point, owing to the sample size of extreme events). This is cause for concern, since an increasing frequency of such events is predicted under climate change scenarios (Uotila et al., 2007). Future work investigating the influence of fine-scale colony topography (e.g. McDowall & Lynch, 2017) on precipitation-driven nest failure would be a valuable contribution to our understanding of these processes and may provide insights regarding how marginal nesting sites within colonies may relate to a terrestrial carrying capacity. While the impacts of the marine environment on the demographics of the *Pygoscelis* spp. penguins has been given much attention, non-marine environmental forcing also appears to play a role in the breeding productivity of this species.

This study serves as an important step in camera-based survivorship methodology and will be important in understanding how other species are responding to these and other stressors. For instance, while Gentoo penguin populations are generally stable or increasing in the Antarctic Peninsula region, other *Pygoscelis* spp. (Chinstrap and Adélie) penguins are decreasing in abundance in the region. Given the propensity of the Chinstrap

penguin and Adélie penguin to lay only a single clutch (while the Gentoo penguin may re-lay following a nest failure), extreme climatic events may have a larger demographic impact for these species. These methods may be adapted to a number of species that breed in a stationary location over the course of a breeding season.

The apparent non-linear relationship between precipitation and breeding success illustrates the importance of both intra-annual dynamics and extreme events in ecological systems and highlights the need for sustained monitoring of these systems. Seasonally averaged breeding success measures (e.g. chicks fledged per nest) do not provide enough information to understand the temporal dynamics of nest failure, and therefore are difficult to interpret with respect to identifying key drivers. Because of the substantial 'noise' present, ecologists rely on aggregation to better understand ecological dynamics and the effect of extreme events (Che-Castaldo et al., 2017). Continuous, high-resolution monitoring, here facilitated by remote time-lapse cameras, coupled with appropriate statistical modelling frameworks, such as that presented here, may provide another solution that can help tease apart factors of supposed importance and identification of the impacts of discrete events, such as extreme weather.

Acknowledgements

We thank all the staff members of Quark Expeditions, One Ocean Expeditions, Golden Fleece Expeditions, and Cheesemans' Ecology Safaris for the field support. We thank Christian Che-Castaldo for helpful input on the capture-recapture model.

Data Availability Statement

Data will be archived on the Dryad data repository upon acceptance.

Funding

The camera network was established using funding from the Darwin Initiative Challenge Fund, Darwin Plus (DPLUS02; to TH). Support for cameras was provided by the UK Foreign & Commonwealth Office (HJL), the John Ellerman Foundation and donations from guests on Quark Expeditions vessels. Funding for this work was provided by NASA (NESSF fellowship #NNX16AO27H; to CY), a NERC DTP studentship (FMJ), the John Ellerman Foundation (TH), and NSF (#1255058; to HJL).

Supporting Information

Supplementary analyses (Appendix S1, S2) are available online. Code to reproduce all analyses are available on GitHub (https://github.com/Penguin-Watch/penguin_watch_model).

Author Contributions

CY, FMJ, HJL, and TH conceived the study. TH established the camera network. All authors contributed to data annotation. CY and HJL conceived of and performed the statistical analyses. CY and FMJ led the writing of the manuscript. CY compiled the Supplementary Materials. All authors contributed to drafts and gave final approval for publication.

References

Bender, N. A., Crosbie, K., & Lynch, H. J. (2016). Patterns of tourism in the Antarctic Peninsula region: a 20-year analysis. *Antarctic Science*, *28*(3), 194–203. https://doi.org/10.1017/S0954102016000031

Bivand, R., Keitt, T., & Rowlingson, B. (2019). *Rgdal: Bindings for the "Geospatial" Data Abstraction Library*.

Bivand, R., Pebesma, E., & Gomez-Rubio, V. (2008). *Applied spatial data analysis with R*. New York: Springer.

Bivand, R., & Rundel, C. (2019). *Rgeos: Interface to Geometry Engine - Open Source ('GEOS')*.

Black, C., Rey, A. R., & Hart, T. (2017). Peeking into the bleak midwinter: investigating nonbreeding strategies of Gentoo Penguins using a camera network. *The Auk*, *134*(3), 520–529. https://doi.org/10.1642/AUK-16-69.1

Boersma, P. D., & Rebstock, G. A. (2014). Climate change increases reproductive failure in magellanic penguins. *PLoS ONE*, *9*(1). https://doi.org/10.1371/journal.pone.0085602

Borboroglu, P. G., & Boersma, P. D. (Eds.). (2013). *Penguins: Natural History and Conservation*. Seattle: University of Washington Press.

Brooks, S. P., & Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, *7*(4), 434–455. https://doi.org/10.1080/10618600.1998.10474787

Cairns, D. K. (1988). Seabirds as Indicators of Marine Food Supplies. *Biological Oceanography*, *5*(4), 261–271.

Chapman, E. W., Hofmann, E. E., Patterson, D. L., Ribic, C. A., & Fraser, W. R. (2011). Marine and terrestrial factors affecting Adélie penguin *Pygoscelis adeliae* chick growth and recruitment off the western Antarctic Peninsula. *Marine Ecology Progress Series*, *436*, 273–289. https://doi.org/10.3354/meps09242

Che-Castaldo, C., Jenouvrier, S., Youngflesh, C., Shoemaker, K. T., Humphries, G., McDowall, P., … Lynch, H. J. (2017). Pan-Antarctic analysis aggregating spatial estimates of Adélie penguin abundance reveals robust dynamics despite stochastic noise. *Nature Communications*, *8*(832). https://doi.org/10.1038/s41467-017-00890-0

Cimino, M. A., Fraser, W. R., Patterson-Fraser, D. L., Saba, V. S., & Oliver, M. J. (2014). Large-scale climate and local weather drive interannual variability in Adélie penguin chick fledging mass. *Marine Ecology Progress Series*, *513*, 253–268. https://doi.org/10.3354/meps10928

Clucas, G. V., Dunn, M. J., Dyke, G., Emslie, S. D., Levy, H., Naveen, R., … Hart, T. (2014). A reversal of fortunes: climate change "winners" and "losers" in Antarctic Peninsula penguins. *Scientific Reports*, *4*, 1–7. https://doi.org/10.1038/srep05024

Commission for the Conservation of Antarctic Marine Living Rosources. (2003). *Report on the twenty-second meeting of the Scientific Committee.* Hobart, Australia.

Commission for the Conservation of Antarctic Marine Living Rosources. (2018). *Krill Fishery Report 2018*. Hobart, Australia.

Commission for the Conservation of Antarctic Marine Living Rosources. (2019). *Statistical Bulletin, Vol. 31*. Hobart, Australia.

Conn, P. B., Johnson, D. S., Fritz, L. W., & Fadely, B. S. (2014). Examining the utility of fishery and survey data to detect prey removal effects on Steller sea lions (*Eumetopias jubatus*). *Canadian Journal of Fisheries and Aquatic Sciences*, *71*(8), 1229–1242. https://doi.org/10.1139/cjfas-2013-0602

Croxall, J. P., Reid, K., & Prince, P. A. (1999). Diet, provisioning and productivity responses of marine predators to differences in availability of Antarctic krill. *Marine Ecology Progress Series*, *177*, 115–131. https://doi.org/10.3354/meps177115

Cury, P. M., Boyd, I. L., Bonhommeau, S., Anker-Nilssen, T., Crawford, R. J. M., Furness, R. W., … Sydeman, W. J. (2011). Global Seabird Response to Forage Fish Depletion - One-Third for the Birds. *Science*, *334*(6063), 1703–1706. https://doi.org/10.1126/science.1212928

Dunn, E. K. (1975). The role of environmental factors in the growth of tern chicks. *Journal of Animal Ecology*, *44*(3), 743–754.

Dunn, M. J., Forcada, J., Jackson, J. A., Waluda, C. M., Nichol, C., & Trathan, P. N. (2019). A long-term study of gentoo penguin (*Pygoscelis papua*) population trends at a major Antarctic tourist site, Goudier Island, Port Lockroy. *Biodiversity and Conservation*, *28*, 37–53. https://doi.org/10.1007/s10531-018-1635-6

Eikenaar, C., Berg, M. L., & Komdeur, J. (2003). Experimental evidence for the influence of food availability on incubation attendance and hatching asynchrony in the Australian reed warbler *Acrocephalus australis*. *Journal of Avian Biology*, *34*(4), 419–427. https://doi.org/10.1111/j.0908-8857.2003.03147.x

Field, J. C., MacCall, A. D., Bradley, R. W., & Sydeman, W. J. (2010). Estimating the impacts of fishing on dependent predators: A case study in the California Current. *Ecological Applications*, *20*(8), 2223–2236. https://doi.org/10.1890/09-0428.1

Frederiksen, M., Jensen, H., Daunt, F., Mavor, R. A., & Wanless, S. (2008). Differential effects of a local industrial sand lance fishery on seabird breeding performance. *Ecological Applications*, *18*(3), 701–710. https://doi.org/10.1890/07-0797.1

Friedman, M. (1937). The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association*, *32*(200), 675–701.

Gelman, A, Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2013). *Bayesian data analysis*. CRC Press.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.

Hijmans, R. (2019). *Raster: geographic data analysis and modeling*.

Hilborn, R., Amoroso, R. O., Bogazzi, E., Jensen, O. P., Parma, A. M., Szuwalski, C., & Walters, C. J. (2017). When does fishing forage species affect their predators? *Fisheries Research*, *191*, 211–221. https://doi.org/10.1016/j.fishres.2017.01.008

Hinke, J. T., Barbosa, A., Emmerson, L. M., Hart, T., Juáres, M. A., Korczak-Abshire, M., …

Southwell, C. (2018). Estimating nest-level phenology and reproductive success of colonial seabirds using time-lapse cameras. *Methods in Ecology and Evolution*, *9*(8), 1853–1863. https://doi.org/10.1111/2041-210X.13015

Hinke, J. T., Cossio, A. M., Goebel, M. E., Reiss, C. S., Trivelpiece, W. Z., & Watters, G. M. (2017). Identifying Risk: Concurrent Overlap of the Antarctic Krill Fishery with Krill-Dependent Predators in the Scotia Sea. *PLoS ONE*, *12*(1), : e0170132. https://doi.org/10.1371/journal.pone.0170132

Hinke, J. T., Polito, M. J., Reiss, C. S., Trivelpiece, S. G., & Trivelpiece, W. Z. (2012). Flexible reproductive timing can buffer reproductive success of *Pygoscelis* spp. penguins in the Antarctic Peninsula region. *Marine Ecology Progress Series*, *454*, 91–104. https://doi.org/10.3354/meps09633

Hinke, J. T., Salwicka, K., Trivelpiece, S. G., Watters, G. M., & Trivelpiece, W. Z. (2007). Divergent responses of *Pygoscelis* penguins reveal a common environmental driver. *Oecologia*, *153*, 845–855. https://doi.org/10.1007/s00442-007-0781-4

Holmes, N. D. (2007). Comparing King, Gentoo, and Royal Penguin Responses to Pedestrian Visitation. *Journal of Wildlife Management*, *71*(8), 2575–2582. https://doi.org/10.2193/2005-715

Holmes, N. D., Giese, M., Achurch, H., Robinson, S., & Kriwoken, L. K. (2006). Behaviour and breeding success of gentoo penguins *Pygoscelis papua* in areas of low and high human activity. *Polar Biology*, *29*(5), 399–412. https://doi.org/10.1007/s00300-005-0070-9

Jones, F. M., Allen, C., Arteta, C., Arthur, J., Black, C., Emmerson, L. M., … Hart, T. (2018). Time-lapse imagery and volunteer classifications from the Zooniverse Penguin Watch project. *Scientific Data*, *5*(1), 180124. https://doi.org/10.1038/sdata.2018.124

Kawaguchi, S., Nicol, S., & Press, A. J. (2009). Direct effects of climate change on the Antarctic krill fishery. *Fisheries Management and Ecology*, *16*(5), 424–427. https://doi.org/10.1111/j.1365-2400.2009.00686.x

Kirchgäßner, A. (2011). An analysis of precipitation data from the Antarctic base

Faraday/Vernadsky. *International Journal of Climatology*, *31*(3), 404–414. https://doi.org/10.1002/joc.2083

Konarzewski, M., & Taylor, J. R. E. (1989). The Influence of Weather Conditions on Growth of Little Auk *Alle alle* Chicks. *Nordic Society Oikos*, *20*(2), 112–116.

Lescroël, A., Ridoux, V., & Bost, C. A. (2004). Spatial and temporal variation in the diet of the gentoo penguin (*Pygoscelis papua*) at Kerguelen Islands. *Polar Biology*, *27*(4), 206–216. https://doi.org/10.1007/s00300-003-0571-3

Lustick, S., & Adams, J. (1977). Seasonal variation in the effects of wetting on the energetics and survival of starlings (*Sturnus vulgaris*). *Comparative Biochemistry and Physiology -- Part A: Physiology*, *56*, 173–177. https://doi.org/10.1016/0300-9629(77)90181-5

Lynch, H. J., Fagan, W. F., & Naveen, R. (2010). Population trends and reproductive success at a frequently visited penguin colony on the western Antarctic Peninsula. *Polar Biology*, *33*, 493–503. https://doi.org/10.1007/s00300-009-0726-y

Lynch, H. J., Naveen, R., & Fagan, W. F. (2008). Censuses of penguin, Blue-eyed Shag *Phalacrocorax atriceps* and Southern Giant Petrel *Macronectes giganteus* populations on the Antarctic Peninsula, 2001-2007. *Marine Ornithology*, *36*, 83–97.

Lynch, H. J., Naveen, R., Trathan, P. N., & Fagan, W. F. (2012). Spatially integrated assessment reveals widespread changes in penguin populations on the Antarctic Peninsula. *Ecology*, *93*(6), 1367–1377. https://doi.org/10.1890/11-1588.1

Lynch, M. A., Youngflesh, C., Agha, N. H., Ottinger, M. A., & Lynch, H. J. (2019). Tourism and stress hormone measures in Gentoo Penguins on the Antarctic Peninsula. *Polar Biology*, *42*(7), 1299–1306. https://doi.org/10.1007/s00300-019-02518-z

Massom, R. A., Stammerjohn, S. E., Smith, R. C., Pook, M. J., Iannuzzi, R. A., Adams, N., … Krouse, H. R. (2006). Extreme Anomalous Atmospheric Circulation in the West Antarctic Peninsula Region in Austral Spring and Summer 2001/02 , and its Profound Impact on Sea Ice and Biota. *Journal of Climate*, *19*(15), 3544–3571.

Mauget, R., Garcia, V., & Jouventin, P. (1995). Endocrine Basis of the Reproductive

Pattern of the Gentoo penguin (*Pygoscelis papua*): Winter Breeding and Extended Laying Period in Northern Populations. *General and Comparative Endocrinology*, *98*(2), 177–184. https://doi.org/10.1006/gcen.1995.1058

Mayewski, P. A., Meredith, M. P., Summerhayes, C. P., Turner, J., Worby, A., Barrett, P. J., … Van Ommen, T. (2009). State of the Antarctic and Southern Ocean Climate System. *Reviews of Geophysics*, *47*. https://doi.org/10.1029/2007RG000231

McDowall, P., & Lynch, H. J. (2017). Ultra-Fine Scale Spatially-Integrated Mapping of Habitat and Occupancy Using Structure-From-Motion. *Plos One*, *12*(1), e0166773.

Miller, A. K., Karnovsky, N. J., & Trivelpiece, W. Z. (2009). Flexible foraging strategies of gentoo penguins *Pygoscelis papua* over 5 years in the South Shetland Islands, Antarctica. *Marine Biology*, *156*(12), 2527–2537. https://doi.org/10.1007/s00227-009-1277-z

Moss, D. (1979). Growth of nestling sparrowhawks (*Accipiter nisus*). *Journal of Zoology*, *187*, 297–314.

Nemoto, T., Okiyama, M., Iwasaki, N., & Kikuchi, T. (1988). Squid as Predators on Krill (*Euphausia superba*) and Prey for Sperm Whales in the Southern Ocean. In D. Sahrhage (Ed.), *Antarctic Ocean and Resources Variability*. https://doi.org/10.1002/2016GL069677.

Newbery, K. B., & Southwell, C. (2009). An automated camera system for remote monitoring in polar environments. *Cold Regions Science and Technology*, *55*(1), 47–51. https://doi.org/10.1016/j.coldregions.2008.06.001

Nichols, J. D. (1992). Capture-Recapture Models. *Bioscience*, *42*(2), 94–102.

Nicol, S., Foster, J., & Kawaguchi, S. (2012). The fishery for Antarctic krill - recent developments. *Fish and Fisheries*, *13*(1), 30–40. https://doi.org/10.1111/j.1467-2979.2011.00406.x

Pfeiffer, S., & Peter, H.-U. (2004). Ecological studies toward the management of an Antarctic tourist landing site (Penguin Island, South Shetland Islands). *Polar Record*, *40*(215), 345–353. https://doi.org/10.1017/S0032247404003845

Piatt, J. F., Harding, A. M. A., Shultz, M., Speckman, S. G., van Pelt, T. I., Drew, G. S., & Kettle, A. B. (2007). Seabirds as indicators of marine food supplies: Cairns revisited. *Marine Ecology Progress Series*, *352*, 221–234. https://doi.org/10.3354/meps07078

Plummer, M. (2003). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling*.

Plummer, M. (2016). *Rjags: Bayesian graphical models using MCMC*.

Polito, M. J., Trivelpiece, W. Z., Patterson, W. P., Karnovsky, N. J., Reiss, C. S., & Emslie, S. D. (2015). Contrasting specialist and generalist patterns facilitate foraging niche partitioning in sympatric populations of *Pygoscelis* penguins. *Marine Ecology Progress Series*, *519*, 221–237. https://doi.org/10.3354/meps11095

R Core Team. (2019). *R: A language and environment for statistical computing.* Retrieved from https://www.r-project.org/

Ratcliffe, N., Adlard, S., Stowasser, G., & McGill, R. (2018). Dietary divergence is associated with increased intra-specific competition in a marine predator. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-25318-7

Ropert-Coudert, Y., Kato, A., Meyer, X., Pellé, M., Macintosh, A. J. J., Angelier, F., … Raclot, T. (2015). A complete breeding failure in an Adélie penguin colony correlates with unusual and extreme environmental events. *Ecography*, *38*(2), 111–113. https://doi.org/10.1111/ecog.01182

Sherley, R. B., Barham, B. J., Barham, P. J., Campbell, K. J., Crawford, R. J. M., Grigg, J., … Votier, S. C. (2018). Bayesian inference reveals positive but subtle effects of experimental fishery closures on marine predator demographics. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1871), 20172443. https://doi.org/10.1098/rspb.2017.2443

Sherley, R. B., Underhill, L. G., Barham, B. J., Barham, P. J., Coetzee, J. C., Crawford, R. J. M., … Upfold, L. (2013). Influence of local and regional prey availability on breeding performance of African penguins *Spheniscus demersus*. *Marine Ecology Progress Series*, *473*, 291–301. https://doi.org/10.3354/meps10070

Southwell, C., & Emmerson, L. (2015). Remotely-operating camera network expands Antarctic seabird observations of key breeding parameters for ecosystem monitoring and management. *Journal for Nature Conservation*, *23*, 1–8. https://doi.org/10.1016/j.jnc.2014.11.002

Stammerjohn, S., Massom, R., Rind, D., & Martinson, D. (2012). Regions of rapid sea ice change: An inter-hemispheric seasonal comparison. *Geophysical Research Letters*, *39*(6), L06501. https://doi.org/10.1029/2012GL050874

Tanton, J. L., Reid, K., Croxall, J. P., & Trathan, P. N. (2004). Winter distribution and behaviour of gentoo penguins *Pygoscelis papua* at South Georgia. *Polar Biology*, *27*(5), 299–303. https://doi.org/10.1007/s00300-004-0592-6

Thyen, S., & Becker, P. H. (2006). Effects of individual life-history traits and weather on reproductive output of Black-headed Gulls *Larus ridibundus* breeding in the Wadden Sea, 1991-97. *Bird Study*, *53*(2), 132–141. https://doi.org/10.1080/00063650609461426

Trathan, P. N., Forcada, J., Atkinson, R., Downie, R. H., & Shears, J. R. (2008). Population assessments of gentoo penguins (*Pygoscelis papua*) breeding at an important Antarctic tourist site, Goudier Island, Port Lockroy, Palmer Archipelago, Antarctica. *Biological Conservation*, *141*(12), 3019–3028. https://doi.org/10.1016/j.biocon.2008.09.006

Turner, J., Lachlan-Cope, T., Colwell, S., & Marshall, G. J. (2005). A positive trend in western Antarctic Peninsula precipitation over the last 50 years reflecting regional and Antarctic-wide atmospheric circulation changes. *Annals of Glaciology*, *41*, 85–91. https://doi.org/10.3189/172756405781813177

Turner, R. (2019). *Deldir: Delaunay triangulation and Dirichlet (Voronoi) tesselation*.

Uotila, P., Lynch, A. H., Cassano, J. J., & Cullather, R. I. (2007). Changes in Antarctic net precipitation in the 21st century based on Intergovernmental Panel on Climate Change (IPCC) model scenarios. *Journal of Geophysical Research Atmospheres*, *112*(D10). https://doi.org/10.1029/2006JD007482

Urbanek, S. (2014). *Jpeg: Read and write JPEG images*.

Volkman, N. J., Presler, P., & Trivelpiece, W. (1980). Diets of Pygoscelid Penguins at King George Island, Antarctica. *The Condor*, *82*, 373–378. https://doi.org/10.2307/1367558

Watters, G. M., Hinke, J. T., & Reiss, C. S. (2020). Long-term observations from Antarctica demonstrate that mismatched scales of fisheries management and predator-prey interaction lead to erroneous conclusions about precaution. *Scientific Reports*, *10*. https://doi.org/10.1038/s41598-020-59223-9

Wickham, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis.* (Springer-Verlag, Ed.). New York.

Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string operations*.

Wickham, H., François, R., Henry, L., & Müller, K. (2018). *Dplyr: A grammar of data manipulation*.

Youngflesh, C. (2018a). MCMCvis: Tools to Visualize, Manipulate, and Summarize MCMC Output. *Journal of Open Source Software*, *3*(24), 640. https://doi.org/10.21105/joss.00640

Youngflesh, C. (2018b). Precipitation could spell peril for penguins. *Frontiers in Ecology and the Environment*, *16*(7), 380–380.

Chapter 6: Discussion and Conclusions

i) New large-scale approaches to ecological monitoring in the Polar Regions

Monitoring studies can provide crucial insight into short- (Edgar & Barrett, 1997) and long-term (Johnson et al., 2006; Mossman et al., 1998) population change. Past small-scale (i.e. localised) monitoring initiatives have allowed assessment of changing population abundance (Croxall et al., 1990; Holmes et al., 1986), demographics (Coulson et al., 2001) and phenology (Burthe et al., 2011; Charmantier et al., 2008; Frederiksen et al., 2004), which in turn can lead to recommendations for conservation policy and management (Woinarski et al., 2010). However, extrapolation of these data to a wider region – or even to species-level – is not necessarily (or indeed likely to be) an accurate representation of the underlying truth, with population trends varying markedly across a species' distribution (Clucas et al., 2014; Jenouvrier et al., 2006; Lynch et al., 2012b). For example, even within a relatively small geographic area there can be 'edge effects', with variables such as nest predation rate and juvenile survival differing between neighbouring regions of a population (Anderson & Boutin, 2002; Kolbe & Janzen, 2002). Similarly, behaviour (e.g. the magnitude and direction of behavioural responses to threats) may vary across a species' distribution (Holmes et al., 2006), and ethological studies which attempt to gain insight into behaviour via small-scale remote monitoring programs (e.g. Bett & Rice, 1993) may erroneously interpret a localised behaviour as a species-wide trait. Since monitoring of a small region may lead to a skewed view of population metrics and behavioural ecology – with potential implications for conservation success – there is a clear requirement for studies with greater geographical breadth.

The use of large-scale ground-based monitoring networks is a burgeoning practice, made possible by advances in remote camera technologies and internal software (Appendix 1; Black, 2018; Swann et al., 2011). In the Polar Regions, some of the most difficult and expensive environments to operate in, they have already been used to great effect – to examine breeding success (Hinke et al., 2018; Southwell & Emmerson, 2015), attendance patterns (Huffeldt & Merkel, 2013), and phenology (Black et al., 2018; Hinke et al., 2018; Julitta et al., 2014), as well as to investigate behaviours which usually go unobserved (Black et al., 2017). In the past, satellite monitoring has revealed and counted new Emperor penguin colonies (Barber-Meyer et al., 2007; Fretwell & Trathan, 2009; Lynch et al., 2012a), in addition to uncovering previously unknown behavioural flexibility in this species (Fretwell et al., 2014). More recently, Unmanned Aerial Vehicles (UAVs) have been used to collect granular data for ecological studies, matching the resolution of image data to the scale of ecological processes (Anderson & Gaston, 2013). In Antarctica, UAVs have been used to estimate the size and abundance of krill-predators (Goebel et al., 2015), and there is evidence that counts derived using UAVs are more accurate than traditional ground counts (Hodgson et al., 2018). Ground-based networks now hold the potential to make similar breakthroughs, though in reality the future of polar monitoring likely lies in multimodal approaches – as demonstrated by Borowicz et al. (2018), who discovered and counted an Adélie penguin 'hotspot' on the Danger Islands using a combination of satellite-, UAV- and ground-based observation techniques.

As large-scale monitoring projects become more popular, it is timely to note that physical construction of a network is just the initial step in a wider data-collection framework (Appendix 1; Lindenmayer & Likens, 2009). As such, these studies may begin to have more in common with 'big data' projects than traditional ecological practices, meriting new

analytical approaches and processing pipelines (Peters et al., 2014; Soranno & Schimel, 2014). With this comes a requirement for integrated processing frameworks (which should have the potential to be adopted by new projects (Mattmann, 2013)), as well as a responsibility for data-curation and sharing (Faniel & Zimmerman, 2011; Hampton et al., 2013). In this thesis I have worked to develop a data-processing framework for *Penguin Watch* (see Fig. 1; section *ii*), which bridges the gap between image collection and investigation of biological hypotheses (Chapters 2-4). I have used this framework to investigate Chinstrap phenology and crèching behaviour, and present key dates for a number of new seasons and sites (Chapter 4). I have also used raw time-lapse imagery, utilising the high temporal resolution data collected by the network to assess breeding success in Gentoo penguins via a 'virtual' mark-recapture study (Chapter 5). This represents the most comprehensive examination of Gentoo demographics to date in terms of spatial scale.

ii) The need for integrated data-processing frameworks

Remote time-lapse cameras are an invaluable tool for ecological monitoring in hard-to-reach locations, allowing high-frequency observations at a low cost (Appendix 1; Hinke et al., 2018; Jones et al., 2018; Southwell & Emmerson, 2015). They provide a practical solution to the logistical challenges associated with large-scale monitoring initiatives, with fully autonomous devices requiring minimal servicing and restricted only by battery life and storage capacity (Jones et al., 2018). As advances in computing technology continue, it is possible that cameras equipped with on-board decision-making features will become common-place, leading to higher quality data and reduced energy requirements

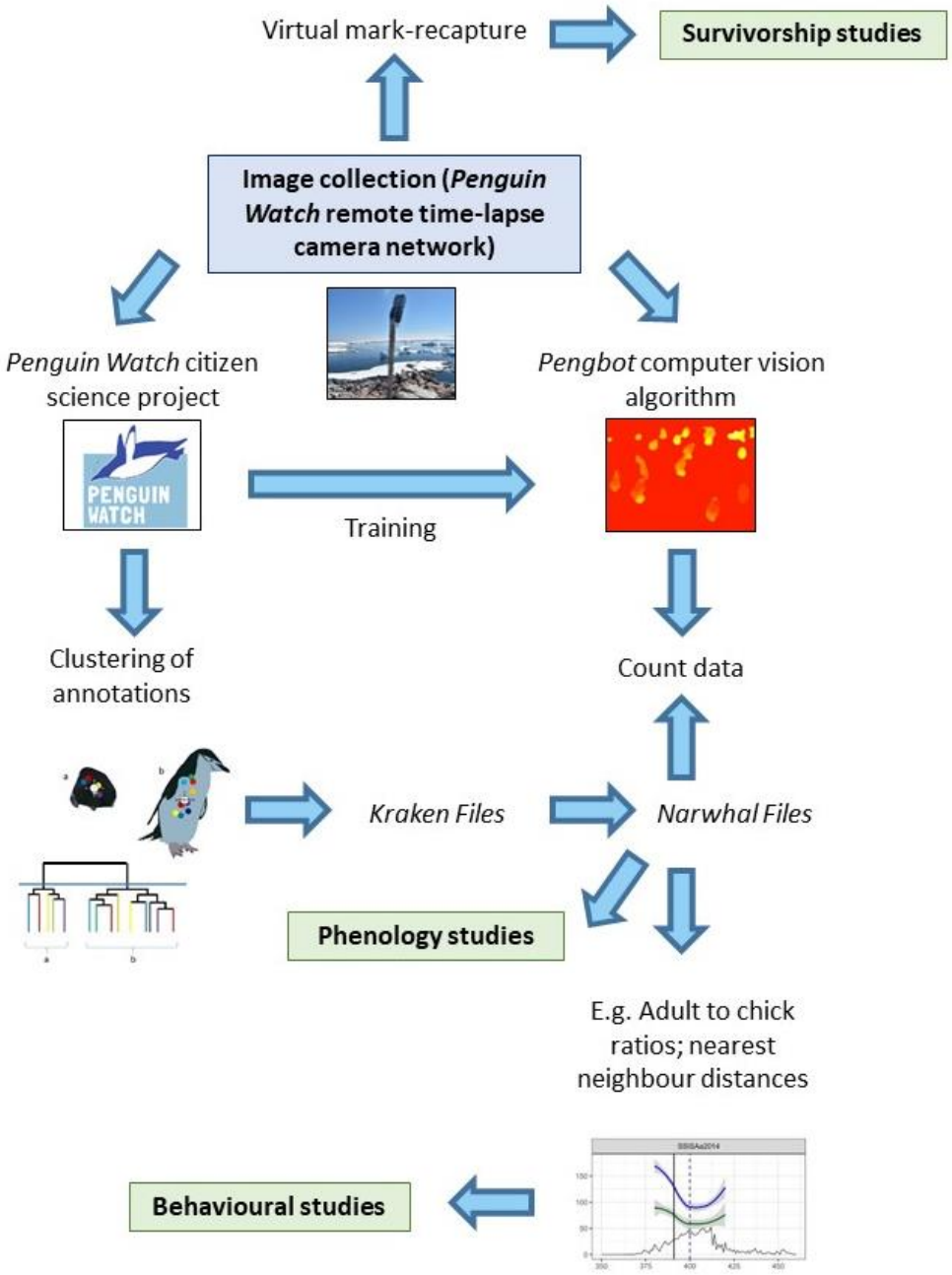


Figure 1: Flow diagram showing the integrated data-processing workflow for *Penguin Watch*. Camera photo: F. M. Jones.

(Appendix 1). For example, a device may have the ability to capture more photographs during a period of interest, such as egg lay or chick hatch, or to halt image capture when the lens is obscured (Appendix 1). Southwell and Emmerson (2015) demonstrate the utility of time-lapse cameras for estimating both population counts and phenological dates.

Although there is a degree of error associated with this approach (as compared to ground-based methods (Southwell & Emmerson, 2015)), this can be true of any empirical study. Even ground-based observations, which cameras attempt to emulate, can be influenced by human error and sampling effects. It is therefore important to assess the trade-off between error rate (i.e. the difference between results obtained using a gold standard and alternative technique) and benefits of the study method, and here there is no doubt that the error rate is outweighed by increased monitoring coverage, potential for decreased disturbance to colonies once cameras have been constructed (e.g. through 'virtual' rather than manual mark-recapture analyses (Chapter 5)), reduced financial costs and increased logistical ease.

While time-lapse cameras present a solution to the challenges associated with ecological monitoring in a harsh environment, they also create an additional challenge – a 'data deluge', which may prove a barrier to progress for small research teams (Faniel & Zimmerman, 2011; Fegraus et al., 2011; Krishnappa & Turner, 2014; Swanson et al., 2015). 'Big data' approaches are becoming increasingly common across a range of biological disciplines (Marx, 2013), including genomics (O'Driscoll et al., 2013; Stephens et al., 2015), biological oceanography (Robinson et al., 2017) and disease biology (Greene & Troyanskaya, 2012; Hay et al., 2013). Alongside this paradigm shift comes an awareness that new techniques are required to process these data, which will likely stem from interdisciplinary approaches and changing attitudes towards data-sharing (Hampton et al., 2013; LaDeau et al., 2017). As biological questions arise that require a 'big data' approach, it only seems logical that lessons can be learnt from fields which have long-incorporated these techniques, such as astronomy and physics (Marx, 2013). Furthermore, it seems that 'shared tools', such as algorithms that can be applied to a wide-range of study systems,

may be the ultimate goal (Mattmann, 2013). In order to achieve this integrated framework, funding agencies must first invest in initiatives that lay the foundations for such cross-discipline research (e.g. the development of file-format translations), whilst also promoting inter-disciplinary training programs for research teams (Mattmann, 2013).

Another key requirement is the development and expansion of data-sharing platforms (Borgman et al., 2007; Faniel & Zimmerman, 2011; Hampton et al., 2013). While open-access publications and datasets are encouraged within the scientific community, there is still a degree of reticence when it comes to making hard-earned results publically available, perhaps because of concerns over data quality, ethics and ownership (Borgman et al., 2007). However, a culture-shift towards increased collaboration – between intra- and inter-disciplinary research groups, professional scientists and students, and researchers and policy-makers – would likely lead to an accelerated rate of discovery and policy-implementation (Borgman et al., 2007; Hampton et al., 2013). In the case of ecological studies, where data relate to a specific day (and hence are unrepeatable), there is an obvious application for data-sharing in the continued development of long-term studies (Borgman et al., 2007). There is clear demand for open-access data (the European Bioinformatics Institute (EBI) received ~9 million data-query requests per day during 2012, representing a 60% increase from the previous year (Marx, 2013)), yet there is a paucity of effective tools for data curation and management (Gray, 2007). Gray (2007) describes the "data iceberg" – the concept that the majority of data remain out of view (i.e. unpublished or unarchived (at least not systematically)), with only the tip of the iceberg represented as a formal publication. Data collection may require specialist instrumentation or fieldwork in remote locations, and can be highly expensive to gather (both financially and in terms of research time). As such, it is wasteful for these data to be

stored out-of-view (or for there to be a substantial lag between data collection and their publication), and the call for improved tools for efficient data curation, visualisation and analysis should be addressed as a priority.

In this thesis I have presented and validated two alternative data-processing streams, both of which can be applied to 'big data' studies. The first of these uses citizen science (the *Penguin Watch* online project, hosted by Zooniverse (Jones et al., 2018)), and the second uses computer vision (the *Pengbot* machine-learning algorithm (Arteta et al., 2016)). While each method comes with its own caveats, both hold huge potential as ecological research tools, with applications extending beyond the boundaries of penguin biology and time-lapse imagery. To address the call for open-access data, I have created two online databases (stored in the Dryad Digital Repository), making more than 60,000 *Penguin Watch* time-lapse photographs publically available, alongside associated metadata, citizen science classifications and count data (Data Citations 1 & 2). These have been systematically archived in standardized formats to facilitate use in future studies (Hampton et al., 2013), and in the accompanying 'Data Descriptor' papers (Chapters 2 & 3; Jones et al., 2018; 2019; Jones et al., in review) I encourage the use of *Penguin Watch* data by ecologists and computer scientists. As previously discussed, there is also increasing demand for shared data-processing tools, and in this thesis I outline a number of new analytical methods (Fig. 1), and provide access to any required software or code. These include ways to produce and visualise population time-series (using *Kraken Files*, *Narwhal Files* and *Narwhal Plots* (Chapter 3)), and to analyse the spatial distribution of point clouds (e.g. nearest neighbour distance metrics (Chapter 3)) – tools which could be applied to a range of data types (including UAV-captured imagery) and species (especially colonially breeding seabirds and mammals). In the immediate future, these techniques should be

applied to the most recent *Penguin Watch* datasets, and tested on 1) aerial photographs processed in *Penguin Watch* (collected by the World Wide Fund for Nature (WWF)) and 2) imagery from its sister project *Seabird Watch* (www.seabirdwatch.org).

iii) *Penguin Watch* and *Pengbot* – strengths and caveats

Both the *Penguin Watch* citizen science project and the *Pengbot* computer vision algorithm have associated caveats, but an awareness of these allows us to employ each strategy in the most appropriate way. For example, reliability testing (i.e. comparing *Penguin Watch*- and *Pengbot*- derived counts to expert-derived estimates) has shown that each method becomes less accurate when processing images with more penguins, likely owing to over-crowding and occlusion of individuals (Chapters 2 & 3; Jones et al., 2018; Jones et al., in review). This issue is difficult to address using *Pengbot*, which uses the sum of pixel densities to generate counts and cannot distinguish between overlapping individuals (Arteta et al., 2016; Jones et al., in review). However, *Penguin Watch* data accuracy could perhaps be improved by highlighting a 'region of interest' within images, and tasking volunteers with counting the penguins within that zone only. It is likely that volunteers begin to feel tired or bored when faced with overly-crowded, difficult images, meaning a reduction in the overall number of penguins to count may lead to more reliable estimates. In future, image sets known to contain a greater number of nests (i.e. those with larger colonies, or where the camera is positioned further away from the nests) should be segmented and passed to the *Penguin Watch* processing stream, while less-crowded images are analysed by *Pengbot*.

Similarly, underestimation of chicks appears to be an issue for both platforms. *Penguin Watch* volunteers may misidentify or miss chicks, particularly when they are small and partially-concealed by a parent in the nest (Jones et al., 2018; Southwell & Emmerson, 2015). In Jones et al. (2018; Chapter 2) I suggest that chicks may also go uncounted as a result of the *Penguin Watch* workflow, with volunteers given the option of moving on to the next photograph after counting 30 individuals. Since publishing this paper, we have altered the workflow, and while volunteers are always able to move to the next photograph (with the option of saying there were 'too many [penguins] to count'), there is no automatic prompt to stop classifying after 30 individuals have been annotated. In addition, we have added more information to the *Penguin Watch* field guide, including images of chicks and breeding cycle diagrams for each relevant species, so volunteers know better when to look out for young chicks. Early identification of chicks is particularly important for phenological studies (currently, 'chick hatch' is a less reliable parameter to extract than 'adult arrival' – see Chapter 4), and further reliability analyses should be carried out to assess whether the improvements to the workflow have increased the accuracy of chick counts. In future, it is hoped that the *Pengbot* algorithm will be able to distinguish between adults and chicks, at which point computer vision will also be a useful tool for comparing 'chick hatch' dates between colonies and seasons.

Despite these caveats, both systems have proved to be powerful processing tools. Time-lapse cameras are unlikely to be able to capture an entire breeding colony within their field-of-view, meaning the imagery they produce is more likely to be used to study behaviour (Black et al., 2016, Black et al., 2017; Haftorn, 1986; Zitterbart et al., 2011), phenology (Chapter 4; Hinke et al., 2018; Penney, 1968; Southwell & Emmerson, 2015) and breeding success (Chapter 5; Hinke et al., 2018; Southwell & Emmerson, 2015) than

to collect census data. As such, being able to extract population trends from imagery is of greater importance than simply raw counts, and in Chapter 3 I demonstrate that population trends are preserved irrespective of the processing method used (i.e. *Penguin Watch*, *Pengbot*, or expert-annotation). In addition, citizen science and – in particular – computer vision allow these trends to be extracted within time-frames that would be impossible using manual annotation. This is essential when handling a network of ~80 cameras, where new data are retrieved each year.

Citizen science has also been used to provide novel solutions to difficult problems. For example, chick fledge date and peak crèche are difficult parameters to constrain to a specific date, even when directly observing colonies. Using *Penguin Watch* data we have been able to define a new proxy for peak crèche based on adult to chick ratios (see Chapter 4 and section *iii*), and estimate fledge date using a 'wisdom of the crowd' logic (Chapter 4; Galton, 1907). Furthermore, *Penguin Watch* has been a powerful tool for public outreach and education, as discussed in Appendix 8.

Given these advantages, it is likely that the golden age of 'big data' processing lies in a combination of approaches, rather than simply being dominated by machine-learning and Artificial Intelligence (AI) strategies (Appendix 1). Furthermore, detailed classifications – such as those obtained using citizen science – are required for continued machine-learning development (Arteta et al., 2016). In terms of *Penguin Watch*, I make the following recommendations:

1) All data should first be processed by *Pengbot*, allowing images containing no animals to be separated from those that do. Reducing the number of 'empty' images that are uploaded to *Penguin Watch* could increase efficiency and make

best use of volunteer effort. However, it is important to retain a proportion of the blank images, since mean session-length (i.e. the amount of time spent classifying) has been shown to increase with the addition of blank images, with volunteers motivated by the "intermittent reward" of discovery (Bowyer et al., 2015).

2) Images containing high numbers of penguins (>20) should be uploaded to *Penguin Watch*, with a 'region of interest' highlighted if necessary (see above).
3) Images captured during the chick-rearing period should be uploaded to *Penguin Watch*, and should be viewed by 15 different volunteers as opposed to 10, to increase the likelihood of chicks being labelled.
4) If a sample of time-lapse images is to be uploaded to *Penguin Watch* (e.g. five images for each day monitored), these should be evenly spread across the daytime period (e.g. 6am, 10am, 2pm, 6pm, 10pm), rather than clustered (e.g. 10am, 11am, 12pm, 1pm, 2pm). This should capture a greater range of behaviours and attendance patterns, as discussed in Chapter 4. Analyses should be carried out to determine the optimal number of images to upload for each day.
5) A portion of images should always be annotated via both pipelines, to ensure continued cross-validation of methods.

iv) *Investigating phenology and breeding success*

Once an effective data-processing pipeline is in place, large-scale monitoring programmes can be used to investigate biological hypotheses over spatial and temporal scales that were previously unmanageable. For example, studies may examine a species over the entirety of its distribution, potentially allowing us to disentangle the effects of threats that

are only present in certain parts of their range (Appendix 1). Moreover, with careful selection of study species, these investigations may shed light on the health of the wider environment, particularly if data are shared between research groups. The health of seabird populations, including penguin colonies, can act as an effective proxy for the health of the wider marine environment (Einoder, 2009; Huffeldt & Merkel, 2013), owing to a number of factors, including their longevity, wide geographical distribution, and varied life histories, which incorporate aspects of both the coastal and marine environments (Croxall et al., 2012; Furness & Camphuysen, 1997; Paleczny et al., 2015; Parsons et al., 2008). In this thesis I use *Penguin Watch* imagery to examine the phenology (Chapter 4) and breeding success (Chapter 5) of Chinstrap and Gentoo penguins, respectively, incorporating data from a substantial portion of their range in the South Atlantic sector of the Southern Ocean. Through examination of the impact of key anthropogenic stressors on Gentoo reproductive success, we can in turn make inferences about the potential effect of these threats on the wider Southern Ocean ecosystem.

Phenology

Shifting phenological parameters are of particular interest to biologists, as they can indicate changes in climatic conditions, and serve as an 'early warning system' for threatened species or ecosystems (Edwards & Richardson, 2004). Uneven phenological shifts within a food web can lead to 'trophic mismatch' (Burthe et al., 2012; Doiron et al., 2015), where, for example, peak resource demand by predators becomes decoupled from peak prey abundance. In birds, a mismatch between the timing of peak breeding and peak food availability can have a significant detrimental effect on offspring and adult fitness (Campbell et al., 2019; Cresswell & McCleery, 2003; Dunn, 2004; Van Noordwijk et al.,

1995). Furthermore, differing degrees of phenological change amongst sympatric species can lead to intensified competitive interactions, which is cause for concern in the *Pygoscelis* penguins, where congeneric species often overlap at breeding colonies (and will increasingly do so if Gentoo populations continue to move southwards in the Antarctic Peninsula region) (Lynch et al., 2012b;c).

Amongst the *Pygoscelis* species, Chinstrap phenology has been the least studied historically. Until recently, investigations were mainly limited to the South Shetland and South Orkney Islands, although a notable exception to this is the study carried out at Waterboat Point (Antarctic Peninsula, 64°48'S, 62°43'W) by Thomas Bagshawe and Maxime Charles Lester in 1921-1922, where they referred to them as the 'ringed' or 'Antarctic' penguins (see Introduction, section *v*, (Bagshawe, 1938)). Before questions relating to phenological change can be considered, it is necessary to reduce this knowledge gap, and create a database of phenological dates which is supplemented each year. Using the *Penguin Watch* network, Black et al. (2018) started this process, presenting dates for a number of new seasons and sites (including the South Sandwich Islands, home to roughly two million breeding pairs of Chinstrap penguins, which is approximately half of the global population (Hart & Convey, 2018; Lynch et al., 2016)). In Chapter 4 I build on this work, contributing dates for a number of new sites (including two new colonies on the Antarctic Peninsula) and years. In order to enhance the scale and repeatability of the process, I show how citizen science-derived data can be used to find key dates. While some manual fine-tuning is still required, this methodology has the potential to dramatically reduce the time-required to extract dates, and also allowed the behaviour associated with chick crèche to be investigated (Chapter 4).

Since the start of chick crèche (or the post-guard period) is difficult to constrain to one particular day, a parameter for 'peak crèche' – based on chick to adult ratio – can be incorporated into the processing pipeline. A clear peak is seen in the chick to adult ratio each season (Appendix 5), coinciding with the period when a number of adults are at sea foraging, and chicks have left the nest and are clearly visible. Another proxy for the post-guard phase is a decrease in average chick second or third nearest neighbour distance (see Chapter 3), representative of huddling, which is considered a diagnostic behaviour of this stage (Wilson, 2009). However, it should be noted that we found marked variation in the degree of huddling between sites and years, and in some cases aggregations were not apparent, even by eye. Taking this into account, studies must be explicit in their definition of crèche (e.g. 'when 50% of chicks have been left unguarded'), and ideally avoid using aggregation behaviour as an indicator.

Through examination of changes in nest-to-adult distances, we support the idea that the onset of crèche occurs gradually across a colony (owing to, for example, asynchrony in hatch date (Moreno et al., 1997; Penteriani et al., 2003; Viñuela et al., 1996)), and it is likely that more aggregation behaviour occurs when colonies are more synchronous in their breeding activity (Emmerson et al., 2011). It would be interesting to use nearest neighbour distances to test whether a shorter time period between the first and last egg lay is an important determinant of aggregation behaviour compared to other key drivers (e.g. adult aggression (Le Bohec et al., 2005; Penteriani et al., 2003), predation intensity (Wilson, 2009) and weather conditions (Black et al., 2016)). Colonies further south tend to be more synchronous in their breeding cycle than those in the north, as they are more temporally constrained by favourable breeding conditions (Croxall, 1984), meaning a latitudinal trend may be observed. However, Black et al. (2016) identified more huddling

behaviour (a greater number of, and larger, aggregations) in South Georgia colonies than on Petermann Island on the Antarctic Peninsula, which contradicts this hypothesis. It is therefore likely that factors such as predation intensity and weather conditions play a greater role in determining huddling behaviour, and it would be interesting to further attempt to parse their relative contributions through future large-scale monitoring studies.

A larger dataset of sites and years is required before the key drivers of phenological timings can be assessed statistically, and deviations from a norm can be detected, and this will be a priority for future *Penguin Watch* research. However, it is interesting to note that the most easterly colony examined (at Saunders Island, South Sandwich Islands (57.78°S, 26.45°W)) departed 24 days after the penultimate colony, suggesting that longitude may be an important determinant of this parameter. The late departure does not appear to be the result of delayed breeding, with 'chick hatch' taking place three days before the colony at Orne Harbour in the same season (2013-2014). Instead, it seems to be the result of extended periods between chick hatch and fledge (69 days, compared to a mean average of 56 days at sites on the South Shetland Islands and Antarctic Peninsula; n = 4), and between chick fledge and adult departure (73 days). Interestingly, this latter stage varies substantially in its duration between sites and years (from 36 days at Booth Island (2016-2017) to 74 days on the Tetrad Islands (2016-2017)). It would be valuable to repeat the analysis using a different definition of departure (for example, when 80% of the colony has left) to see whether this high-level of variability is retained. Plasticity in the duration of key phenological stages has previously been demonstrated (Black et al., 2018), with variation reported across latitudes and between years. Additional data will allow this to be explored further, and the relative influences of location, sea surface temperature, photoperiod and

food availability to be assessed. For example, just as winter breeding in Gentoo penguins on the Crozet Islands is thought to be driven by timing of peak food abundance in the region (see Bost & Jouventin, 1991), timing of food availability in the South Sandwich Islands may be a key determinant of the prolonged breeding and later departure observed here.

Breeding success

Time-lapse photography can also be used as a tool to carry out traditional ecological studies in novel ways. One such example is the use of images to perform 'virtual' mark-recapture analyses, as demonstrated in Chapter 5. Here we utilise the spatial and temporal breadth of the *Penguin Watch* network, and in particular the high frequency of image capture, to examine chick survival (i.e. breeding success) in Gentoo penguin colonies across the Antarctic Peninsula and South Georgia. Such studies are only possible when individuals are recognisable, for example through tagging, natural markings, or fixed position (Appendix 1). Individual Gentoo penguins cannot be distinguished by eye, but chicks will remain in the nest (i.e. a fixed position) until the end of the brood-guard stage, meaning they can be identified to sibling-level during this phase.

The study presented in Chapter 5 is the most comprehensive analysis of Gentoo demographics to date, and covers nearly all of the species' range in the Atlantic sector of the Southern Ocean. As such, it allows investigation of the impact of key environmental stressors across a portion of the Southern Ocean that is of primary concern owing to its increasing levels of human activity (Bender et al., 2016; Nicol et al., 2012). The study found that precipitation levels, krill fishing intensity, and tourism visitation had little impact on breeding success, but identified a potential highly non-linear relationship between

precipitation and breeding success (although this effect was not subject to statistical analyses here, owing to the small sample size of extreme events). During heavy snowfall events egg/chick mortality increased markedly, presumably owing to drowning or chick wetting and hypothermia (Cimino et al., 2014; Lustick & Adams, 1977; Massom et al., 2006; Thyen & Becker, 2006). Precipitation levels are projected to increase across Antarctica and the Southern Ocean over the 21st Century (Uotila et al., 2007), and since storms have been linked to reproductive failure in a number of other penguin and bird species (Boersma & Rebstock, 2014; Cimino et al., 2014; Thyen & Becker, 2006), the results of this study have implications for the wider Southern Ocean ecosystem.

If Gentoo penguins are used as 'marine sentinels' (Boersma, 2008), we can interpret that, at present, climate change (i.e. an indirect anthropogenic impact) is a greater threat to Southern Ocean species than direct anthropogenic stressors (i.e. the krill fishery or tourism). However, this is an over-simplified view. Firstly, with respect to the krill fishery, it may be the case that currently a higher krill catch stems from higher krill abundance, with both penguins and the fishing industry 'winners' during these years, making it difficult to examine the true impact of krill depletion (Conn et al., 2014; Sherley et al., 2013). Furthermore, Gentoo populations may be affected during a different part of their annual cycle, such as the non-breeding winter period, which merits further study. In addition, the variable life-histories of the *Pygoscelis* penguins must be taken into account, and it is likely that the krill-specialist Adélie and Chinstrap penguins (Lishman, 1985; Lynnes et al., 2004) will be more substantially impacted by intense krill harvesting than Gentoo penguins, which have a generalist foraging strategy (Lescroël et al., 2004). Therefore, now that the virtual mark-recapture methodology has been devised and validated, it should be applied to Adélie and Chinstrap colony images as a priority.

Fisheries closure experiments (see e.g. Daunt et al., 2008; Sherley et al., 2015) may help to fully elucidate the impact of the krill fishery on populations of *Pygoscelis* species (and indeed other Southern Ocean predators). Previous experiments have detected positive impacts of closures on predator populations, such as decreased foraging effort in breeding African penguins (*Spheniscus demersus*; Pichegru et al., 2010; 2012), and increased breeding productivity in Black-legged Kittiwakes (*Rissa tridactyla*; Frederiksen et al., 2008). In order to be informative, these experiments should take place over multiple years, allowing the impacts of confounding variables (such as extreme weather events) to be disentangled from the effects of the fishery closure. Furthermore, during analyses it is important to account for factors such as natural variation in prey biomass, and fishing activity over spatial scales beyond those of the immediate study area (Sherley et al., 2015). In addition, caution is required when deciding on where to position marine no-take zones, as fishing activity may be displaced to areas which are equally – or indeed more – vulnerable to its effects. Vessels may also adopt a "fishing the line" approach, concentrating their harvesting effort at the boundary of the protected area (Kellner et al., 2007), which in some cases may at least partially negate any positive impact of the closure itself (Pichegru et al., 2012). Nevertheless, if executed effectively, such experiments may play an important role in fully understanding the true impact of the krill fishery. Additional experiments could be carried out to assess the relative contributions of exploitative versus interference competition (Watters et al., 2020). Watters et al. (2020) suggest that interference competition (where fishing activity has changed the structure or distribution of krill swarms) has occurred along the Antarctic Peninsula, and propose 'open net' experiments to assess its impact. They suggest that, using modern acoustic equipment, the amount of krill that would have been caught (if nets were closed) could be estimated.

Therefore, vessels could operate in an area until, for example, >10% of local krill biomass was “caught” (albeit with open nets), and penguin performance could be assessed and compared to a control experiment where nets were closed (Watters et al., 2020).

In terms of tourism, it is important to remember than direct human visitation may have numerous effects (e.g. physiological stress and behavioural changes (Coetzee & Chown, 2016)) that may not be translated into depressed breeding success. However, the findings of Lynch et al. 2019 suggest that current tourism levels on the Antarctic Peninsula are not causing elevated stress levels in Gentoo penguins, and support our conclusion that tourism visitation is not a significant threat to Southern Ocean penguins at present. Of course, this is not to say that increasing numbers of tourists, or increasing levels of krill harvesting, will not have an impact in the future and continual assessment of these potential threats is of primary importance.

Conclusions

Remote time-lapse cameras are an invaluable tool for ecological research, allowing populations to be monitored across large spatial and temporal scales at high resolution (both in terms of image quality and frequency of image capture). However, camera networks cannot act in isolation, and require a well thought out data-processing pipeline to a) annotate raw imagery and b) extract biologically meaningful metrics, which enables subsequent investigation of biological hypotheses. In this thesis I have demonstrated that citizen science (the *Penguin Watch* online project) and computer vision (the *Pengbot* model) are attractive data-processing tools, which can be used in combination to maximise efficiency and minimise error. Counts derived using these methods can be used to generate population time series, and – in the case of citizen science – spatial data can be used to investigate movement and behaviour, such as chick huddling.

Baseline data spanning species' distributions are essential for disentangling the effects of threats (e.g. climate change) from natural variation within and between populations. Here I use the *Penguin Watch* network to present new data relating to a) Chinstrap penguin phenology and b) Gentoo penguin breeding success, contributing to our understanding of the behavioural ecology and reproductive success of these species in the Scotia Arc region. Our chick survival data suggest that, at present, three key anthropogenic stressors in the Southern Ocean (precipitation, krill fishing activity and tourism) are having little effect on Gentoo penguin breeding success, although concerns are raised regarding severe precipitation events, with implications for a number of Southern Ocean bird species. Continued assessment of these threats is vital, and repetition of our mark-recapture methodology on Chinstrap and Adélie penguin colonies is a key research priority.

Chapter 7: References
(For Chapters 1 (Introduction) & 6 (Discussion and Conclusions)

Abram, N. J., Mulvaney, R., Wolff, E. W., Triest, J., Kipfstuhl, S., Trusel, L. D., … Arrowsmith, C. (2013). Acceleration of snow melt in an Antarctic Peninsula ice core during the twentieth century. *Nature Geoscience*, *6*(5), 404–411. https://doi.org/10.1038/ngeo1787

Ainley, D. G., Ballard, G., Karl, B. J., & Dugger, K. M. (2005). Leopard seal predation rates at penguin colonies of different size. *Antarctic Science*, *17*(03), 335–340. https://doi.org/10.1017/S0954102005002750

Ancel, A., Gilbert, C., Poulin, N., Beaulieu, M., & Thierry, B. (2015). New insights into the huddling dynamics of emperor penguins. *Animal Behaviour*, *110*, 91–98. https://doi.org/10.1016/j.anbehav.2015.09.019

Anderson, D. W., & Keith, J. O. (1980). The human influence on seabird nesting success: Conservation implications. *Biological Conservation*, *18*(1), 65–80. https://doi.org/10.1016/0006-3207(80)90067-1

Anderson, E. M., & Boutin, S. (2002). Edge effects on survival and behaviour of juvenile red squirrels (*Tamiasciurus hudsonicus*). *Canadian Journal of Zoology*, *80*(6), 1038–1046.

Anderson, K., & Gaston, K. J. (2013). Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment*, *11*(3), 138–146. https://doi.org/10.1890/120150

Anderson, O. R. J., Small, C. J., Croxall, J. P., Dunn, E. K., Sullivan, B. J., Yates, O., & Black, A. (2011). Global seabird bycatch in longline fisheries. *Endangered Species Research*, *14*(2), 91–106. https://doi.org/10.3354/esr00347

Antarctic Treaty Secretariat. *The Antarctic Treaty*. , (1959).

Arteta, C., Lempitsky, V., & Zisserman, A. (2016). Counting in the Wild. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer Vision - ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9911* (pp. 483–498). https://doi.org/https://doi.org/10.1007/978-3-319-46478-7_30

Antarctic Treaty Consultative Parties (ATCP). (2006). Site guidelines for visitors (Resolution 2). *Twenty-Ninth Antarctic Treaty Consultative Meeting*, 243. Retrieved from https://www.ats.aq/seleccion.htm

Atkinson, A., Hill, S. L., Pakhomov, E. A., Siegel, V., Reiss, C. S., Loeb, V. J., … Sailley, S. F. (2019). Krill (*Euphausia superba*) distribution contracts southward during rapid regional warming. *Nature Climate Change*, *9*(2), 142–147. https://doi.org/10.1038/s41558-018-0370-z

Auld, T. D., & Leishman, M. R. (2015). Ecosystem risk assessment for Gnarled Mossy Cloud Forest, Lord Howe Island, Australia. *Austral Ecology*, *40*(4), 364–372. https://doi.org/10.1111/aec.12202

Australian Antarctic Division (AAD). (2004). Antarctic territorial claims. Retrieved September 15, 2019, from History of the Antarctic Treaty website: http://www.antarctica.gov.au/law-and-treaty/history/antarctic-territorial-claims

Australian Antarctic Division (AAD). (2016). History of the Antarctic Treaty. Retrieved September 15, 2019, from Antarctic Treaty System website: http://www.antarctica.gov.au/law-and-treaty/history

Australian Antarctic Division (AAD). (2019b). Convention on the Conservation of Antarctic Marine Living Resources. Retrieved September 15, 2019, from http://www.antarctica.gov.au/law-and-treaty/ccamlr

Australian Antarctic Division (AAD). (2019a). Protocol on Environmental Protection to the Antarctic Treaty (The Madrid Protocol). Retrieved September 15, 2019, from Antarctic Treaty System website: http://www.antarctica.gov.au/law-and-treaty/the-madrid-protocol

Bagshawe, T. W. (1938). Notes on the habits of the gentoo and ringed or Antarctic penguins. *Transactions of the Zoological Society of London*, *24*(3), 185–307.

Baker, A. J., Pereira, S. L., Haddrath, O. P., & Edge, K.-A. (2006). Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1582), 11–17. https://doi.org/10.1098/rspb.2005.3260

Ballard, G., & Ainley, D. G. (2005). Killer whale harrassment of Adélie penguins at Ross Island. *Antarctic Science*, *17*(03), 385–386. https://doi.org/10.1017/S0954102005002828

Ballard, G., Toniolo, V., Ainley, D. G., Parkinson, C. L., Arrigo, K. R., & Trathan, P. N. (2010). Responding to climate change: Adélie penguins confront astronomical and ocean boundaries. *Ecology*, *91*(7), 2056–2069. https://doi.org/10.1016/j.dib.2016.02.036

Barber-Meyer, S. M., Kooyman, G. L., & Ponganis, P. J. (2007). Estimating the relative abundance of emperor penguins at inaccessible colonies using satellite imagery. *Polar Biology*, *30*(12), 1565–1570. https://doi.org/10.1007/s00300-007-0317-8

BBS. (2018). North American Breeding Bird Survey. Retrieved September 23, 2019, from https://www.pwrc.usgs.gov/bbs/

Bender, N. A., Crosbie, K., & Lynch, H. J. (2016). Patterns of tourism in the Antarctic Peninsula region: a 20-year analysis. *Antarctic Science*, *28*(3), 194–203. https://doi.org/10.1017/S0954102016000031

Bett, B. J., & Rice, A. L. (1993). The feeding behaviour of an abyssal echiuran revealed by in situ time-lapse photography. *Deep-Sea Research Part I*, *40*(9), 1767–1779. https://doi.org/10.1016/0967-0637(93)90031-W

Bingham, M. (1998). The distribution, abundance and population trends of gentoo, rockhopper and king penguins in the Falkland Islands. *Oryx*, *32*(3), 223–232. https://doi.org/10.1046/j.1365-3008.1998.d01-39.x

Black, C., Collen, B., Johnston, D., & Hart, T. (2016). Why Huddle? Ecological Drivers of Chick Aggregations in Gentoo Penguins, *Pygoscelis papua*, across Latitudes. *Plos One*, *11*(2), e0145676. https://doi.org/10.1371/journal.pone.0145676

Black, C., Collen, B., Lunn, D., Filby, D., Winnard, S., & Hart, T. (2018). Time-lapse cameras reveal latitude and season influence breeding phenology durations in penguins. *Ecology and Evolution*, *8*(16), 8286–8296. https://doi.org/10.1002/ece3.4160

Black, C. E. (2016). A comprehensive review of the phenology of *Pygoscelis* penguins. *Polar Biology*, *39*(3), 405–432. https://doi.org/10.1007/s00300-015-1807-8

Black, C. E. (2018). Spying on seabirds: a review of time-lapse photography capabilities

and limitations. *Seabird*, *31*, 1–14.

Black, C., Rey, A. R., & Hart, T. (2017). Peeking into the bleak midwinter: investigating nonbreeding strategies of Gentoo Penguins using a camera network. *The Auk*, *134*(3), 520–529. https://doi.org/10.1642/AUK-16-69.1

Bland, L. M., Regan, T. J., Dinh, M. N., Ferrari, R., Keith, D. A., Lester, R., … Nicholson, E. (2017). Using multiple lines of evidence to assess the risk of ecosystem collapse. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1863), 20170660. https://doi.org/10.1098/rspb.2017.0660

Boersma, P. D. (2008). Penguins as Marine Sentinels. *BioScience*, *58*(7), 597–607. https://doi.org/10.1641/b580707

Boersma, P. D., Clark, J. A., & Hillgarth, N. (2002). Seabird Conservation. In E. A. Schreiber & J. Burger (Eds.), *Biology of Marine Birds* (pp. 563–566). CRC Press.

Boersma, P. D., & Rebstock, G. A. (2014). Climate change increases reproductive failure in magellanic penguins. *PLoS ONE*, *9*(1). https://doi.org/10.1371/journal.pone.0085602

Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009). Citizen science: a developing tool for expanding science knowledge and scientific literacy. *BioScience*, *59*(11), 977–984. https://doi.org/10.1525/bio.2009.59.11.9

Borboroglu, P. G., & Boersma, P. D. (Eds.). (2013). *Penguins: Natural History and Conservation*. Seattle: University of Washington Press.

Borgman, C. L., Wallis, J. C., & Enyedy, N. (2007). Little Science Confronts the Data Deluge: Habitat ecology, Embedded Sensor Networks, and Digital Libraries. *International Journal on Digital Libraries*, *7*(1–2), 17–30. https://doi.org/10.1007/s00799-007-0022-9

Borowicz, A., McDowall, P., Youngflesh, C., Sayre-McCord, T., Clucas, G., Herman, R., … Lynch, H. J. (2018). Multi-modal survey of Adélie penguin mega-colonies reveals the Danger Islands as a seabird hotspot. *Scientific Reports*, *8*(3926), DOI:10.1038/s41598-018-22313-w. https://doi.org/10.1038/s41598-018-22313-w

Bost, C. A., Thiebot, J. B., Pinaud, D., Cherel, Y., & Trathan, P. N. (2009). Where do penguins go during the inter-breeding period? Using geolocation to track the winter

dispersion of the macaroni penguin. *Biology Letters*, *5*(4), 473–476. https://doi.org/10.1098/rsbl.2009.0265

Bost, C.A., & Jouventin, P. (1991). The breeding performance of the Gentoo Penguin *Pygoscelis papua* at the northern edge of its range. *Ibis*, *133*(1), 14–25.

Bost, C.A., & Jouventin, P. (1990). Evolutionary Ecology of Gentoo Penguins (*Pygoscelis papua*). In Lloyd S. Davis & J. T. Darby (Eds.), *Penguin Biology* (pp. 85–112). San Diego, California: Academic Press.

Bowyer, A., Maidel, V., Lintott, C., Swanson, A., & Miller, G. (2015). This Image Intentionally Left Blank: Mundane Images Increase Citizen Science Participation. *Proceedings of the 2015 Conference on Human Computation & Crowdsourcing*, 2–4. https://doi.org/10.13140/RG.2.2.35844.53121

Bricher, P. K., Lucieer, A., & Woehler, E. J. (2008). Population trends of Adélie penguin (*Pygoscelis adeliae*) breeding colonies: A spatial analysis of the effects of snow accumulation and human activities. *Polar Biology*, *31*(11), 1397–1407. https://doi.org/10.1007/s00300-008-0479-z

British Antarctic Survey (BAS). (2015a). Antarctica Act (1994). Retrieved September 16, 2019, from Environmental Protection website: https://www.bas.ac.uk/about/antarctica/environmental-protection/antarctic-act-1994/

British Antarctic Survey (BAS). (2015b). The Antarctic Treaty Explained. Retrieved September 15, 2019, from The Antarctic Treaty website: https://www.bas.ac.uk/about/antarctica/the-antarctic-treaty/the-antarctic-treaty-explained/

Burthe, S., Butler, A., Searle, K. R., Hall, S. J. G., Thackeray, S. J., & Wanless, S. (2011). Demographic consequences of increased winter births in a large aseasonally breeding mammal (*Bos taurus*) in response to climate change. *Journal of Animal Ecology*, *80*(6), 1134–1144. https://doi.org/10.1111/j.1365-2656.2011.01865.x

Burthe, S., Daunt, F., Butler, A., Elston, D. A., Frederiksen, M., Johns, D., … Wanless, S. (2012). Phenological trends and trophic mismatch across multiple levels of a North Sea pelagic food web. *Marine Ecology Progress Series*, *454*, 119–133.

https://doi.org/10.3354/meps09520

Campbell, K. J., Steinfurth, A., Underhill, L. G., Coetzee, J. C., Dyer, B. M., Ludynia, K., … Sherley, R. B. (2019). Local forage fish abundance influences foraging effort and offspring condition in an endangered marine predator. *Journal of Applied Ecology*, *56*(7), 1751–1760. https://doi.org/10.1111/1365-2664.13409

Carlini, A. R., Coria, N. R., Santos, M. M., Libertelli, M. M., & Donini, G. (2007). Breeding success and population trends in Adélie penguins in areas with low and high levels of human disturbance. *Polar Biology*, *30*(7), 917–924. https://doi.org/10.1007/s00300-006-0251-1

CCAMLR. *Convention on the Conservation of Antarctic Marine Living Resources*. (1980).

Chapman, E. W., Hofmann, E. E., Patterson, D. L., Ribic, C. A., & Fraser, W. R. (2011). Marine and terrestrial factors affecting Adélie penguin *Pygoscelis adeliae* chick growth and recruitment off the western Antarctic Peninsula. *Marine Ecology Progress Series*, *436*, 273–289. https://doi.org/10.3354/meps09242

Charmantier, A., McCleery, R. H., Cole, L. R., Perrins, C., Kruuk, L. E. B., & Sheldon, B. C. (2008). Adaptive Phenotypic Plasticity in Response to Climate Change in a Wild Bird Population. *Science*, *320*(5877), 800–803. https://doi.org/10.1126/science.1157174

Cherel, Y., Verdon, C., & Ridoux, V. (1993). Seasonal importance of oceanic myctophids in king penguin diet at Crozet Islands. *Polar Biology*, *13*(5), 355–357. https://doi.org/10.1007/BF00238362

Christian, C., Lintott, C., Smith, A., Fortson, L., & Bamford, S. (2012). Citizen science: contributions to astronomy research. *ArXiv Preprint ArXiv:1202.2577*. Retrieved from http://arxiv.org/abs/1202.2577

Cimino, M. A., Fraser, W. R., Patterson-Fraser, D. L., Saba, V. S., & Oliver, M. J. (2014). Large-scale climate and local weather drive interannual variability in Adélie penguin chick fledging mass. *Marine Ecology Progress Series*, *513*, 253–268. https://doi.org/10.3354/meps10928

Cimino, M. A., Lynch, H. J., Saba, V. S., & Oliver, M. J. (2016). Projected asymmetric

response of Adélie penguins to Antarctic climate change. *Scientific Reports*, *6*, 28785. https://doi.org/10.1038/srep28785

Clement, J. (2019). Global digital population as of July 2019 (in millions). Retrieved September 23, 2019, from Statista website: https://www.statista.com/statistics/617136/digital-population-worldwide/

Clucas, G. V., Dunn, M. J., Dyke, G., Emslie, S. D., Levy, H., Naveen, R., … Hart, T. (2014). A reversal of fortunes: climate change "winners" and "losers" in Antarctic Peninsula penguins. *Scientific Reports*, *4*, 1–7. https://doi.org/10.1038/srep05024

Cobley, N. D., & Shears, J. R. (1999). Breeding performance of gentoo penguins (*Pygoscelis papua*) at a colony exposed to high levels of human disturbance. *Polar Biology*, *21*(6), 355–360. https://doi.org/10.1007/s003000050373

Coetzee, B. W. T., & Chown, S. L. (2016). A meta-analysis of human disturbance impacts on Antarctic wildlife. *Biological Reviews of the Cambridge Philosophical Society*, *91*(3), 578–596. https://doi.org/10.1111/brv.12184

Collins, P. M., Green, J. A., Dodd, S., Shaw, P. J. A., & Halsey, L. G. (2014). Predation of Black-legged Kittiwake Chicks *Rissa tridactyla* by a Peregrine Falcon *Falco peregrinus*: Insights from Time-lapse Cameras. *The Wilson Journal of Ornithology*, *126*(1), 158–161.

Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR). (2018). Fisheries: Krill fisheries. Retrieved February 20, 2019, from https://www.ccamlr.org/en/fisheries/krill-fisheries

Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR). (2019). About CCAMLR. Retrieved September 16, 2019, from https://www.ccamlr.org/en/organisation

Conn, P. B., Johnson, D. S., Fritz, L. W., & Fadely, B. S. (2014). Examining the utility of fishery and survey data to detect prey removal effects on Steller sea lions (*Eumetopias jubatus*). *Canadian Journal of Fisheries and Aquatic Sciences*, *71*(8), 1229–1242. https://doi.org/10.1139/cjfas-2013-0602

Coulson, T., Catchpole, E. A., Albon, S. D., Morgan, B. J. T., Pemberton, J. M., Clutton-Brock, T. H., … Grenfell, B. T. (2001). Age, Sex, Density, Winter Weather, and Population Crashes in Soay sheep. *Science*, *292*(5521), 1528–1531. https://doi.org/10.1126/science.292.5521.1528

Cowan, C. (2001). John Lachlan Cope's Expedition to Graham Land, 1920-22. Retrieved September 18, 2019, from Scott Polar Research Institute - Online Resources - Expeditions website: https://www.spri.cam.ac.uk/resources/expeditions/blax/

Crawford, R. J. M., Cooper, J., du Toit, M., Greyling, M. D., Hanise, B., Holness, C. L., … Wolfaardt, A. C. (2003). Population and breeding of the gentoo penguin *Pygoscelis papua* at Marion Island, 1994/95 - 2002/03. *African Journal of Marine Science*, *25*, 463–474.

Cresswell, W., & McCleery, R. (2003). How great tits maintain synchronization of their hatch date with food supply in response to long-term variability in temperature. *Journal of Animal Ecology*, *72*(2), 356–366. https://doi.org/10.1046/j.1365-2656.2003.00701.x

Croxall, J. P. (1984). Seabirds. In R. M. Laws (Ed.), *Antarctic Ecology, Vol. 2* (pp. 533–619). New York: Academic Press.

Croxall, J. P., Rothery, P., Pickering, S. P. C., & Prince, P. A. (1990). Reproductive Performance, Recruitment and Survival of Wandering Albatrosses *Diomedea exulans* at Bird Island, South Georgia. *Journal of Animal Ecology*, *59*(2), 775–796.

Croxall, J. P., Trathan, P. N., & Murphy, E. J. (2002). Environmental change and Antarctic seabird populations. *Science*, *297*(5586), 1510–1514. https://doi.org/10.1126/science.1071987

Croxall, J. P., Butchart, S. H. M., Lascelles, B., Stattersfield, A. J., Sullivan, B., Symes, A., & Taylor, P. (2012). Seabird conservation status, threats and priority actions: a global assessment. *Bird Conservation International*, *22*(1), 1–34.

Croxall, J. P., & Kirkwood, E. D. (1979). *The Distribution of Penguins on the Antarctic Peninsula and Islands of the Scotia Sea*. Balogh Scientific Books.

Crutsinger, G. M., Short, J., & Sollenberger, R. (2016). The future of UAVs in ecology: an insider perspective from the Silicon Valley drone industry. *Journal of Unmanned Vehicle*

Systems, *4*, 161–168. https://doi.org/10.1139/juvs-2016-0008

Cruz, I. (2007). Avian taphonomy: observations at two Magellanic penguin (*Spheniscus magellanicus*) breeding colonies and their implications for the fossil record. *Journal of Archaeological Science*, *34*(8), 1252–1261. https://doi.org/10.1016/j.jas.2006.10.016

Culik, B., Adelung, D., & Woakes, A. J. (1990). The Effect of Disturbance on the Heart Rate and Behaviour of Adelie Penguins (*Pygoscelis adeliae*) During the Breeding Season. In K. R. Kerry & G. Hempel (Eds.), *Antarctic Ecosystems* (pp. 177–182). Springer Berlin Heidelberg.

Culik, B. M., & Wilson, R. P. (1995). Penguins disturbed by tourists. *Nature*, *376*(6538), 301–302. https://doi.org/10.1038/376301b0

Daunt, F., Wanless, S., Greenstreet, S. P. R., Jensen, H., Hamer, K. C., & Harris, M. P. (2008). The impact of the sandeel fishery closure on seabird food consumption, distribution, and productivity in the northwestern North Sea. *Canadian Journal of Fisheries and Aquatic Sciences*, *65*(3), 362–381. https://doi.org/10.1139/F07-164

Davis, L. S. (1982). Creching behaviour of Adelie penguin chicks (*Pygoscelis adeliae*). *New Zealand Journal of Zoology*, *9*(2), 279–285. https://doi.org/10.1080/03014223.1982.10423856

Davis, L. S., & Darby, J. T. (1990). *Penguin Biology* (Lloyd S. Davis & J. T. Darby, Eds.). San Diego, California: Academic Press.

Davis, L. S., & Renner, M. (2003). *Penguins*. London: T & A D Poyser.

Derksen, D. V. (1977). A Quantitative Analysis of the Incubation Behavior of the Adélie Penguin. *The Auk*, *94*(3), 552–566. https://doi.org/10.1093/auk/94.3.552

Descamps, S., Tarroux, A., Varpe, Ø., Yoccoz, N. G., Tveraa, T., & Lorentsen, S.-H. (2015). Demographic effects of extreme weather events: snow storms, breeding success, and population growth rate in a long-lived Antarctic seabird. *Ecology and Evolution*, *5*(2), 314–325. https://doi.org/10.1002/ece3.1357

Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., … Purcell, K. (2012). The current state of citizen science as a tool for ecological research and public

engagement. *Frontiers in Ecology and the Environment*, *10*(6), 291–297. https://doi.org/10.1890/110236

Dickinson, J. L., Zuckerberg, B., & Bonter, D. N. (2010). Citizen science as an ecological research tool: challenges and benefits. *The Annual Review of Ecology, Evolution and Systematics*, *41*, 149–172. https://doi.org/10.1146/annurev-ecolsys-102209-144636

Doiron, M., Gauthier, G., & Lévesque, E. (2015). Trophic mismatch and its effects on the growth of young in an Arctic herbivore. *Global Change Biology*, *21*(12), 4364–4376. https://doi.org/10.1111/gcb.13057

Ducklow, H. W., Fraser, W. R., Meredith, M. P., Stammerjohn, S. E., Doney, S. C., Martinson, D. G., … Amsler, C. D. (2013). West Antarctic Peninsula: An ice-dependent coastal marine ecosystem in transition. *Oceanography*, *26*(3), 190–203. https://doi.org/http://dx.doi.org/10.5670/oceanog.2013.62

Ducklow, H. W., Baker, K., Martinson, D. G., Quetin, L. B., Ross, R. M., Smith, R. C., … Fraser, W. (2007). Marine pelagic ecosystems: the West Antarctic Peninsula. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*, 67–94. https://doi.org/10.1098/rstb.2006.1955

Dunn, E. K. (1975). The role of environmental factors in the growth of tern chicks. *Journal of Animal Ecology*, *44*(3), 743–754.

Dunn, M. J., Forcada, J., Jackson, J. A., Waluda, C. M., Nichol, C., & Trathan, P. N. (2019). A long-term study of gentoo penguin (*Pygoscelis papua*) population trends at a major Antarctic tourist site, Goudier Island, Port Lockroy. *Biodiversity and Conservation*, *28*, 37–53. https://doi.org/10.1007/s10531-018-1635-6

Dunn, M. J., Jackson, J. A., Adlard, S., Lynnes, A. S., Briggs, D. R., Fox, D., & Waluda, C. M. (2016). Population size and decadal trends of three penguin species nesting at Signy Island, South Orkney Islands. *PLoS ONE*, *11*(10), 1-15 e0164025. https://doi.org/10.1371/journal.pone.0164025

Dunn, P. (2004). Breeding Dates and Reproductive Performance. *Advances in Ecological Research*, *35*, 69–87.

Dunn, P. O., & Winkler, D. W. (1999). Climate change has affected the breeding date of tree swallows throughout North America. *Proceedings of the Royal Society B: Biological Sciences*, *266*(1437), 2487–2490. https://doi.org/10.1098/rspb.1999.0950

Edgar, G. J., & Barrett, N. S. (1997). Short term monitoring of biotic change in Tasmanian marine reserves. *Journal of Experimental Marine Biology and Ecology*, *213*(2), 261–279. https://doi.org/10.1016/S0022-0981(96)02769-4

Edwards, M., & Richardson, A. J. (2004). Impact of climate change on marine pelagic phenology and trophic mismatch. *Nature, 430*(7002), 881–884.

Einoder, L. D. (2009). A review of the use of seabirds as indicators in fisheries and ecosystem management. *Fisheries Research*, *95*(1), 6–13. https://doi.org/10.1016/j.fishres.2008.09.024

Emmerson, L., Pike, R., & Southwell, C. (2011). Reproductive consequences of environment-driven variation in Adélie penguin breeding phenology. *Marine Ecology Progress Series*, *440*, 203–216. https://doi.org/10.3354/meps09265

Emslie, S. D., Karnovsky, N., & Trivelpiece, W. (1995). Avian predation at penguin colonies on King George Island, Antarctica. *The Wilson Bulletin*, *107*(2), 317–327.

EyesOnALZ. (2019). Stall Catchers. Retrieved September 22, 2019, from https://stallcatchers.com/main

Faniel, I. M., & Zimmerman, A. (2011). Beyond the Data Deluge : A Research Agenda for Large-Scale Data Sharing and Reuse. *The International Journal of Digital Curation*, *6*(1), 58–69.

Fegraus, E. H., Lin, K., Ahumada, J. A., Baru, C., Chandra, S., & Youn, C. (2011). Data acquisition and management software for camera trap data: A case study from the TEAM Network. *Ecological Informatics*, *6*(6), 345–353. https://doi.org/10.1016/j.ecoinf.2011.06.003

Flemming, S. A., Lalas, C., & van Heezik, Y. (2013). Little penguin (*Eudyptula minor*) diet at three breeding colonies in New Zealand. *New Zealand Journal of Ecology*, *37*(2), 199–205.

Forcada, J., Trathan, P. N., Reid, K., Murphy, E. J., & Croxall, J. P. (2006). Contrasting population changes in sympatric penguin species in association with climate warming. *Global Change Biology*, *12*, 411–423. https://doi.org/10.1111/j.1365-2486.2006.01108.x

Fraser, W. R., Carlson, J. C., Duley, P. A., Holm, E. J., & Patterson, D. L. (1999). Using kite-based Aerial Photography for Conducting Adélie Penguin Censuses in Antarctica. *Waterbirds: The International Journal of Waterbird Biology*, *22*(3), 435–440.

Frederiksen, M., Daunt, F., Harris, M. P., & Wanless, S. (2008). The demographic impact of extreme events: stochastic weather drives survival and population dynamics in a long-lived seabird. *Journal of Animal Ecology*, *77*(5), 1020–1029. https://doi.org/10.1111/j.1365-2656.2007.0

Frederiksen, M., Harris, M. P., Daunt, F., Peter, R., & Sarah, W. (2004). Scale-dependent climate signals drive breeding phenology of three seabird species. *Global Change Biology*, *10*(7), 1214–1221. https://doi.org/10.1111/j.1365-2486.2004.00794.x

Frederiksen, M., Jensen, H., Daunt, F., Mavor, R. A., & Wanless, S. (2008). Differential effects of a local industrial sand lance fishery on seabird breeding performance. *Ecological Applications*, *18*(3), 701–710. https://doi.org/10.1890/07-0797.1

Freer, J. J., Mable, B. K., Clucas, G., Rogers, A. D., Polito, M. J., Dunn, M., … Hart, T. (2015). Limited genetic differentiation among chinstrap penguin (*Pygoscelis antarctica*) colonies in the Scotia Arc and Western Antarctic Peninsula. *Polar Biology*, *38*(9), 1493–1502. https://doi.org/10.1007/s00300-015-1711-2

Fretwell, P. T., LaRue, M. A., Morin, P., Kooyman, G. L., Wienecke, B., Ratcliffe, N., … Trathan, P. N. (2012). An Emperor Penguin Population Estimate: The First Global, Synoptic Survey of a Species from Space. *PLoS ONE*, *7*(4), e33751 doi: 10.1371. https://doi.org/10.1371/journal.pone.0033751

Fretwell, P. T., & Trathan, P. N. (2009). Penguins from space: Faecal stains reveal the location of emperor penguin colonies. *Global Ecology and Biogeography*, *18*(5), 543–552. https://doi.org/10.1111/j.1466-8238.2009.00467.x

Fretwell, P. T., & Trathan, P. N. (2019). Emperors on thin ice: three years of breeding failure at Halley Bay. *Antarctic Science*, *31*(3), 133–138.

https://doi.org/10.1017/S0954102019000099

Fretwell, P. T., Trathan, P. N., Wienecke, B., & Kooyman, G. L. (2014). Emperor Penguins Breeding on Iceshelves. *PLoS ONE*, *9*(1), e85285. https://doi.org/10.1371/journal.pone.0085285

Furness, R. W., & Camphuysen, Kees, (C.J.). (1997). Seabirds as monitors of the marine environment. *ICES Journal of Marine Science*, *54*(4), 726–737. https://doi.org/10.1006/jmsc.1997.0243

Galton, F. (1907). Vox Populi (The Wisdom of Crowds). *Nature*, *75*(7), 450–451. https://doi.org/https://dx.doi.org/10.1002/clc.22367

Gianuca, D., Phillips, R. A., Townley, S., & Votier, S. C. (2017). Global patterns of sex- and age-specific variation in seabird bycatch. *Biological Conservation*, *205*, 60–76. https://doi.org/10.1016/j.biocon.2016.11.028

Giese, M. (1996). Effects of human activity on Adelie penguin *Pygoscelis adeliae* breeding success. *Biological Conservation*, *75*, 157–164. https://doi.org/10.1016/0006-3207(95)00060-7

Gilbert, C., Robertson, G., Le Maho, Y., Naito, Y., & Ancel, A. (2006). Huddling behavior in emperor penguins: Dynamics of huddling. *Physiology and Behavior*, *88*(4–5), 479–488. https://doi.org/10.1016/j.physbeh.2006.04.024

Goebel, M. E., Perryman, W. L., Hinke, J. T., Krause, D. J., Hann, N. A., Gardner, S., & LeRoi, D. J. (2015). A small unmanned aerial system for estimating abundance and size of Antarctic predators. *Polar Biology*, *38*(5), 619–630. https://doi.org/10.1007/s00300-014-1625-4

Gonsamo, A., Chen, J. M., & Wu, C. (2013). Citizen Science: linking the recent rapid advances of plant flowering in Canada with climate variability. *Scientific Reports*, *3*(2239), 1–5. https://doi.org/10.1038/srep02239

Gray, J. (2007). Jim Gray on eScience: A Transformed Scientific Method. Transcribed and edited by Tony Hey, Stewart Tansley and Kristin Tolle, 2009. *The Fourth Paradigm*, 17–31. https://doi.org/10.1038/embor.2010.47

Green, J. A., Butler, P. J., Woakes, A. J., & Boyd, I. L. (2003). Energetics of diving in macaroni penguins. *Journal of Experimental Biology*, *206*(1), 43–57. https://doi.org/10.1242/jeb.00059

Greene, C. S., & Troyanskaya, O. G. (2012). Chapter 2: Data-Driven View of Disease Biology. *PLoS Computational Biology*, *8*(12), e1002816. https://doi.org/10.1371/journal.pcbi.1002816

Gura, T. (2013). Citizen science: amateur experts. *Nature*, *496*(7444), 259–261. https://doi.org/10.1038/nj7444-259a

Gwynn, A. M. (1993). Clutch size in *Eudyptes* penguins. *Emu-Austral Ornithology*, *93*(4), 287–290.

Haftorn, S. (1986). A quantitative analysis of the behaviour of the Chinstrap penguin *Pygoscelis antarctica* and Macaroni penguin *Eudyptes chrysolophus* on Bouvetøya during the late incubation and early nestling periods. *Polar Research*, *4*(1), 33–45. https://doi.org/10.1111/j.1751-8369.1986.tb00516.x

Haftorn, S., Somme, L., & Gray, J. S. (1981). A census of penguins and seals on Bouvetoya. In *Bouvetøya, South Atlantic Ocean - Results from the Norweigan Antarctic Research Expeditions 1976/77 and 1978/79* (Vol. 175, pp. 29–35). Retrieved from https://core.ac.uk/download/pdf/52073995.pdf#page=31

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., … Porter, J. H. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment*, *11*(3), 156–162. https://doi.org/10.1890/120103

Happywhale. (2019). Your Whales as Individuals. Retrieved September 22, 2019, from https://happywhale.com/home

Hart, T., & Convey, P. (2018). The South Sandwich Islands–a community of meta-populations across all trophic levels. *Biodiversity*, *19*(1–2), 20–33. https://doi.org/10.1080/14888386.2018.1464952

Hart, T., Jones, F. M., Black, C., Lintott, C. J., Youngflesh, C., Lynch, H. J., … Jessopp, M. (in review). Time-lapse imagery as a low-cost monitoring option for colonial species.

Hay, S. I., George, D. B., Moyes, C. L., & Brownstein, J. S. (2013). Big Data Opportunities for Global Infectious Disease Surveillance. *PLoS Medicine*, *10*(4), e1001413. https://doi.org/10.1371/journal.pmed.1001413

Hill, A., Guralnick, R., Smith, A., Sallans, A., Gillespie, R., Denslow, M., … Fortson, L. (2012). The notes from nature tool for unlocking biodiversity records from museum records through citizen science. *ZooKeys*, *209*, 219–233. https://doi.org/10.3897/zookeys.209.3472

Hindell, M. A. (1988). The diet of the king penguin *Aptenodytes patagonicus* at Macquarie Island. *Ibis*, *130*(2), 193–203. https://doi.org/10.1111/j.1474-919X.1988.tb00970.x

Hinke, J. T., Barbosa, A., Emmerson, L. M., Hart, T., Juáres, M. A., Korczak-Abshire, M., … Southwell, C. (2018). Estimating nest-level phenology and reproductive success of colonial seabirds using time-lapse cameras. *Methods in Ecology and Evolution*, *9*(8), 1853–1863. https://doi.org/10.1111/2041-210X.13015

Hinke, J. T., Salwicka, K., Trivelpiece, S. G., Watters, G. M., & Trivelpiece, W. Z. (2007). Divergent responses of *Pygoscelis* penguins reveal a common environmental driver. *Oecologia*, *153*, 845–855. https://doi.org/10.1007/s00442-007-0781-4

Hodgson, J. C., Mott, R., Baylis, S. M., Pham, T. T., Wotherspoon, S., Kilpatrick, A. D., … Koh, L. P. (2018). Drones count wildlife more accurately and precisely than humans. *Methods in Ecology and Evolution*, *9*(5), 1160–1167. https://doi.org/10.1111/2041-210X.12974

Hof, A. R., & Bright, P. W. (2016). Quantifying the long-term decline of the West European hedgehog in England by subsampling citizen-science datasets. *European Journal of Wildlife Research*, *62*(4), 407–413. https://doi.org/10.1007/s10344-016-1013-1

Holden, J., Yanuar, A., & Martyr, D. J. (2003). The Asian Tapir in Kerinci Seblat National Park, Sumatra: evidence collected through photo-trapping. *Oryx*, *37*(01), 34–40. https://doi.org/10.1017/S0030605303000097

Holmes, N. D., Giese, M., Achurch, H., Robinson, S., & Kriwoken, L. K. (2006). Behaviour

and breeding success of gentoo penguins *Pygoscelis papua* in areas of low and high human activity. *Polar Biology*, *29*, 399–412. https://doi.org/10.1007/s00300-005-0070-9

Holmes, R. T., Sherry, T. W., & Sturges, F. W. (1986). Bird Community Dynamics in a Temperate Deciduous Forest: Long-Term Trends at Hubbard Brook. *Ecological Monographs*, *56*(3), 201–220. https://doi.org/10.2307/2937074

Howard, E., & Davis, A. K. (2004). Documenting the Spring Movements of Monarch Butterflies with Journey North, a Citizen Science Program. In K. S. Oberhauser & M. J. Solensky (Eds.), *The Monarch Butterfly. Biology and Conservation.* (pp. 105–114). New York: Cornell University Press.

Huffeldt, N. P., & Merkel, F. R. (2013). Remote Time-lapse Photography as a Monitoring Tool for Colonial Breeding Seabirds: A Case Study Using Thick-billed Murres (*Uria lomvia*). *Waterbirds: The International Journal of Waterbird Biology*, *36*(3), 330–341.

Humphries, G. R. W., Naveen, R., Schwaller, M., Che-Castaldo, C., McDowall, P., Schrimpf, M., & Lynch, H. J. (2017). Mapping Application for Penguin Populations and Projected Dynamics (MAPPPD): Data and tools for dynamic management and decision support. *Polar Record*, *53*(269), 160–166. https://doi.org/10.1017/S0032247417000055

Ide, R., & Oguma, H. (2013). A cost-effective monitoring method using digital time-lapse cameras for detecting temporal and spatial variations of snowmelt and vegetation phenology in alpine ecosystems. *Ecological Informatics*, *16*, 25–34. https://doi.org/10.1016/j.ecoinf.2013.04.003

iSPEX. (2019). Measure Aerosols with your Smartphone. Retrieved September 25, 2019, from http://ispex.nl/en/vragen/

IUCN-CEM. (2016). The IUCN Red List of Ecosystems. Version 2016-1. Retrieved February 20, 2019, from https://iucnrle.org/

IUCN. (2018). IUCN Red List of Threatened Species. Version 2018-2. Retrieved February 25, 2019, from https://www.iucnredlist.org/search?query=penguins&searchType=species

Ivošević, B., Han, Y. G., Cho, Y., & Kwon, O. (2015). The use of conservation drones in

ecology and wildlife research. *Journal of Ecology and Environment*, *38*(1), 113–118. https://doi.org/10.5141/ecoenv.2015.012

Jennett, C., Furniss, D. J., Iacovides, I., Wiseman, S., Gould, S. J. J., & Cox, A. L. (2014). Exploring Citizen Psych-Science and the Motivations of Errordiary Volunteers. *Human Computation*, *1*(2), 201–220. https://doi.org/10.15346/hc.v1i2.10

Jenouvrier, S., Barbraud, C., & Weimerskirch, H. (2006). Sea ice affects the population dynamics of Adélie penguins in Terre Adélie. *Polar Biology*, *29*(5), 413–423. https://doi.org/10.1007/s00300-005-0073-6

Johnson, L., Camp, R. J., Brink, K. W., & Banko, P. C. (2006). Long-Term Population Monitoring: Lessons Learned From an Endangered Passerine in Hawai'i. *Wildlife Society Bulletin*, *34*(4), 1055–1063. https://doi.org/10.2193/0091-7648(2006)34[1055:lpmllf]2.0.co;2

Jones, F. M., Allen, C., Arteta, C., Arthur, J., Black, C., Emmerson, L. M., … Hart, T. (2018). Time-lapse imagery and volunteer classifications from the Zooniverse Penguin Watch project. *Scientific Data*, *5*(1), 180124. https://doi.org/10.1038/sdata.2018.124

Jones, F. M., Allen, C., Arteta, C., Arthur, J., Black, C., Emmerson, L. M., … Hart, T. (2019). Author Correction: Time-lapse imagery and volunteer classifications from the Zooniverse Penguin Watch Project. *Scientific Data*, *6*(1), 1–2.

Jones, F. M., Arteta, C., Zisserman, A., Lempitsky, V., Lintott, C. J., & Hart, T. (in review). Processing citizen science- and machine- annotated time-lapse imagery for biologically meaningful metrics. *Scientific Data*.

Julitta, T., Cremonese, E., Migliavacca, M., Colombo, R., Galvagno, M., Siniscalco, C., … Menzel, A. (2014). Using digital camera images to analyse snowmelt and phenology of a subalpine grassland. *Agricultural and Forest Meteorology*, *198–199*, 116–125. https://doi.org/10.1016/j.agrformet.2014.08.007

Kalan, A. K., Hohmann, G., Arandjelovic, C. B., McCarthy, M. S., Agbor, A., Angedakin, S., … Kühl, H. S. (2019). Novelty response of wild African apes to camera traps. *Current Biology*, *29*(7), 1211–1217.

Kawaguchi, S., Nicol, S., & Press, A. J. (2009). Direct effects of climate change on the Antarctic krill fishery. *Fisheries Management and Ecology*, *16*(5), 424–427. https://doi.org/10.1111/j.1365-2400.2009.00686.x

Keith, D. (2013). Scientific foundations for an IUCN Red List of Ecosystems. *Plos One Supplementary Material.*, *8*(5), 50–63. https://doi.org/10.1371/journal.pone.0062111.s002

Kellner, J. B., Tetreault, I., Gaines, S. D., & Nisbet, R. M. (2007). Fishing the line near marine reserves in single and multispecies fisheries. *Ecological Applications*, *17*(4), 1039–1054. https://doi.org/10.1890/05-1845

Kirchgäßner, A. (2011). An analysis of precipitation data from the Antarctic base Faraday/Vernadsky. *International Journal of Climatology*, *31*(3), 404–414. https://doi.org/10.1002/joc.2083

Koh, L. P., & Wich, S. A. (2012). Dawn of drone ecology: low-cost autonomous aerial vehicles for conservation. *Tropical Conservation Science*, *5*(2), 121–132. https://doi.org/10.1177/194008291200500202

Kolbe, J. J., & Janzen, F. J. (2002). Spatial and temporal dynamics of turtle nest predation: edge effects. *Oikos*, *99*(3), 538–544. https://doi.org/10.1034/j.1600-0706.2002.11853.x

Konarzewski, M., & Taylor, J. R. E. (1989). The Influence of Weather Conditions on Growth of Little Auk *Alle alle* Chicks. *Nordic Society Oikos*, *20*(2), 112–116.

Kooyman, G. L., & Mullins, J. L. (1990). Ross Sea Emperor Penguin Breeding Populations Estimated by Aerial Photography. In K. R. Kerry & G. Hempel (Eds.), *Antarctic Ecosystems. Ecological Change and Conservation.* (pp. 169–176). https://doi.org/10.1007/978-3-642-84074-6_17

Krishnappa, Y. S., & Turner, W. C. (2014). Software for minimalistic data management in large camera trap studies. *Ecological Informatics*, *24*, 11–16. https://doi.org/10.1016/j.ecoinf.2014.06.004

LaDeau, S. L., Han, B. A., Rosi-Marshall, E. J., & Weathers, K. C. (2017). The Next Decade of Big Data in Ecosystem Science. *Ecosystems*, *20*(2), 274–283.

https://doi.org/10.1007/s10021-016-0075-y

Land-Zandstra, A. M., Devilee, J. L. A., Snik, F., Buurmeijer, F., & van den Broek, J. M. (2016). Citizen science on a smartphone: Participants' motivations and learning. *Public Understanding of Science*, *25*(1), 45–60. https://doi.org/10.1177/0963662515602406

Laskin, D. N., & McDermid, G. J. (2016). Evaluating the level of agreement between human and time-lapse camera observations of understory plant phenology at multiple scales. *Ecological Informatics*, *33*, 1–9. https://doi.org/10.1016/j.ecoinf.2016.02.005

Le Bohec, C., Gauthier-Clerc, M., & Le Maho, Y. (2005). The adaptive significance of crèches in the king penguin. *Animal Behaviour*, *70*(3), 527–538. https://doi.org/10.1016/j.anbehav.2004.11.012

Le Maho, Y. (1977). The emperor penguin: a strategy to live and breed in the cold: morphology, physiology, ecology, and behavior distinguish the polar emperor penguin from the other penguin species, particularly from its close relative, the king penguin. *American Scientist*, *65*(6), 680–693. Retrieved from http://www.jstor.org/stable/pdf/27848168.pdf

Lescroël, A., & Bost, C. A. (2006). Recent decrease in gentoo penguin populations at Iles Kerguelen. *Antarctic Science*, *18*(2), 171–174. https://doi.org/10.1017/S0954102006000198

Lescroël, A., Ridoux, V., & Bost, C. A. (2004). Spatial and temporal variation in the diet of the gentoo penguin (*Pygoscelis papua*) at Kerguelen Islands. *Polar Biology*, *27*(4), 206–216. https://doi.org/10.1007/s00300-003-0571-3

Levy, H., Clucas, G. V., Rogers, A. D., Leache, A. D., Ciborowski, K. L., Polito, M. J., … Hart, T. (2016). Population structure and phylogeography of the Gentoo Penguin (*Pygoscelis papua*) across the Scotia Arc. *Ecology and Evolution*, 1834–1853. https://doi.org/10.1002/ece3.1929

Lindenmayer, D. B., & Likens, G. E. (2009). Adaptive monitoring: a new paradigm for long-term research and monitoring. *Trends in Ecology and Evolution*, *24*(9), 482–486. https://doi.org/10.1016/j.tree.2009.03.005

Lintott, C. J., Schawinski, K., Keel, W., Van Arkel, H., Bennert, N., Edmondson, E., … Vandenberg, J. (2009). Galaxy Zoo: Hanny's Voorwerp, a quasar light echo? *Monthly Notices of the Royal Astronomical Society*, *399*(1), 129–140. https://doi.org/10.1111/j.1365-2966.2009.15299.x

Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., … Van den Berg, J. (2008). Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*, *389*(3), 1179–1189. https://doi.org/10.1111/j.1365-2966.2008.13689.x

Lishman, G. S. (1985). The food and feeding ecology of Adelie penguins (*Pygoscelis adeliae*) and chinstrap penguins (*P. antarctica*) at Signy Island, South Orkney Islands. *Journal of Zoology*, *205*(2), 245–263.

Louw, G. J. (1992). Functional anatomy of the penguin flipper. *Journal of the South African Veterinary Association*, *63*(3), 113–120.

Lustick, S., & Adams, J. (1977). Seasonal variation in the effects of wetting on the energetics and survival of starlings (*Sturnus vulgaris*). *Comparative Biochemistry and Physiology -- Part A: Physiology*, *56*, 173–177. https://doi.org/10.1016/0300-9629(77)90181-5

Lynch, H. J., Crosbie, K., Fagan, W. F., & Naveen, R. (2010b). Spatial patterns of tour ship traffic in the Antarctic Peninsula region. *Antarctic Science*, *22*(2), 123–130. https://doi.org/10.1017/S0954102009990654

Lynch, H. J. (2013). Gentoo Penguin. In P. G. Borboroglu & P. D. Boersma (Eds.), *Penguins: Natural History and Conservation* (pp. 73–88). Seattle: University of Washington Press.

Lynch, H. J., Fagan, W. F., & Naveen, R. (2010a). Population trends and reproductive success at a frequently visited penguin colony on the western Antarctic Peninsula. *Polar Biology*, *33*, 493–503. https://doi.org/10.1007/s00300-009-0726-y

Lynch, H. J., Fagan, W. F., Naveen, R., Trivelpiece, S. G., & Trivelpiece, W. Z. (2012c). Differential advancement of breeding phenology in response to climate may alter staggered breeding among sympatric pygoscelid penguins. *Marine Ecology Progress Series*, *454*, 135–145. https://doi.org/10.3354/meps09252

Lynch, H. J., Naveen, R., & Fagan, W. F. (2008). Censuses of penguin, Blue-eyed Shag *Phalacrocorax atriceps* and Southern Giant Petrel *Macronectes giganteus* populations on the Antarctic Peninsula, 2001-2007. *Marine Ornithology*, *36*, 83–97.

Lynch, H. J., Naveen, R., Trathan, P. N., & Fagan, W. F. (2012b). Spatially integrated assessment reveals widespread changes in penguin populations on the Antarctic Peninsula. *Ecology*, *93*(6), 1367–1377. https://doi.org/10.1890/11-1588.1

Lynch, H. J., White, R., Black, A. D., & Naveen, R. (2012a). Detection, differentiation, and abundance estimation of penguin species by high-resolution satellite imagery. *Polar Biology*, *35*(6), 963–968. https://doi.org/10.1007/s00300-011-1138-3

Lynch, H. J., White, R., Naveen, R., Black, A., Meixler, M. S., & Fagan, W. F. (2016). In stark contrast to widespread declines along the Scotia Arc, a survey of the South Sandwich Islands finds a robust seabird community. *Polar Biology*, *39*(9), 1615–1625. https://doi.org/10.1007/s00300-015-1886-6

Lynch, M. A., Youngflesh, C., Agha, N. H., Ottinger, M. A., & Lynch, H. J. (2019). Tourism and stress hormone measures in Gentoo Penguins on the Antarctic Peninsula. *Polar Biology*, *42*(7), 1299–1306. https://doi.org/10.1007/s00300-019-02518-z

Lynnes, A. S., Reid, K., & Croxall, J. P. (2004). Diet and reproductive success of Adelie and chinstrap penguins: linking response of predators to prey population dynamics. *Polar Biology*, *27*(9), 544–554. https://doi.org/10.1007/s00300-004-0617-1

Maksym, T., Stammerjohn, S. E., Ackley, S., & Massom, R. (2012). Antarctic sea ice - A polar opposite? *Oceanography*, *25*(3), 140–151. https://doi.org/10.5670/oceanog.2012.88

Mallory, M. L., Gilchrist, H. G., Janssen, M., Major, H. L., Merkel, F., Provencher, J. F., & Strøm, H. (2018). Financial costs of conducting science in the Arctic: examples from seabird research. *Arctic Science*, *4*(4), 624–633. https://doi.org/10.1139/as-2017-0019

Marx, V. (2013). The big challenges of big data. *Nature*, *498*, 255–260. https://doi.org/10.1038/498255a

Massaro, M., & Davis, L. S. (2005). Differences in egg size, shell thickness, pore density,

pore diameter and water vapour conductance between first and second eggs of Snares Penguins *Eudyptes robustus* and their influence on hatching asynchrony. *Ibis*, *147*(2), 251–258. https://doi.org/10.1111/j.1474-919x.2005.00379.x

Massaro, M., Davis, L. S., & Davidson, R. S. (2006). Plasticity of brood patch development and its influence on incubation periods in the yellow-eyed penguin *Megadyptes antipodes*: An experimental approach. *Journal of Avian Biology*, *37*(5), 497–506. https://doi.org/10.1111/j.0908-8857.2006.03783.x

Massom, R. A., Stammerjohn, S. E., Smith, R. C., Pook, M. J., Iannuzzi, R. A., Adams, N., … Krouse, H. R. (2006). Extreme Anomalous Atmospheric Circulation in the West Antarctic Peninsula Region in Austral Spring and Summer 2001/02 , and its Profound Impact on Sea Ice and Biota. *Journal of Climate*, *19*(15), 3544–3571.

Masters, K., Oh, E. Y., Simmons, B., Lintott, C., Graham, G., Greenhill, A., & Holmes, K. (2016). Science learning via participation in online citizen science. *ArXiv Preprint ArXiv:1601.05973*.

Mattmann, C. A. (2013). A vision for data science. *Nature*, *493*(7433), 473–475. https://doi.org/10.1038/493473a

McMahon, K. W., Michelson, C. I., Hart, T., McCarthy, M. D., Patterson, W. P., & Polito, M. J. (in review). Divergent Trophics Responses of Sympatric Penguin Species to Historic Anthropogenic Exploitation and Recent Climate Change. *PNAS*.

McNeill, S., Barton, K., Lyver, P., & Pairman, D. (2011). Semi-automated penguin counting from digital aerial photographs. *International Geoscience and Remote Sensing Symposium (IGARSS), IEEE*, 4312–4315. https://doi.org/10.1109/IGARSS.2011.6050185

Merkel, F. R., Johansen, K. L., & Kristensen, A. J. (2016). Use of time-lapse photography and digital image analysis to estimate breeding success of a cliff-nesting seabird. *Journal of Field Ornithology*, *87*(1), 84–95. https://doi.org/10.1111/jofo.12143

Meyer, B., Freier, U., Grimm, V., Groeneveld, J., Hunt, B. P. V., Kerwath, S., … Yilmaz, N. I. (2017). The winter pack-ice zone provides a sheltered but food-poor habitat for larval Antarctic krill. *Nature Ecology and Evolution*, *1*(12), 1853–1861. https://doi.org/10.1038/s41559-017-0368-3

Millow, C. J., Mackintosh, S. A., Lewison, R. L., Dodder, N. G., & Hoh, E. (2015). Identifying Bioaccumulative Halogenated Organic Compounds Using a Nontargeted Analytical Approach: Seabirds as Sentinels. *Plos One*, *10*(5), e0127205. https://doi.org/10.1371/journal.pone.0127205

Miyazaki, M., & Waas, J. R. (2003). Correlations between body size, defensive behaviour and reproductive success in male Little Blue Penguins *Eudyptula minor*: implications for female choice. *Ibis*, *145*(1), 98–105. https://doi.org/10.1046/j.1474-919X.2003.00126.x

Montevecchi, W. A. (2002). Interactions between Fisheries and Seabirds. In E. A. Schreiber & J. Burger (Eds.), *Biology of Marine Birds* (pp. 529–534). CRC Press.

Moreno, J., Carrascal, L. M., Sanz, J. J., Amat, J. A., & Cuervo, J. J. (1994). Hatching asynchrony, sibling hierarchies and brood reduction in the Chinstrap penguin *Pygoscelis antarctica*. *Polar Biology*, *14*(1), 21–30. https://doi.org/10.1007/BF00240268

Moreno, J., Barbosa, A., Potti, J., & Merino, S. (1997). The Effects of Hatching Date and Parental Quality on Chick Growth and Creching Age in the Chinstrap Penguin (*Pygoscelis antarctica*): A Field Experiment. *The Auk*, *114*(1), 47–54.

Mossman, M. J., Hartman, L. M., Hay, R., Sauer, J. R., & Dhuey, B. J. (1998). Monitoring and long-term trends in Wisconsin frog and toad populations. In *Status and Conservation of MidWestern Amphibians* (pp. 169–198).

Mulvaney, R., Abram, N. J., Hindmarsh, R. C. A., Arrowsmith, C., Fleet, L., Triest, J., … Foord, S. (2012). Recent Antarctic Peninsula warming relative to Holocene climate and ice-shelf history. *Nature*, *489*(7414), 141. https://doi.org/10.1038/nature11391

Murphy, R. C. (1936). *Oceanic Birds of South America Vol 1*. Retrieved from https://archive.org/stream/oceanicbirdsofso01murp#page/n0/mode/2up

Nagai, S., Yoshitake, S., Inoue, T., Suzuki, R., Muraoka, H., Nasahara, K. N., & Saitoh, T. M. (2014). Year-to-year blooming phenology observation using time-lapse digital camera images. *Journal of Agricultural Meteorology*, *70*(3), 163–170. https://doi.org/10.2480/agrmet.D-13-00021

Naveen, R., Forrest, S. C., Dagit, R. G., Blight, L. K., Trivelpiece, W. Z., & Trivelpiece, S. G.

(2000). Censuses of penguin, blue-eyed shag, and southern giant petrel populations in the Antarctic Peninsula region, 1994-2000. *Polar Record*, *36*(199), 323–334. https://doi.org/10.1017/S0032247400016818

Naveen, R., Lynch, H. J., Forrest, S., Mueller, T., & Polito, M. (2012). First direct, site-wide penguin survey at Deception Island, Antarctica, suggests significant declines in breeding chinstrap penguins. *Polar Biology*, *35*(12), 1879–1888. https://doi.org/10.1007/s00300-012-1230-3

Newbery, K. B. (2016). *Penguin Nest Picture Analyser*. Retrieved from https://sourceforge.net/projects/nestpicanalyser/

Newbery, K. B., & Southwell, C. (2009). An automated camera system for remote monitoring in polar environments. *Cold Regions Science and Technology*, *55*(1), 47–51. https://doi.org/10.1016/j.coldregions.2008.06.001

Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S., & Crowston, K. (2012). The future of citizen science: Emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment*, *10*(6), 298–304. https://doi.org/10.1890/110294

Nicol, S., Foster, J., & Kawaguchi, S. (2012). The fishery for Antarctic krill - recent developments. *Fish and Fisheries*, *13*(1), 30–40. https://doi.org/10.1111/j.1467-2979.2011.00406.x

Nimon, A. J., Schroter, R. C., & Stonehouse, B. (1995). Heart rate of disturbed penguins. *Nature*, *374*(6521), 415.

NoiseTube. (2017). NoiseTube. Retrieved September 25, 2019, from http://www.noisetube.net/index.html#&panel1-1

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(25), E5716–E5725. https://doi.org/10.1073/pnas.1719367115

O'Connor, P. M., & Claessens, L. P. A. M. (2005). Basic avian pulmonary design and flow-

through ventilation in non-avian theropod dinosaurs. *Nature*, *436*(7048), 253–256. https://doi.org/10.1038/nature03716

O'Driscoll, A., Daugelaite, J., & Sleator, R. D. (2013). "Big data", Hadoop and cloud computing in genomics. *Journal of Biomedical Informatics*, *46*(5), 774–781. https://doi.org/10.1016/j.jbi.2013.07.001

Oceanites. (2019). Antarctic Site Inventory. Retrieved September 19, 2019, from Oceanites Research Portal website: https://oceanites.org/research-portal/antarctic-site-inventory/

Olsson, O. (1996). Seasonal Effects of Timing and Reproduction in the King Penguin : A Unique Breeding Cycle. *Journal of Avian Biology*, *27*(1), 7–14.

Otley, H., Tansell, J., & Scofield, P. (2017). A comprehensive demographic assessment of the endangered Fiordland crested penguin *Eudyptes pachyrhynchus*. *New Zealand Journal of Zoology*, 1–19. https://doi.org/10.1080/03014223.2017.1284135

Overeem, R. L., Peucker (nee Mitchelson), A. J., Austin, C. M., Dann, P., & Burridge, C. P. (2008). Contrasting genetic structuring between colonies of the World's smallest penguin, *Eudyptula minor* (Aves: Spheniscidae). *Conservation Genetics*, *9*(4), 893–905. https://doi.org/10.1007/s10592-007-9414-z

Paleczny, M., Hammill, E., Karpouzi, V., & Pauly, D. (2015). Population Trend of the World's Monitored Seabirds, 1950-2010. *PLoS ONE*, *10*(6), 1–11. https://doi.org/10.1371/journal.pone.0129342

Parsons, M., Mitchell, I., Butler, A., Ratcliffe, N., Frederiksen, M., Foster, S., & Reid, J. B. (2008). Seabirds as indicators of the marine environment. *ICES Journal of Marine Science*, *65*(8), 1520–1526. https://doi.org/10.1093/icesjms/fsn155

Peirano, A., Bordone, A., Marini, S., Piazza, P., & Schiaparelli, S. (2016). Short Note - A simple time-lapse apparatus for monitoring macrozoobenthos activity in Antarctica. *Antarctic Science*, *28*(6), 473–474. https://doi.org/10.1017/S0954102016000377

Penney, R. L., & Lowry, G. (1967). Leopard seal predation of Adelie penguins. *Ecology*, *48*(5), 878–882.

Penney, R. L. (1968). Territorial and social behaviour in the Adélie Penguin. In O. L. Austin (Ed.), *Antarctic Research Series. Vol 12: Antarctic Bird Studies* (pp. 83–131). American Geophysical Union of the National Academy of Sciences - National Research Council.

Penteriani, V., Viñuela, J., Belliure, J., Bustamante, J., & Ferrer, M. (2003). Causal and functional correlates of brood amalgamation in the chinstrap penguin *Pygoscelis antarctica*: parental decision and adult aggressiveness. *Polar Biology*, *26*(8), 538–544. https://doi.org/10.1007/s00300-003-0517-9

Peters, D. P. C., Havstad, K. M., Cushing, J., Tweedie, C., Fuentes, O., & Villanueva-Rosales, N. (2014). Harnessing the power of big data: infusing the scientific method with machine learning to transform ecology. *Ecosphere*, *5*(6), 1–15. https://doi.org/10.1890/ES13-00359.1

Pettorelli, N., Lobora, A. L., Msuha, M. J., Foley, C., & Durant, S. M. (2010). Carnivore biodiversity in Tanzania: revealing the distribution patterns of secretive mammals using camera traps. *Animal Conservation*, *13*(2), 131–139. https://doi.org/10.1111/j.1469-1795.2009.00309.x

Pichegru, L., Grémillet, D., Crawford, R. J. M., & Ryan, P. G. (2010). Marine no-take zone rapidly benefits endangered penguin. *Biology Letters*, *6*(4), 498–501. https://doi.org/10.1098/rsbl.2009.0913

Pichegru, L., Ryan, P. G., van Eeden, R., Reid, T., Grémillet, D., & Wanless, R. (2012). Industrial fishing, no-take zones and endangered penguins. *Biological Conservation*, *156*, 117–125. https://doi.org/10.1016/j.biocon.2011.12.013

Pitman, R. L., & Durban, J. W. (2010). Killer whale predation on penguins in Antarctica. *Polar Biology*, *33*(11), 1589–1594. https://doi.org/10.1007/s00300-010-0853-5

Poisbleau, M., Demongin, L., Strange, I. J., Otley, H., & Quillfeldt, P. (2008). Aspects of the breeding biology of the southern rockhopper penguin *Eudyptes c. chrysocome* and new consideration on the intrinsic capacity of the A-egg. *Polar Biology*, *31*(8), 925–932. https://doi.org/10.1007/s00300-008-0431-2

Project Noah. (2019). Project Noah. Retrieved September 23, 2019, from https://www.projectnoah.org/organisms

Pütz, K., Ingham, R. J., & Smith, J. G. (2000). Satellite tracking of the winter migration of Magellanic Penguins *Spheniscus magellanicus* breeding in the Falkland Islands. *Ibis*, *142*(4), 614–622. https://doi.org/10.1111/j.1474-919x.2000.tb04461.x

Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Cardamone, C., Murray, P., … Vandenberg, J. (2013). Galaxy Zoo: motivations of citizen scientists. *Astronomy Education Review*, *12*(1), 1–41. https://doi.org/10.3847/AER2011021

Rankin, C. H., Abrams, T., Barry, R. J., Bhatnagar, S., Clayton, D. F., Colombo, J., … Thompson, R. F. (2009). Habituation revisited: An updated and revised description of the behavioral characteristics of habituation. *Neurobiology of Learning and Memory*, *92*(2), 135–138. https://doi.org/10.1016/j.nlm.2008.09.012

Ratcliffe, N., Guihen, D., Robst, J., Crofts, S., Stanworth, A., & Enderlein, P. (2015). A protocol for the aerial survey of penguin colonies using UAVs. *Journal of Unmanned Vehicle Systems*, *3*(3), 95–101. https://doi.org/10.1139/juvs-2015-0006

Regel, J., & Pütz, K. (1997). Effect of human disturbance on body temperature and energy expenditure in penguins. *Polar Biology*, *18*, 246–253.

Richardson, A. D., Klosterman, S., & Toomey, M. (2013). Near-Surface Sensor-Derived Phenology. In M. D. Schwartz (Ed.), *Phenology: An Integrative Environmental Science* (Vol. 9789400769, pp. 413–430). https://doi.org/10.1007/978-94-007-6925-0

Robinson, K. L., Luo, J. Y., Sponaugle, S., Guigand, C., & Cowen, R. K. (2017). A Tale of Two Crowds: Public Engagement in Plankton Classification. *Frontiers in Marine Science*, *4*(82), 1–7. https://doi.org/10.3389/fmars.2017.00082

Rogers, T., & Bryden, M. M. (1995). Predation of Adelie penguins (*Pygoscelis adeliae*) by leopard seals (*Hydrurga leptonyx*) in Prydz Bay Antarctica. *Canadian Journal of Zoology*, *73*(5), 1001–1004.

Ropert-Coudert, Y., Kato, A., Meyer, X., Pellé, M., Macintosh, A. J. J., Angelier, F., … Raclot, T. (2015). A complete breeding failure in an Adélie penguin colony correlates with unusual and extreme environmental events. *Ecography*, *38*(2), 111–113. https://doi.org/10.1111/ecog.01182

Schippers, P., Snep, R. P. H., Schotman, A. G. M., Jochem, R., Stienen, E. W. M., & Slim, P. A. (2009). Seabird metapopulations: searching for alternative breeding habitats. *Population Ecology*, *51*(4), 459–470. https://doi.org/10.1007/s10144-009-0159-z

Schwaller, M. R., Benninghoff, W. S., & Olson, C. E. (1984). Prospects for satellite remote sensing of Adelie penguin rookeries. *International Journal of Remote Sensing*, *5*(5), 849–853. https://doi.org/10.1080/01431168408948868

Sherley, R. B., Underhill, L. G., Barham, B. J., Barham, P. J., Coetzee, J. C., Crawford, R. J. M., … Upfold, L. (2013). Influence of local and regional prey availability on breeding performance of African penguins *Spheniscus demersus*. *Marine Ecology Progress Series*, *473*, 291–301. https://doi.org/10.3354/meps10070

Sherley, R. B., Winker, H., Altwegg, R., Van Der Lingen, C. D., Votier, S. C., & Crawford, R. J. M. (2015). Bottom-up effects of a no-take zone on endangered penguin demographics. *Biology Letters*, *11*(7), 20150237. https://doi.org/10.1098/rsbl.2015.0237

Shuttleworth, S., Lintott, C., O'Donnell, J., & Tweddle, J. (2017). Data for the "Orchid Observers" project. Retrieved September 20, 2019, from https://ora.ox.ac.uk/objects/uuid:5a1de8b9-8f22-4f14-8bd3-aea1ac3d4734

Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology and Evolution*, *24*(9), 467–471. https://doi.org/10.1016/j.tree.2009.03.017

Soranno, P. A., & Schimel, D. S. (2014). Macrosystems ecology: big data, big ecology. *Frontiers in Ecology and the Environment*, *12*(1), 3. https://doi.org/10.1890/1540-9295-12.1.3

Southwell, C., & Emmerson, L. (2015). Remotely-operating camera network expands Antarctic seabird observations of key breeding parameters for ecosystem monitoring and management. *Journal for Nature Conservation*, *23*, 1–8. https://doi.org/10.1016/j.jnc.2014.11.002

Stammerjohn, S., Massom, R., Rind, D., & Martinson, D. (2012). Regions of rapid sea ice change: An inter-hemispheric seasonal comparison. *Geophysical Research Letters*, *39*(6), L06501. https://doi.org/10.1029/2012GL050874

Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., … Robinson, G. E. (2015). Big Data: Astronomical or Genomical? *PLoS Biology*, *13*(7), e1002195. https://doi.org/10.1371/journal.pbio.1002195

Sullivan, T. N., Wang, B., Espinosa, H. D., & Meyers, M. A. (2017). Extreme lightweight structures: avian feathers and bones. *Materials Today*, *20*(7), 377–391. https://doi.org/10.1016/j.mattod.2017.02.004

Swann, D. E., Kawanishi, K., & Palmer, J. (2011). Evaluating Types and Features of Camera Traps in Ecological Studies: A Guide for Researchers. In A. F. O'Connell, J. D. Nichols, & K. U. Karanth (Eds.), *Camera Traps in Animal Ecology* (pp. 27–43). https://doi.org/https://doi.org/10.1007/978-4-431-99495-4_3

Swanson, A., Kosmala, M., Lintott, C., & Packer, C. (2016). A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. *Conservation Biology*, *30*(3), 520–531. https://doi.org/10.1111/cobi.12695

Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A., & Packer, C. (2015). Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Scientific Data*, *2*, 150026. https://doi.org/10.1038/sdata.2015.26

Tagg, N., McCarthy, M., Dieguez, P., Bocksberger, G., Willie, J., Mundry, R., … Boesch, C. (2018). Nocturnal activity in wild chimpanzees (*Pan troglodytes*): Evidence for flexible sleeping patterns and insights into human evolution. *American Journal of Physical Anthropology*, *166*(3), 510–529. https://doi.org/10.1002/ajpa.23478

Takahashi, A., Ito, M., Nagai, M., Thiebot, J. B., Mitamura, H., Noda, T., … Watanabe, Y. Y. (2018). Migratory movements and winter diving activity of Adélie penguins in East Antarctica. *Marine Ecology Progress Series*, *589*, 227–239.

Taylor, R. H., Wilson, P. R., & Thomas, B. W. (1990). Status and trends of Adélie penguin populations in the Ross Sea region. *Polar Record*, *26*(159), 293–304. https://doi.org/10.1017/S0032247400011803

Thomas, D. B., Fordyce, R. E., & Gordon, K. C. (2013). Evidence for a krill-rich diet from non-destructive analyses of penguin bone. *Journal of Avian Biology*, *44*(2), 203–207.

https://doi.org/10.1111/j.1600-048X.2012.00095.x

Thyen, S., & Becker, P. H. (2006). Effects of individual life-history traits and weather on reproductive output of Black-headed Gulls *Larus ridibundus* breeding in the Wadden Sea, 1991-97. *Bird Study*, *53*(2), 132–141. https://doi.org/10.1080/00063650609461426

Trathan, P.N., Forcada, J., Atkinson, R., Downie, R. H., & Shears, J. R. (2008). Population assessments of gentoo penguins (*Pygoscelis papua*) breeding at an important Antarctic tourist site, Goudier Island, Port Lockroy, Palmer Archipelago, Antarctica. *Biological Conservation*, *141*(12), 3019–3028. https://doi.org/10.1016/j.biocon.2008.09.006

Trathan, P. N., & Ballard, G. (2013). Adelie Penguin. In P. G. Borboroglu & P. D. Boersman (Eds.), *Penguins: Natural History and Conservation* (pp. 37–57). Seattle: University of Washington Press.

Trathan, P. N., Garcia-Borboroglu, P., Boersma, D., Bost, C.-A., Crawford, R. J. M., Crossin, G. T., … Wienecke, B. (2015). Pollution, habitat loss, fishing, and climate change as critical threats to penguins. *Conservation Biology*, *29*(1), 31–41. https://doi.org/10.1111/cobi.12349

Trathan, P. N. (2004). Image analysis of color aerial photography to estimate penguin population size. *Wildlife Society Bulletin*, *32*(2), 332–343. https://doi.org/10.2193/0091-7648(2004)32[332:iaocap]2.0.co;2

Trivelpiece, W., & Trivelpiece, S. (2013). Chinstrap Penguin. In P. G. Borboroglu & P. D. Boersma (Eds.), *Penguins: Natural History and Conservation* (pp. 59–71). Seattle: University of Washington Press.

Trivelpiece, W. Z., Buckelew, S., Reiss, C., & Trivelpiece, S. G. (2007). The winter distribution of chinstrap penguins from two breeding sites in the South Shetland Islands of Antarctica. *Polar Biology*, *30*(10), 1231–1237. https://doi.org/10.1007/s00300-007-0283-1

Trivelpiece, W. Z., Trivelpiece, S. G., & Volkman, N. J. (1987). Ecological segregation of Adélie, gentoo and chinstrap penguins at King George Island, Antarctica. *Ecology*, *68*(2), 351–361.

Trumbull, D. J., Bonney, R., Bascom, D., & Cabral, A. (2000). Thinking scientifically during participation in a citizen science project. *Science Education*, *84*(2), 265–275.

United Kingdom Government. *Antarctic Act*. , (1994).

Uotila, P., Lynch, A. H., Cassano, J. J., & Cullather, R. I. (2007). Changes in Antarctic net precipitation in the 21st century based on Intergovernmental Panel on Climate Change (IPCC) model scenarios. *Journal of Geophysical Research Atmospheres*, *112*(D10). https://doi.org/10.1029/2006JD007482

Van den Broeke, M. R. (2005). Strong surface melting preceded collapse of Antarctic Peninsula ice shelf. *Geophysical Research Letters*, *32*(12), L12815. https://doi.org/10.1029/2005GL023247

Van Noordwijk, A. J., McCleery, R. H., & Perrins, C. M. (1995). Selection for the Timing of Great Tit Breeding in Relation to Caterpillar Growth and Temperature. *Journal of Animal Ecology*, *64*(4), 451–458.

Vianna, J. A., Noll, D., Dantas, G. P. M., Petry, M. V., Barbosa, A., González-Acuña, D., … Poulin, E. (2017). Marked phylogeographic structure of Gentoo penguin reveals an ongoing diversification process along the Southern Ocean. *Molecular Phylogenetics and Evolution*, *107*, 486–498. https://doi.org/10.1016/j.ympev.2016.12.003

Viblanc, V. A., Smith, A. D., Gineste, B., & Groscolas, R. (2012). Coping with continuous human disturbance in the wild: insights from penguin heart rate response to various stressors. *BMC Ecology*, *12*(1), 1–11. https://doi.org/10.1186/1472-6785-12-10

Viñuela, J., Moreno, J., Carrascal, L. M., Sanz, J. J., Amat, J. A., Ferrer, M., … Cuervo, J. J. (1996). The effect of hatching date on parental care, chick growth, and chick mortality in the chinstrap penguin *Pygoscelis antarctica*. *Journal of Zoology*, *240*(1), 51–58. https://doi.org/10.1111/j.1469-7998.1996.tb05485.x

Votier, S. C., Hatchwell, B. J., Beckerman, A., McCleery, R. H., Hunter, F. M., Pellatt, J., … Birkhead, T. R. (2005). Oil pollution and climate have wide-scale impacts on seabird demographics. *Ecology Letters*, *8*(11), 1157–1164. https://doi.org/10.1111/j.1461-0248.2005.00818.x

Walker, C. H. (1990). Persistent pollutants in fish-eating sea birds - bioaccumulation, metabolism and effects. *Aquatic Toxicology*, *17*(4), 293–324. https://doi.org/10.1016/0166-445X(90)90014-G

Watters, G. M., Hinke, J. T., & Reiss, C. S. (2020). Long-term observations from Antarctica demonstrate that mismatched scales of fisheries management and predator-prey interaction lead to erroneous conclusions about precaution. *Scientific Reports*, *10*. https://doi.org/10.1038/s41598-020-59223-9

Weller, M. W., & Derksen, D. V. (1972). Use of time-lapse photography to study nesting activities of birds. *The Auk*, *89*(1), 196–200.

Wienecke, B., Robertson, G., Kirkwood, R., & Lawton, K. (2007). Extreme dives by free-ranging emperor penguins. *Polar Biology*, *30*(2), 133–142. https://doi.org/10.1007/s00300-006-0168-8

Wilcox, C., Van Sebille, E., & Hardesty, B. D. (2015). Threat of plastic pollution to seabirds is global, pervasive, and increasing. *Proceedings of the National Academy of Sciences*, *112*(38), 11899–11904. https://doi.org/10.1073/pnas.1502108112

Williams, T. D. (1990b). Annual variation in breeding biology of gentoo penguins, *Pygoscelis papua*, at Bird Island, South Georgia. *Journal of Zoology*, *222*(2), 247–258.

Williams, T. D., & Rothery, P. (1990). Factors affecting variation in foraging and activity patterns of gentoo penguins (*Pygoscelis papua*) during the breeding season at Bird Island, South Georgia. *Journal of Applied Ecology*, *27*(3), 1042–1054.

Williams, T. D. (1990a). Growth and survival in macaroni penguin, *Eudyptes chrysolophus*, A- and B-chicks: do females maximise investment in the large B-egg? *Oikos*, *59*(3), 349–354.

Wilson, D. (2009). Causes and Benefits of Chick Aggregations in Penguins. *The Auk*, *126*(3), 688–693. https://doi.org/10.1525/auk.2009.9709

Wilson, R. P., Culik, B., Danfeld, R., & Adelung, D. (1991). People in Antarctica - how much do Adélie Penguins *Pygoscelis adeliae* care? *Polar Biology*, *11*(6), 363–370. https://doi.org/10.1007/BF00239688

Wilson, R. P., & Zimmer, I. (2004). Inspiration by Magellanic penguins: reduced swimming effort when under pressure. *Marine Ecology Progress Series*, *278*, 303–307. https://doi.org/10.3354/meps278303

Wilson, S., Anderson, E. M., Wilson, A. S. G., Bertram, D. F., & Arcese, P. (2013). Citizen science reveals an extensive shift in the winter distribution of migratory western grebes. *PLoS ONE*, *8*(6), e65408. https://doi.org/10.1371/journal.pone.0065408

Woehler, E. J., Penney, R. L., Creet, S. M., & Burton, H. R. (1994). Impacts of human visitors on breeding success and long-term population trends in Adélie Penguins at Casey, Antarctica. *Polar Biology*, *14*, 269–274. https://doi.org/10.1007/BF00239175

Woinarski, J. C. Z., Armstrong, M., Brennan, K., Fisher, A., Griffiths, A. D., Hill, B., … Young, S. (2010). Monitoring indicates rapid and severe decline of native small mammals in Kakadu National Park, northern Australia. *Wildlife Research*, *37*(2), 116–126. https://doi.org/10.1071/WR09125

Young, E. (1994). *Skua and Penguin: Predator and Prey*. New York: Cambridge University Press.

Young, E. C. (1990). Long-Term Stability and Human Impact in Antarctic Skuas and Adélie Penguins. In K. R. Kerry & G. Hempel (Eds.), *Antarctic Ecosystems* (pp. 231–236). https://doi.org/10.1007/978-3-642-84074-6_25

Youngflesh, C. (2018). Precipitation could spell peril for penguins. *Frontiers in Ecology and the Environment*, *16*(7), 380–380.

Zitterbart, D. P., Wienecke, B., Butler, J. P., & Fabry, B. (2011). Coordinated Movements Prevent Jamming in an Emperor Penguin Huddle. *PLoS ONE*, *6*(6), e20260. https://doi.org/10.1371/journal.pone.0020260

Zooniverse. (2019). Welcome to the Zooniverse; People-powered research. Retrieved September 23, 2019, from https://www.zooniverse.org/

Data Citations

Data Citation 1: Jones, F.M., Allen, C., Arteta, C., Arthur, J., Black, C., Emmerson, L.M., Freeman, R., Hines, G., Lintott, C.J., Macháčková, Z., Miller, G., Simpson, R., Southwell, C. Torsey, H.R., Zisserman, A. & Hart, T. (2018). Data from: Time-lapse imagery and volunteer classifications from the Zooniverse Penguin Watch project. Dryad Digital Repository. DOI:10.5061/dryad.vv36g

Data Citation 2: Jones, F.M., Arteta, C., Lempitsky, V., Zisserman, A., Lintott, C.J. & Hart, T. Data from: Processing citizen science- and machine- annotated time-lapse imagery for biologically meaningful metrics. Dryad Digital Repository. DOI:10.5061/dryad.94sp17b

Appendix 1: Time-lapse imagery as a low-cost monitoring option for colonial species

Authors

Tom Hart[1], Fiona M. Jones[1], Caitlin Black[1,2], Chris J. Lintott[3], Casey Youngflesh[4], Heather J. Lynch[5], Alasdair Davies[6], Eamonn Maguire[7], Andrew Zisserman[8], Carlos Arteta[8], Peter Barham[9,10], Louise Emmerson[11], Colin Southwell[11], Mark Jessopp[12,13]

Affiliations

1. Department of Zoology, University of Oxford, 11a Mansfield Road, Oxford, OX1 3SZ, UK
2. Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, UK
3. Zooniverse, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH, UK
4. Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, 06269, USA
5. Institute for Advanced Computational Science, Stony Brook University, Stony Brook, NY 11794, USA
6. Zoological Society of London, Outer Circle, Regent's Park, London, NW1 4RY, UK.
7. Pictet Asset Management, Route des Acacias 60, 1211 Geneva 73

8. Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, UK
9. School of Physics, HH Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol, BS8 1TL, UK
10. Department of Biological Sciences, University of Cape Town, Rondbosch, 7710, South Africa
11. Australian Antarctic Division, 203 Channel Highway, Kingston, Tasmania 7050, Australia
12. School of Biological, Earth and Environmental Science, University College Cork, Enterprise Centre, North Mall, Cork, Ireland.
13. MaREI Centre, Environmental Research Institute, University College Cork, Beaufort Building, Ringaskiddy, Co. Cork, Ireland.

Corresponding author: Tom Hart (tom.hart@zoo.ox.ac.uk)

Abstract

Many of the species in decline around the world are subject to different environmental stressors across their range, so large-scale monitoring programmes, with replication of sites, are necessary to disentangle the relative impacts of these threats. At the same time, budgets for long-term monitoring projects are being cut, and studies are coming under fire for lacking statistical power. For the subset of taxa or environments where a single vantage point affords a representative view of individuals or ecological processes, time-lapse cameras can provide a cost-effective way of collecting time series data at scales that would otherwise be impossible. However, time-lapse camera networks create a problem, in that the scale of data collection easily exceeds our ability to process the raw imagery manually. Although there are still issues with parameter estimation (such as working out the timing of nest failure), citizen science and machine learning provide solutions to scaling up data extraction (such as locating all animals in an image). Crucially, citizen science, machine learning-derived classifiers, and the intersection between them, are key to understanding how to establish monitoring systems that are sensitive to – and sufficiently powerful to detect – behavioural changes in the study system. Citizen science works relatively 'out of the box', and we regard it as a first step for many systems until machine learning algorithms have been sufficiently trained to take some of the workload. Using *Penguin Watch* (www.penguinwatch.org) data as a case study, we discuss a complete workflow from data to parameter estimation and interpretation: the use of citizen science and computer vision for image processing, and parameter estimation and individual recognition for investigating biological questions. We discuss which techniques are easily generalizable to a range of questions, and where more work is needed to supplement 'out

of the box' tools. We conclude with a horizon scan of the advances in camera technology necessary to enable more complex time-lapse monitoring, such as on board computer vision and decision making.

Introduction

The world faces both an extinction crisis and a data deficiency (Legg & Nagy, 2006) to understand and disentangle the contrasting or confounded threats to species across their ranges. Remote settings such as mountain ranges, deserts and polar regions are often most data deficient. Some of these data gaps can be filled by advances in technology or a decrease in cost per unit of monitoring networks. Via a review of time-lapse camera analytical techniques and a case study of the monitoring programme *Penguin Watch*, we show how time-lapse photography has come of age for fixed-point monitoring. We discuss the financial implications, the analytical demands of establishing a monitoring network for a new system or taxa, and where the field is going. We discuss the types of analytical techniques that can be applied to such data, including those which can be used 'out of the box', and those that need to be created *de novo*. We conclude with a call for researchers to work with camera developers to incorporate needs specific to time series monitoring into camera design.

Time-lapse photography and photogrammetry techniques have been employed in research since Muybridge first captured stills of a running horse in 1887 (Muybridge, 1887). It can be defined as: any repeated image of the same subject or view that can be used to infer change over time. For example, the surge and retreat of glaciers over large timescales has been documented using this technique, with images captured via repeated manual ground photography as well as time-lapse cameras (Lenzano et al., 2014) and satellite imagery (Cook & Vaughan, 2010). Ground-level and satellite imagery is routinely used to infer everything from snow and vegetation cover (Zhang et al., 2003; 2004) to the

census of animal populations (Fretwell et al., 2012; LaRue et al., 2014). Motion-triggered passive infrared (PIR) cameras are already widely used in wildlife studies (Kays et al., 2009) for population density estimation (Cusack et al., 2015; Lucas et al., 2015), to identify home ranges (Dillon & Kelly, 2008; Gil-Sanchez et al., 2011), and for anti-poaching or bushmeat monitoring (Hegerl et al., 2017). In contrast, time-lapse cameras have been slower to establish as a widely used technique, particularly with any effort to standardise methods. However, time-lapse cameras have enormous potential to capture behaviours at times of the year that are difficult to directly observe, and in remote settings (Black et al., 2017; 2018). Ground-level time-lapse photography was used for avian studies before the digital age (Harris, 1982; Penney, 1968; Weller & Derksen, 1972), but its use is gaining pace due to the ubiquity of affordable off-the-shelf cameras and cheaper electronic storage.

Time-lapse imagery is ideal for monitoring species that aggregate at high densities at some point in their life history, allowing the maximum number of individuals to be observed using a minimal number of camera units. Colonial breeding species (e.g. seabirds, land-breeding seals, bats and meerkats) are obvious targets for time-lapse photography (Southwell & Emmerson, 2015) because they gather in fixed, known locations on an annual basis. Time-lapse sensors may also be particularly valuable when monitoring migratory species (Kays et al., 2009) (e.g. whales, wildebeest or salmon), provided their migration routes are predictable and constrained.

Penguin Watch (www.penguinwatch.org) is one of an increasing number of camera network projects that creates large volumes of imagery that can be processed to test hypotheses on ecological processes, and to estimate parameters of interest (e.g. survival, reproduction) suitable for long-term monitoring (Jones et al., 2018). Fig. 1 shows a

schematic of the *Penguin Watch* monitoring process, from data collection through processing to analysis and interpretation. Expert annotation, citizen science and automation should not be viewed in isolation, but as an integrated network, to fix wildlife monitoring problems.

Data collection

A single remote camera can capture and store thousands of images, and many such cameras are on the market. Traditionally, studies using time-lapse have utilised off the shelf camera units (Bater et al., 2011; Black et al., 2017) that are prioritised for motion triggered detection of animals, or bespoke systems (Richter et al., 2018), usually comprising a digital SLR and some interval controller (Huffeldt & Merkel, 2013; Newbery & Southwell, 2009). When combined into networks, large scale or 'big data' approaches can answer questions that are impossible to answer through more conventional monitoring studies, due to a lack of statistical power. Decisions need to be made regarding battery life, servicing interval, and orientation to best capture the imagery of interest. One of the biggest gaps for monitoring is the ability to have differing rules for data acquisition over a season or year (most cameras record at the same times and intervals regardless of the day or season) and as yet, no study has incorporated image processing into a camera device to help make decisions about frequency of monitoring.

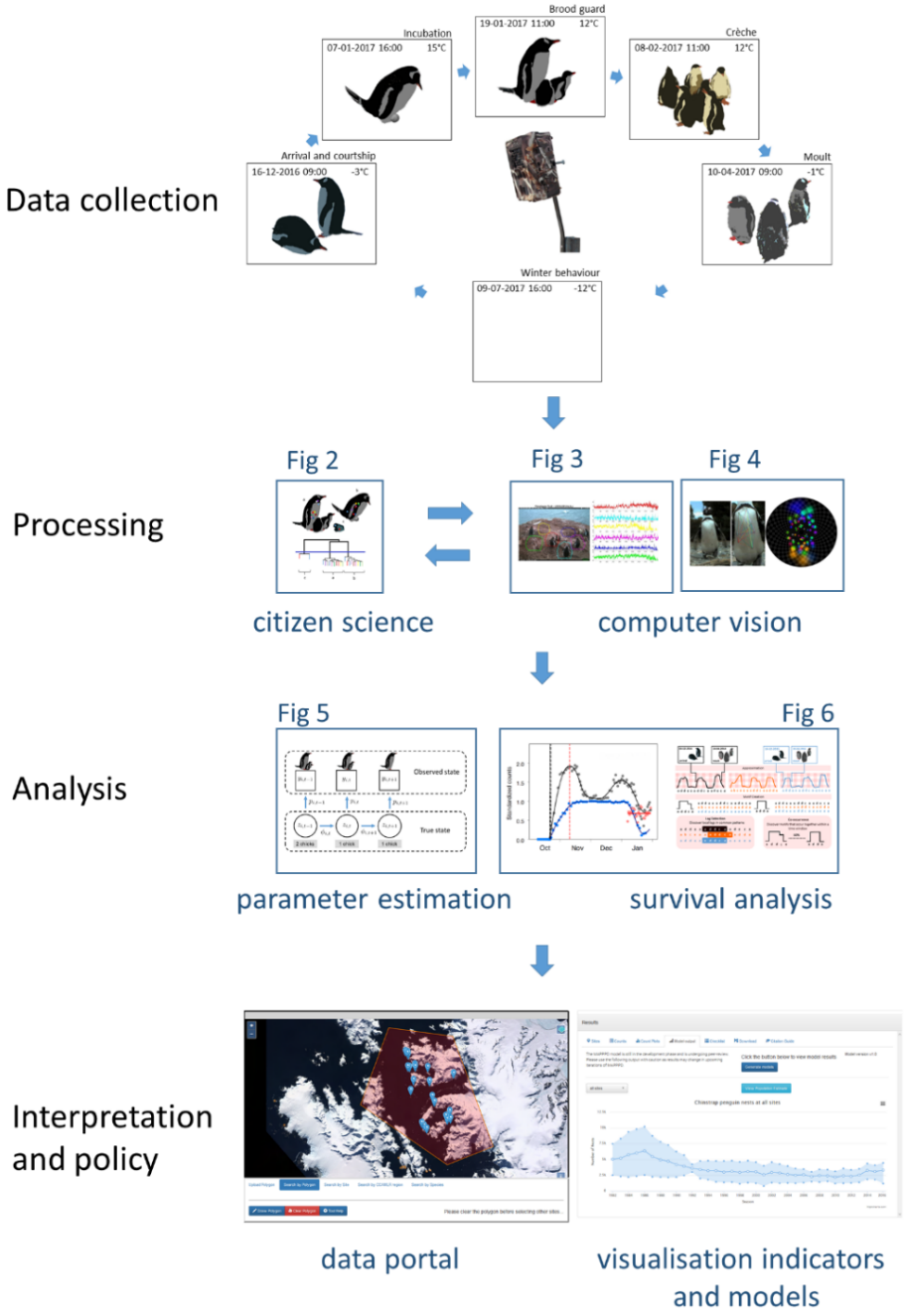


Figure 1: An overview of data collection, processing and analysis. The 'data collection' step shows the phenology of a penguin colony, as viewed by a remote camera (centre) throughout a breeding cycle and over the winter. Cameras can also provide detail on the timing of breeding failures, and possibly even the causative factors (e.g. predation, abandonment, infrequent meal provisioning). Such imagery requires 'processing' and 'analysis' and these can feed back into the 'data collection', either by modification of the network design, or by integration into the cameras whereby the cameras can respond to events that they record. The ultimate goal is for data to feed into conservation 'policy' management decisions.

Data Processing

An inherent problem of large-scale data collection is that – depending on the quantity of imagery captured and the required processing task – it can be impossible to analyse everything via conventional means (e.g. expert manual annotation). Many solutions are available, including citizen science (where non-expert volunteers are used to crowd-source analysis) and computer vision (where a machine learning algorithm is used to process data). At present, we can collect large amounts of imagery, but the cross-utility of analytical techniques means that (with the possible exception of citizen science) tools need to be redeveloped for each new study system. Some techniques exist that require no specific training for the taxa or scenario of interest, such as optic flow analysis used to examine welfare in broiler chickens (Roberts et al., 2011). In contrast, computer vision methods usually require a period of 'training'. Machine learning algorithms must be provided with a subset of pre-processed data in order to learn a task, meaning an alternative method of annotation (e.g. expert- or citizen science- processing) is a prerequisite for this type of analysis. Therefore, regardless of the method that will ultimately be employed for the bulk of image processing, large numbers of human-annotated images will likely be required in the first instance, and these can also be used to calibrate parameters across sites.

Citizen science

As mentioned above, a single remote camera unit can capture and store thousands of images. Hence, camera networks are capable of rapidly collecting datasets that may prove impossible for an individual team of researchers to manually process. Citizen science (Fig.

2) and computer vision (Figs. 3 and 4) both offer solutions to this problem. While they are often viewed as alternatives, they are increasingly being used in synergy, where their differential errors counteract each other (Wright et al., 2017). Furthermore, citizen science classifications provide essential training datasets for machine learning approaches.

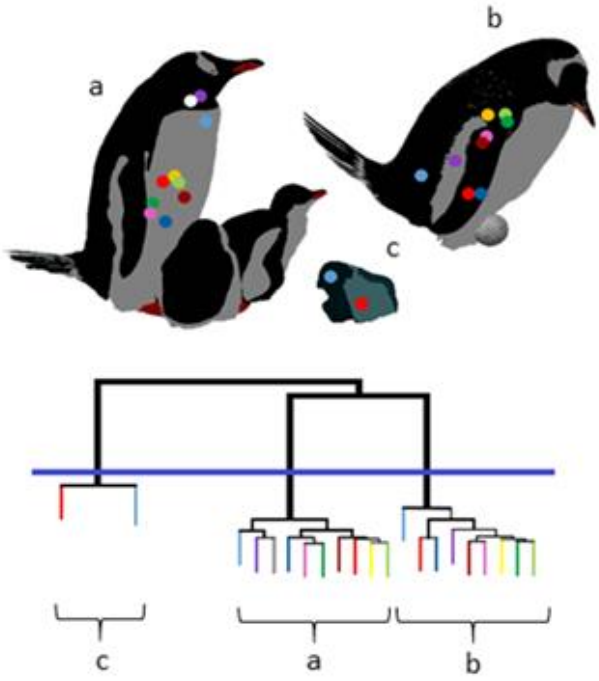


Figure 2: Data reduction of volunteer classifications (clicks on a time-lapse image) aggregated using agglomerative hierarchical clustering to identify penguins ‘a’ and ‘b’. Coloured dots represent ‘clicks’ by ten different volunteers and the coloured branches of the dendrogram show how these markings have been clustered by distance. Object ‘c’ has been (incorrectly) marked by two volunteers, meaning it will be filtered out if an appropriate threshold is applied (e.g. there must be >3 clicks in a cluster for the ‘consensus click’ to be carried forward). Chick markings have been excluded for clarity, and would be clustered separately, as chicks and eggs are annotated with different labels, and potentially analysed using different filtering thresholds.

Successful citizen science projects present volunteers with a simple task, which can be easily learnt via a short tutorial, and produce data of comparable reliability to expert data (Swanson et al., 2015). Regular error-checking and aggregation of multiple users' data – with the potential for filtering of erroneous annotations – are vital to producing reliable output (Fig. 2). *Penguin Watch* is a Zooniverse (www.zooniverse.org) project, first launched in September 2014, which aims to identify the position of – and thereby count – every penguin in each image from a network of over 80 time-lapse cameras positioned across the Southern Ocean. By September 2017 over six million images had been classified by nearly 48,000 registered volunteers and a wealth of anonymous participants.

Within *Penguin Watch*, every image is shown to four volunteers. If at least one person identifies a penguin (or other animal/object), six more volunteers are asked to classify the image (i.e. a total of ten people see the image before it is retired). These numbers could be altered to change the accuracy of the results, and efficiency of data collection. Volunteers are asked to 'tag' individuals, labelling them as 'adult', 'chick', or 'egg' (for penguins) or 'other' (for different fauna, ships or tourists). Volunteers can also highlight unusual features of an image, such as the presence of a rare animal – something that computer vision may not be able to accommodate. After image retirement, the classifications made by each citizen scientist are aggregated using agglomerative hierarchical clustering (Fig. 2), and a single 'consensus click' is produced for each group of markings. This method effectively averages volunteer input, and allows incorrect markings to be filtered out. For example, a threshold could be introduced that discards any 'consensus click' that was drawn from a cluster of fewer than four classifications. In Fig. 2, there are two erroneous markings on object 'c'; these are clustered together (see

dendrogram), but would be filtered out because fewer than four classifications form cluster ‘c’.

Computer vision for counting problems

Modern approaches to ecology often require automated data collection and processing, which is already used widely for birds and mammals (Weinstein, 2018). Automated approaches to image data extraction are rooted in machine learning techniques such as Convolutional Neural Networks (CNN), which have become popular due to their ability to achieve human-level performance on specific recognition tasks. However, before machine learning can automate a task, they require a lot of pre-annotated data to learn, test and validate themselves. In the context of time-lapse monitoring, the task may centre on counting objects in images, which consists of estimating the number of instances of an object appearing in a scene, but could also include looking for colour changes, for example in flowering plants (Graham et al., 2010). Automatic counting in *Penguin Watch* imagery consists of a CNN trained to estimate an object density map (Arteta et al., 2016; Lempitsky & Zisserman, 2010), where pixel density (the proportion of a penguin that each pixel represents) is estimated rather than individuals identified. Summing the ‘penguin pixel’ densities in any given area equals the number of penguins (Fig. 3).

Once trained, the counting algorithm can count previously unseen images, either over the whole image or a region of interest. Training the counting CNN requires a large dataset of images (in the order of 72,000 for this task) (Arteta et al., 2016), with each penguin labelled within it (Jones et al., in review). Although placing dots on an image is an easy type of label to provide, the large number of images required still makes it an extremely labour-

intensive task. Crowdsourcing (via citizen science or by micropayment sites such as Amazon's Mechanical Turk; Mason & Suri, 2012) can overcome this, but can be expensive – both financially (in the case of professional annotation) or in terms of accuracy (variation in annotations can prove challenging). Nevertheless, the combination of the annotations of five to ten volunteers over the same image provides sufficient information to train the counting CNN successfully (Arteta et al., 2016) (Fig. 2), and the CNN continues to learn as more annotations are added.

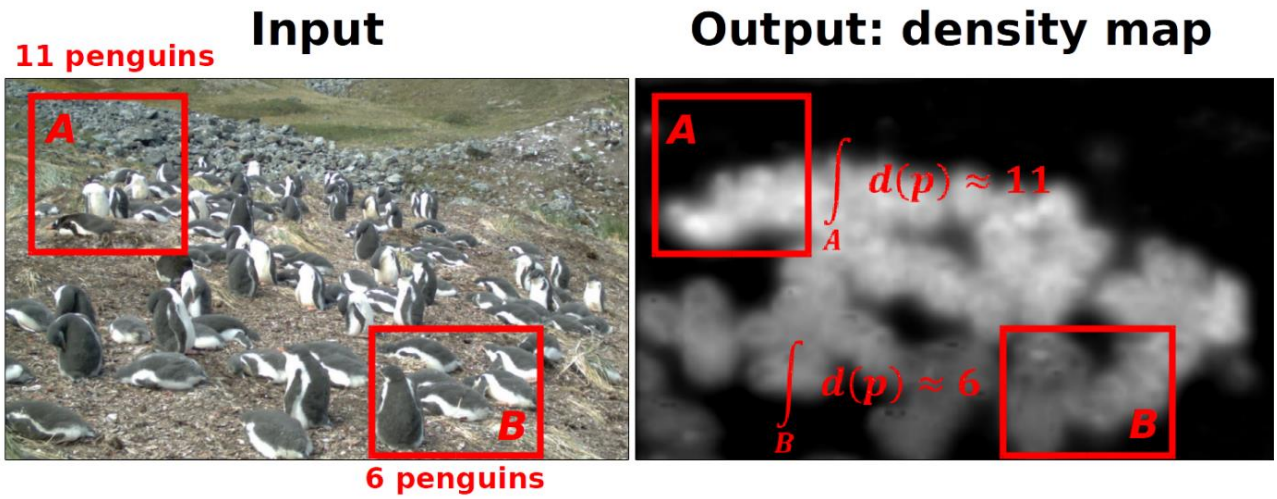


Figure 3: Left) input data, in this case a time-lapse image of a Gentoo penguin (*Pygoscelis papua*) colony, with two regions of interest highlighted in red, and a population count of these two regions; right) output data, showing the same population totals in each box, but derived from pixel densities. Pixels further away from the camera appear brighter as they represent a 'higher density' of penguins (individuals appear smaller in the distance and therefore fewer pixels are required to form one penguin). For the opposite reason, penguin pixels in the foreground appear dimmer (individuals are larger, so more pixels are required to form one penguin).

Computer vision for unique identification

Many animals carry unique patterns that are stable over their lifetime – a leopard cannot change its spots any more than a human can change their fingerprints or iris patterns. The constancy and uniqueness of such patterns makes them ideal to recognize specific individuals in a non-invasive fashion. As such, they can be used to estimate survival, movement and demography using the techniques discussed below. Unique markings have been used to successfully identify individuals in a number of species. In some cases, such as whale sharks (Southwell & Emmerson, 2015) or humpback whales (HappyWhale, 2019), researchers or citizen scientists encounter and photograph animals as part of voyages. In other cases, images are obtained from remote camera traps or CCTV systems (Sherley et al., 2010a; 2010b). Once the images have been obtained, the pattern is extracted, indexed, and matched with a database of known animals. Sometimes this step can be automated, but often it is done entirely manually, or manually with machine assistance (Arzoumanian et al., 2005).

Automated systems of image capture are subject to environmental conditions. For example, in the African penguin system (Sherley et al., 2010a) – where individuals can be identified by a spot pattern – lighting conditions or dirt can change the number of spots detected, and the relative positions of spots can change with orientation and posture of the animal. Therefore, spots (or other features of interest) must be stored as a pattern in a robust form, such as a set of vectors connecting each spot to the every other spot (Fig. 4). In a recognition algorithm, shorter vectors are given higher weight as they are more reliable. Individual identification systems frequently reject many images offered, and require decisions on the false acceptance rate versus true acceptance rate.

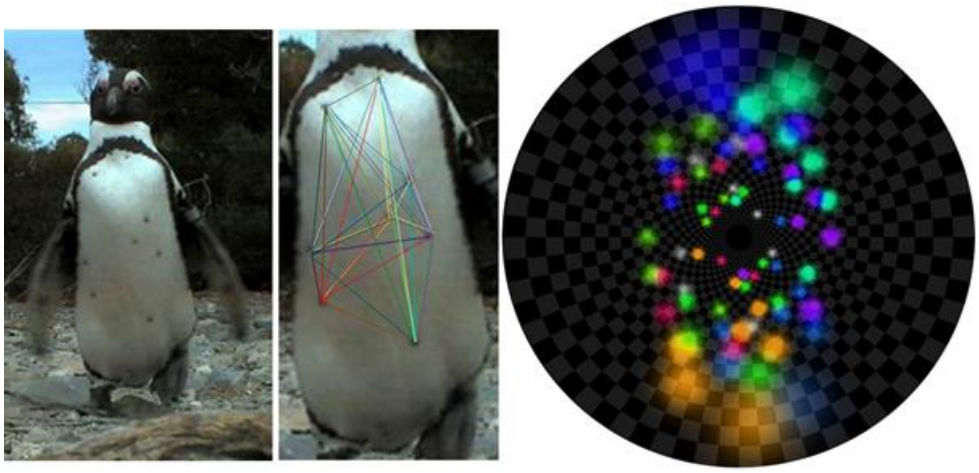


Figure 4: Left) an image of a penguin with unique markings taken by a remote camera system; centre) the chest of the same penguin as identified by a computer vision system and labelled with vectors connecting every pair of spots; right) a representation of the unique identifier for the penguin illustrating the reduced weight given to pairs of spots that have greater separation.

Individual recognition is most efficient in relatively small populations. Once the number of entries in a database becomes large (in the thousands), it becomes increasingly difficult to determine whether images that are not matched come from individuals that are not in the database and hence need to be added, or are poor quality images of individuals already in the database. In addition, the time taken to compare patterns and identify individuals increases with the second power of the number of database entries. These challenges can be in part met by the increases in availability and decreasing cost of cloud computing that provide opportunities to quickly and cheaply process these large datasets.

For some species (such as the *Pygoscelis* penguins monitored through *Penguin Watch*), individual recognition is not possible. However, if a unique pattern is present, and images are of good quality, then time-lapse photography provides a way of non-invasively tracking individuals over time. Furthermore, novel approaches can be adopted when it comes to monitoring unmarked species. For example, individuals may be recognisable during part of their life-cycle, such as prior to fledge – when they are restricted to a nest.

Data Analysis

Survival analysis – following individuals or features of interest

For automated camera networks to be useful in either an ecological or management context, robust demographic parameters must be derived. These should be comparable to (or better than) those obtained by direct observation. Measures such as survivorship (where individuals can be recognised), or breeding success, estimated with larger sample sizes, across larger areas with contrasting environmental conditions or manipulations, can disentangle confounded variables, test hypotheses and permit large-scale monitoring.

Frameworks developed for 'mark-recapture' studies have proved particularly useful in estimating survivorship within camera networks (Lorentzen et al., 2012). Mark-recapture studies mark (or use a unique pattern, such as natural fingerprints (Sherley et al., 2010a) (Fig. 4) or nest position (Youngflesh et al., in review)) individuals in a population and then attempt to recapture those individuals on subsequent occasions to estimate survival. Each of these individuals has a probability of being captured or of dying between capture attempts, along with a probability of recapture. When monitoring nest survival (Fig. 5), we

need to estimate the observation ('capture') probability and survival probabilities of eggs or chicks separately. 'State-space' models have proven useful where the true state (alive or dead) of an individual (chick or egg) for a given nest is modelled as an unobserved state, with the observed state of the individual (captured/not-captured) being the input data for the model at each time step (in our case, time-lapse image). Bayesian state-space models are proven on such data (Black et al., 2018; Youngflesh et al., in review) and allow error estimates to be carried all the way through analyses.

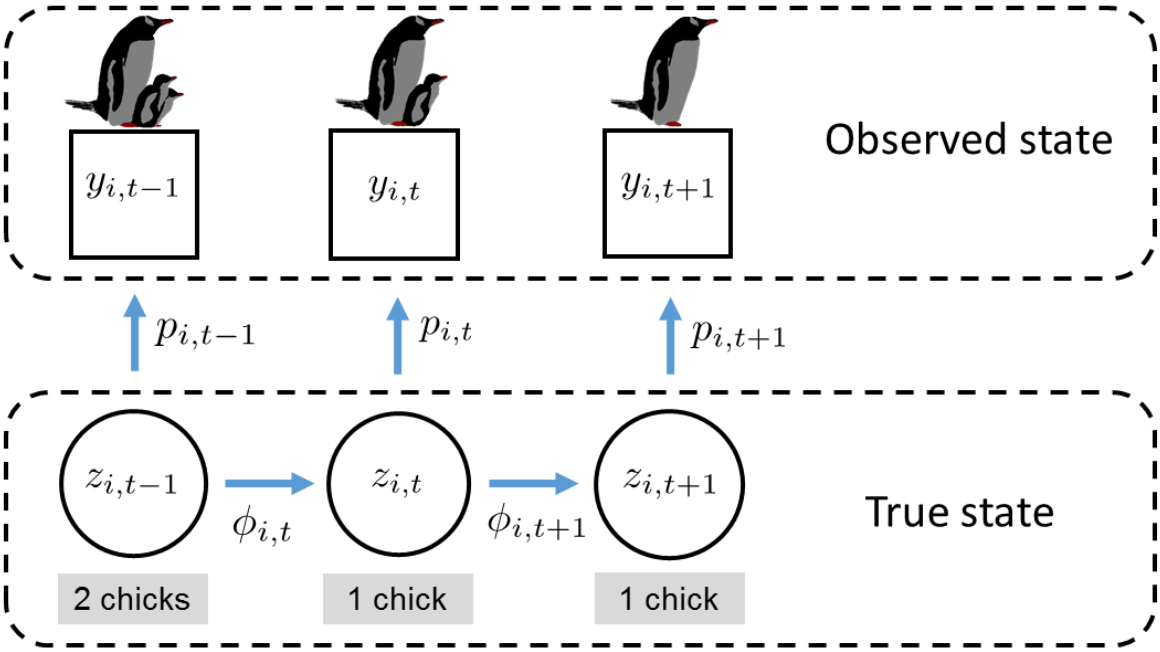


Figure 5: Schematic diagram of a 'virtual' mark-recapture process; $y_{i,t}$ represents the observed state of nest *i* at time step *t*, $z_{i,t}$ represents the true state of a nest, $p_{i,t}$ represents the capture probability of that nest, and $\varphi_{i,t}$ represents the survival probability of a nest from one time step to the next. For example, the final step shows how chick presence is inferred if it is detected in a subsequent observation.

Parameter estimation and pattern recognition

Obtaining metrics of interest, for example the first egg lay date of a penguin or the duration of flowering of an iris, is an important step to analysing large-scale network data. These metrics could be simple and user defined, such as: 'the first arrival of an individual after winter', or 'the date at which any chick is last seen.' Such metrics – e.g. first arrival or peak laying – have been used in direct monitoring programmes for many years (Joint Nature Conservation Committee, 1995). They could also exist in repeating combinations (arrival, followed by courtship, followed by laying), referred to here as 'motifs'. Large datasets from camera networks permit new machine-learning approaches to search for complex patterns and key features. Motif-driven techniques (Fig. 6) look for common patterns to understand: 1) common events occurring across sites; 2) shifts in periodicities over time and 3) co-occurrence of motifs (for example, which motifs happen with a particular time lag).

To find patterns in time series data, we need to first create an approximation, or simplified version, of the data (Fig. 6). Approximations help to remove noise and increase the probability of finding overlapping patterns in a set of series. There are a number of time series approximation techniques available including Fourier transforms, discrete wavelet transformations, and piecewise aggregated or linear approximations. In Fig. 6, we show the symbolic approximation (SAX) approach, which uses alphabetic characters to represent averages of windows (periods of, for instance, three images, three days, five days, or one month) in the time series. The benefit of such an approach is that it enables

search and co-occurrence measures to be used, that have been developed to support DNA and amino acid analyses (Lin et al., 2003; 2007).

These letter transformations are used to discover overlapping string patterns (Fig. 6). The common strings in these sets are used as motif candidates. From a set of motifs, we can understand their occurrence across a dataset, and investigate co-occurrences and lags. For example, a motif would be the pattern of arrival, presence through the breeding period, and departure from a site.

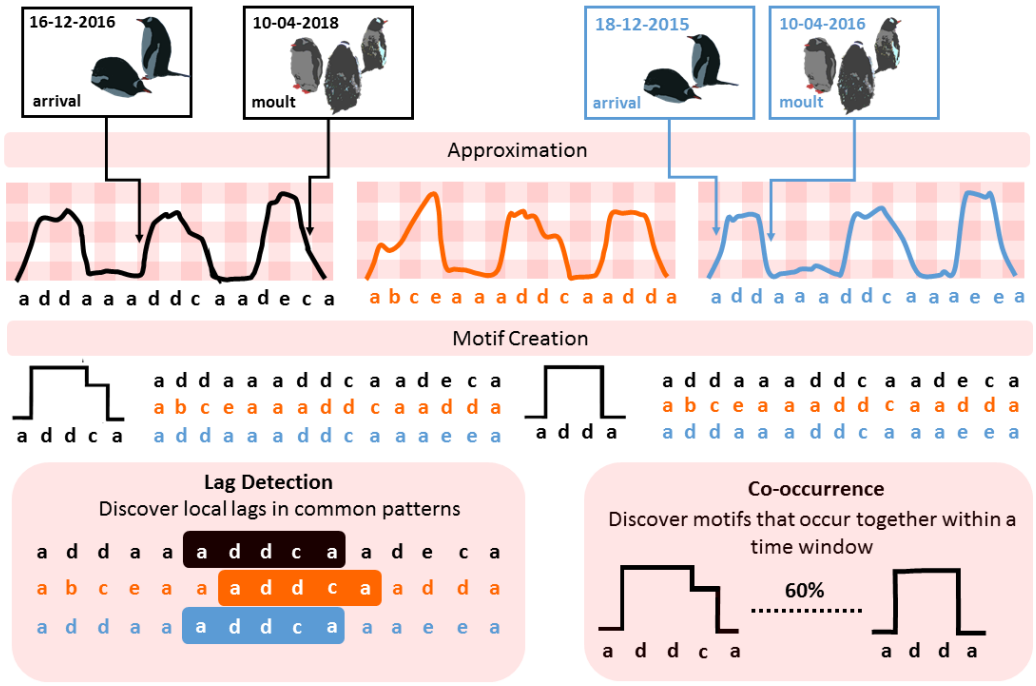


Figure 6: A summary of motif approaches to large-scale data exploration, showing changes in the state of three different colonies through the number of penguins in an image. Three colonies are represented – black, orange and blue. Raw data are approximated or reduced to a sequence of letters based on the number of penguins and classified ‘a-e’ (where ‘a’ is the smallest frequency class with few penguins, and ‘e’ is the largest frequency class where most penguins are present). Common features are identified, detected and compared between sites through lag analyses and co-occurrence.

Discussion

Applications and advantages of time-lapse technologies

Automated data processing requires large amounts of data and training time, yet researchers need to take the leap knowing that these techniques will come and that citizen science can provide a proven first (and ongoing) step.

Historically, the quantity of data produced via camera networks has proved an obstacle for researchers, yet citizen science and computer vision already provide solutions to this issue (Figs. 2, 3 and 4). If constructed carefully – i.e. with an appealing, clearly-explained task – citizen science projects can be used to process hundreds of thousands of data points and actively engage the public in scientific research (Bonney et al., 2009; Dickinson et al., 2012). The use of computer vision is unlikely to render citizen science redundant, rather each can be used where most appropriate. For example, computer vision could be used to analyse 'less interesting' images, and citizen science to validate or train computer vision. Citizen science can also be used to analyse images that computer vision may currently struggle with, such as those containing crowds of individuals (Chapter 3 – Jones et al., in review). Both approaches are generalizable to a wide range of study types – e.g. population censuses (where individuals are 'tagged' or have unique patterns), habitat studies (where animals or plants are identified to species level), and studies on foraging ecology, which could be enhanced with animal borne cameras. Since its launch in September 2014, over six million images have been annotated on the *Penguin Watch* project, a task which would likely have been impossible for an individual research team to complete.

Cameras that can capture time-lapse imagery over periods of interest (e.g. a breeding period) across multiple years allow researchers to address questions relating to seasonal dynamics, breeding phenology, and behaviour. They also facilitate large-scale monitoring studies that may have previously been unmanageable owing to high costs or logistical constraints (e.g. the requirement to frequently return to remote study sites). Co-ordinated monitoring programs over such scales enable studies of ecological phenomena such as climate impacts and other human and non-human induced 'local' impacts.

Time-lapse monitoring can be used to answer a wide range of questions about the study system of interest. These could involve calculating abundance (something easily achievable using citizen science or computer vision), or relate to a metric of interest, such as first arrival (Fig. 1) or nest survival (Fig. 5). For the latter, modelling techniques – such as the mark-recapture model shown in Fig. 5 – and individual recognition (Fig. 4) can be used.

Challenges and future directions

There is huge potential for bespoke camera devices, which are tailored to suit different environments, study species and research fields, and these are likely to become both cheaper and capable of producing higher resolution images in the future. Advances to improve the energy efficiency and operational life of cameras will facilitate higher temporal resolution monitoring, in addition to reducing costs. Miniaturization means we will be able to deploy mobile time-lapse cameras on a wider range of species to investigate behaviour and ecology away from fixed locations or colonies (Kooyman, 2007). Ideally,

cameras will become connected sensors that provide appropriate data and metadata transmitted over great distances at low expense (Kays et al., 2009; Steenweg et al., 2017).

The challenge to the growing researcher community is to determine whether existing solutions can be identified and reapplied to a remote setting, where power supply is limited. For example, the Raspberry Pi microcomputer has a low power consumption, but requires more than can be supplied uninterrupted by most remote power systems. Hence, solutions like 'Instant Detect' (Instant Detect, n.d.), a monitoring system combining audio and visual sensors with camera traps and satellite links, usually have to engineer deep sleep functions. Once cameras are built around microcomputers, some of the computer vision or AI could be integrated into the camera to allow onboard decision making. For instance, dark or blurry images could be discarded prior to transmission, saving money and processing time downstream, or the device could identify the start of a 'period of interest', and capture images at a higher frequency from that point. For example, cameras monitoring nest or den sites could take an image every hour, but trigger an image every minute when patterns of movement consistent with establishment of the breeding site are detected.

No field-ready time-lapse thermal imaging camera exists yet, but cheap laboratory-ready versions (FLIR, 2019) do exist, and could be ruggedized. Paired thermal/visible spectrum imagery may facilitate self-learning in computer vision (and therefore allow rapid development of computer vision to new taxa or questions). Different sensors may already be paired through experimental design; for example, animal borne loggers may be linked with time-lapse nest cameras. One change in experimental design would be to pair cameras on land so that stereo images would allow 3D structure and photogrammetry

(Lavy et al., 2015). This process has already been used to examine nest site preference (McDowall & Lynch, 2017), but could be used on individual animals, for example to examine size class distribution (Durban et al., 2016), or body condition over time (Berger, 2012) and reproductive status (Waite et al., 2007).

Conclusions

The potential use of time-lapse imagery – both in terms of geographic or temporal range and diversity of study species – is vast. Time-lapse technologies enable species or habitats to be monitored over large scales at very little cost, meaning the methodology can be used for new potential research areas with minimal financial risk, by teams who may otherwise lack the resources to do so. Time-lapse camera networks can provide data on site attendance (Black et al., 2018), breeding phenology (Southwell & Emmerson, 2015) and breeding success (Hinke et al., 2018) as well as identify prey and interactions with other species (Gremillet et al., 2010). This can in turn be used to investigate the causes of population declines, demonstrating their use in addressing the 'grand challenges' of ecology and conservation biology. With further developments in camera technology and processing capability, studies covering the entire biogeographic range of species will be possible, and indeed commonplace, providing researchers and decision-makers with the data and tools necessary to implement suitable and effective conservation measures.

Acknowledgements

The authors wish to thank Terra Dawson for the artwork in Figure 1. TH was funded by Darwin Plus and the John Ellerman Foundation for *Penguin Watch*. TH and HJL would like to thank the UK Foreign and Commonwealth Office for Antarctic cameras. FMJ is funded by a NERC studentship. AZ and CA were funded by the EPSRC Programme Grant Seebibyte EP/M013774/1.

Author Contributions

TH, FMJ and MJ conceived the review. AD, AZ, CA, CB, CJL, CS, CY, EM, FMJ, HJL, LE and PB contributed sections according to their expertise. All authors reviewed and approved the manuscript.

References

Arteta, C., Lempitsky, V. & Zisserman, A. (2016). Counting in the Wild. *European Conference on Computer Vision*. 9911 LNCS 483-498.

Arzoumanian, Z., Holmberg, J. & Norman, B. (2005). An astronomical pattern-matching algorithm for computer-aided identification of whale sharks *Rhincodon typus*. *Journal of Applied Ecology 42*(6), 999-1011.

Bater, C. W. et al. (2011). Using digital time-lapse cameras to monitor species-specific understorey and overstorey phenology in support of wildlife habitat assessment. *Environmental Monitoring and Assessment 180*(1-4), 1-13.

Berger, J. (2012). Estimation of body-size traits by photogrammetry in large mammals to inform conservation. *Conservation Biology 26*(5), 769-777.

Black, C., Rey, A. R. & Hart, T. (2017). Peeking into the bleak midwinter: Investigating nonbreeding strategies of Gentoo Penguins using a camera network. *Auk 134*, 520-529.

Black, C., Southwell, C., Emmerson, L., Lunn, D. & Hart, T. (2018). Time-lapse imagery of Adélie reveals differential winter strategies and breeding site occupation. *PLoS ONE, 13*(3) e0193532.

Bonney, R. et al. (2009). Citizen science: a developing tool for expanding science knowledge and scientific literacy. BioScience 59(11) 977-984.

Cook, A. J. & Vaughan, D. G. (2010). Overview of areal changes of the ice shelves on the Antarctic Peninsula over the past 50 years. *Cryosphere 4*(1), 77-98.

Cusack, J. J. et al. (2015). Applying a random encounter model to estimate lion density from camera traps in Serengeti National Park, Tanzania. *The Journal of Wildlife Management 79*(6) 1014-1021.

Dickinson, J. et al. (2012). The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment 10*(6) 291-297.

Dillon, A. & Kelly, M. J. (2008). Ocelot home range, overlap and density: comparing radio telemetry with camera trapping. *Journal of Zoology 275*, 391-398.

Durban, J. W. et al. (2016). Photogrammetry of blue whales with an unmanned hexacopter. *Marine Mammal Science 32*(4), 1510-1515.

FLIR. (2019). FLIR ONE Pro-Series Thermal Imaging Cameras for your Smartphone. URL: https://www.flir.co.uk/flirone/. Retrieved on 05/11/2019.

Fretwell, P. T. et al. (2012). An Emperor Penguin Population Estimate: The First Global, Synoptic Survey of a Species from Space. *PLoS ONE 7*(4) e33751.

Gil-Sanchez, J. M. et al. (2011). The use of camera trapping for estimating Iberian lynx (*Lynx pardinus*) home ranges. *European Journal of Wildlife Research 57*(6), 1203-1211.

Graham, E. A., Riordan, E. C., Yuen, E. M., Estrin, D. & Rundel, P. W. (2010). Public internet-connected cameras used as a cross-continental ground-based plant phenology monitoring system. *Global Change Biology 16*, 3014-3023.

Gremillet, D. et al. (2010). Seabirds, fisheries, and cameras. *Frontiers in Ecology and the Environment 8*, 401-402.

Happy Whale. (2019). Your Whales as Individuals. URL: www.happywhale.com. Retrieved on 22/09/2019.

Harris, M. P. (1982). Promiscuity in the Shag as Shown by Time-Lapse Photography. *Bird Study 29*, 149-154.

Hegerl, C. et al. (2017). Using camera trap data to assess the impact of bushmeat hunting on forest mammals in Tanzania. *Oryx 51*(1), 87-97.

Hinke, J. T. et al. (2018). Estimating nest-level phenology and reproductive success of colonial seabirds using time-lapse cameras. *Methods in Ecology and Evolution 9*(8) 1853-1863.

Huffeldt, N. P. & Merkel, F. R. (2013). Remote Time-lapse Photography as a Monitoring Tool for Colonial Breeding Seabirds: A Case Study Using Thick-billed Murres (*Uria lomvia*). *WATERBIRDS: The International Journal of Waterbird Biology*, 330-341.

Instant Detect. (n.d.). Conservation Technology – Instant Detect. URL: https://www.zsl.org/conservation/how-we-work/conservation-technology/instant-detect. Retrieved on 06/10/2019.

Joint Nature Conservation Committee (Great Britain), Royal Society for the Protection of Birds. Institute of Terrestrial Ecology & Seabird Group. Joint Nature Conservation Committee, Peterborough. (1995).

Jones, F. M. et al. (2018). Time-lapse imagery and volunteer classifications from the Zooniverse Penguin Watch project. *Scientific Data 5*(1) 180124.

Jones, F.M., Arteta, C., Zisserman, A., Lempitsky, V., Lintott, C.J., Hart, T. (in review). Processing citizen science- and machine- annotated time-lapse imagery for biologically meaningful metrics. *Scientific Data*.

Kays, R. et al. (2009). Camera traps as sensor networks for monitoring animal communities. 2009 IEEE 34th Conference on Local Computer Networks.

Kooyman, G. (2007). Animal-borne instrumentation systems and the animals that bear them: then (1939) and now (2007). *Marine Technology Society Journal 41*(4), 6-8.

LaRue, M., Kooyman, G., Lynch, H. L. & Fretwell, P. (2014). Emigration in emperor penguins: implications for interpretation of long-term studies. *Ecography* 38(2) 114-120.

Lavy, A. et al. (2015). A quick, easy and non-intrusive method for underwater volume and surface area evaluation of benthic organisms by 3D computer modelling. *Methods in Ecology and Evolution 6*, 521-531.

Legg, C. J. & Nagy, L. (2006). Why most conservation monitoring is, but need not be, a waste of time. *Journal of Environmental Management 78*(2), 194-199.

Lempitsky, V. & Zisserman, A. (2010). Learning to Count Objects in Images. *Advances in Neural Information Processing Systems*, 1324-1332.

Lenzano, M. G., Lannutti, E., Toth, C., Lenzano, L. & Lovecchio, A. (2014). Assessment of Ice-Dam Collapse by Time-Lapse Photos at the Perito Moreno Glacier, Argentina. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-1*. ISPRS Technical Commission I Symposium 40-1, 211-217.

Lin, J., Keogh, E., Lonardi, S. & Chiu, B. (2003). A symbolic representation of time series, with implications for streaming algorithms. Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery, 2-11 (ACM).

Lin, J., Keogh, E., Wei, L. & Lonardi, S. (2007). Experiencing SAX: a novel symbolic representation of time series. *Data Mining and Knowledge Discovery 15*(2), 107-144.

Lorentzen, E., Choquet, R. & Steen, H. (2012). Modelling state uncertainty with photo series data for the estimation of breeding success in a cliff-nesting seabird. *Journal of Ornithology 152*(2), 477-483.

Lucas, T. C. D., Moorcroft, E. A., Freeman, R., Rowcliffe, J. M. & Jones, K. E. (2015). A generalised random encounter model for estimating animal density with remote sensor data. *Methods in Ecology and Evolution 6*(5), 500-509.

Mason, W. & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods 44*(1), 1-23.

McDowall, P. & Lynch, H. J. (2017). Ultra-Fine Scale Spatially-Integrated Mapping of Habitat and Occupancy Using Structure-From-Motion. *PLoS ONE 12*, e0166773.

Muybridge, E. (1887). Animal locomotion. Da Capo Press.

Newbery, K. & Southwell, C. (2009). An automated camera system for remote monitoring in polar environments. *Cold Regions Science and Technology 55*(1), 47-51.

Penney, R. (1968). Territorial and social behaviour in the Adélie Penguin. *Antarctic Bird Studies 12*, 83-131.

Richter, S. et al. (2018). A remote-controlled observatory for behavioural and ecological research: A case study on emperor penguins. *Methods in Ecology and Evolution 9*(5), 1168-1178.

Roberts, S. J., Lee, H. J., Drake, K. A. & Dawkins, M. S. (2011). Prediction of feather damage in laying hens using optical flows and Markov models. *Journal of the Royal Society Interface 8*(57), 489-499.

Sherley, R. B., Burghardt, T., Barham, P. J. & Cuthill, I. C. in 7th International Penguin Conference (IPC 2010). 116.

Sherley, R. B., Burghardt, T., Barham, P. J., Campbell, N. & Cuthill, I. C. (2010). Spotting the difference: towards fully-automated population monitoring of African penguins *Spheniscus demersus*. *Endangered Species Research 11*(2), 101-111.

Southwell, C. & Emmerson, L. (2015). Remotely-operating camera network expands Antarctic seabird observations of key breeding parameters for ecosystem monitoring and management. *Journal for Nature Conservation 23*, 1-8.

Steenweg, R. et al. (2017). Scaling-up camera traps: monitoring the planet's biodiversity with networks of remote sensors. *Frontiers in Ecology and the Environment 15*(1), 26-34.

Swanson, A. et al. (2015). Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Scientific Data 2* 150026.

Waite, J. N., Schrader, W. J., Mellish, J.-A. E., Horning, M. (2007). Three-dimensional photogrammetry as a tool for estimating morphometrics and body mass of Steller sea lions (*Eumetopias jubatus*). *Canadian Journal of Fisheries and Aquatic Science 64*(2), 296-303.

Weinstein, B. G. (2018). A computer vision for animal ecology. *Journal of Animal Ecology 87*(3) 533-545.

Weller, M.W. & Derksen, D.V. (1972). Use of time-lapse photography to study nesting activities of birds. *The Auk 89*(1) 196-200.

Wright, D. E. et al. (2017). A transient search using combined human and machine classifications. *Monthly Notices of the Royal Astronomical Society 472*(2), 1315-1323.

Zhang, X. et al. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of the Environment 84*(3), 471-475.

Zhang, X., Friedl, M. A., Schaaf, C. B. & Strahler, A. H. (2004). Climate controls on vegetation phenological patterns in northern mid-and high latitudes inferred from MODIS data. *Global Change Biology 10*(7), 1133-1145.

Youngflesh, C., Jones, F.M., Lynch, H.J., Arthur, J., Macháčková, Z., Torsey, H.R. & Hart, T. (in review). Large-scale assessment of intra- and inter-annual nesting success using a remote camera network.

Appendix 2: Penguin Watch Cameras

Region	Location	Species	Camera code	Latitude	Longitude	First season	Status	Data owner
Antarctic Peninsula	Andresen Island	Adélie	ANDRa	66.877	66.660	2015/16	Active	TH
	Booth Island	Gentoo/Chinstrap	PCHAc	65.067	64.024	2011/12	Removed	TH
		Gentoo	BOOTa	65.067	64.037	2011/12	Active	TH
		Gentoo	BOOTb	65.067	64.022	2012/13	Active	TH
		Chinstrap	BOOTc	65.067	64.024	2014/15	Active	TH
		Gentoo (F)	BOOTd	65.067	64.022	2014/15	Active	TH
	Brown Bluff	Gentoo	BROWa	63.522	56.889	2010/11	Removed	TH
		Gentoo	BROWb	63.521	56.889	2014/15	Active	TH
		Adélie	BROWc	63.520	56.893	2014/15	Active	TH
		Adélie	BROWd	63.520	56.893	2014/15	Active	TH
	Cuverville	Gentoo	CUVEa	64.682	62.621	2011/12	Active	TH
		Gentoo	CUVEb	64.682	62.621	2013/14	Active	TH
	Damoy Point	Gentoo	DAMOa	64.818	63.519	2013/14	Active	TH
	Danco Island	Gentoo	DANCa	64.728	62.597	2011/12	Removed	TH
		Gentoo	DANCb	64.728	62.597	2011/12	Active	TH
	Gaston Islands	Chinstrap	GASTa	64.488	61.836	2014/15	Active	TH
	George's Point	Gentoo	GEORa	64.670	62.670	2012/13	Active	TH
		Gentoo	GEORb	64.670	62.670	2013/16	Active	TH
		Gentoo	GEORc	64.670	62.670	2014/15	Active	TH
		Chinstrap	GEORd	64.670	62.670	2015/16	Active	TH
		Chinstrap	GEORe	64.670	62.670	2018/19	Active	TH
	Heroina, Danger Islands	Adélie	HEROa	63.400	54.600	2015/16	Active	TH
		Adélie (F)	HEROb	63.400	54.600	2015/16	Active	TH
	Jougla Point	Gentoo (Shag)	JOUGa	64.826	63.493	2013/14	Active	TH

	Mikkelson Harbour	Gentoo	MIKKa	63.902	60.790	2014/15	Active	TH
	Moot Point	Gentoo	MOOT	65.204	64.075	2014/15	Active	TH
	Neko Harbour	Gentoo	NEKOa	64.844	62.529	2011/12	Active	TH
		Gentoo	NEKOb	64.845	62.530	2011/12	Removed	TH
		Gentoo	NEKOc	64.843	62.528	2012/13	Active	TH
		Gentoo (F)	NEKOd	64.844	63.298	2014/15	Active	TH
	Orne Harbour	Chinstrap	SPIGa	64.631	62.555	2012/13	Active	TH
		Chinstrap	ORNEb	64.631	62.555	Unknown	Unknown	TH
	Petermann Island	Gentoo	PETEa	65.172	64.137	2011/12	Active	TH
		Gentoo	PETEb	65.171	64.136	2011/12	Removed	TH
		Gentoo	PETEc	65.171	64.137	2012/13	Active	TH
		Adélie	PETEd	65.172	64.136	2012/13	Active	TH
		Adélie & Gentoo	PETEe	65.172	64.136	2012/13	Active	TH
		Gentoo (F)	PETEf	65.173	64.138	2012/13	Active	TH
	Port Lockroy, Goudier Island	Gentoo	LOCKa	64.827	63.493	2010/11	Active	TH
		Gentoo	LOCKb	64.827	63.493	2012/13	Active	TH
		Gentoo	LOCKc	64.827	63.493	2014/15	Active	TH
	Tetrad Islands	Chinstrap	TETRa	63.917	60.733	Unknown	Active	TH
	Yalour Islands	Adélie	YALOa	65.235	64.159	2011/12	Active	TH
		Adélie	YALOb	65.233	64.161	2013/14	Active	TH
Falkland Islands	Beaver Island	Gentoo	BEAVa	51.839	61.278	Unknown	Active	TH
	Bluff Cove	Gentoo	BLUFa	51.758	58.069	2014/15	Active	TH
		King	BLUFb	51.758	58.069	2014/15	Active	TH
	Saunders Island	Gentoo	SAUNa	51.306	60.241	2012/13	Active	TH
		Kings	SAUNb	53.306	60.241	Unknown	Removed	TH
		Gentoo (F)	SAUNc	53.306	60.241	Unknown	Active	TH
South Georgia	Cooper Bay	Macaroni	COOPa	54.782	35.803	2014/15	Active	TH
		Macaroni	COOPb	54.782	35.803	2014/15	Active	TH
		Gentoo	COOPc	54.783	35.802	2014/15	Active	TH

		Gentoo	COOPd	54.783	35.802	2014/15	Active	TH
	Elsehul	Seals	ELSEa	54.030	37.965	2015/15	Unknown	TH
	Fortuna	Seals	FORTb	54.150	36.786	2014/15	Active	TH/CF
	Godthul	Gentoo	GODHa	54.283	36.300	2014/15	Active	TH
	Gold Harbour	King	GOLDa	54.620	35.949	2014/15	Active	TH
	Hercules Point	Macaroni	HERCa	54.166	36.667	2014/15	Active	TH
	Maiviken	Gentoo	MAIVb	54.249	36.490	2012/13	Active	TH
		Gentoo	MAIVc	54.249	36.490	2012/13	Active	TH
		Gentoo	MAIVd	54.249	36.490	2012/13	Active	TH
		Gentoo	MAIVe	54.249	36.490	2014/15	Active	TH
	Ocean Harbour	Gentoo	OCEAa	54.341	36.247	2014/15	Removed	TH
		Gentoo	OCEAb	54.341	36.247	2014/15	Removed	TH
	Right Whale Bay	Seals	RIGHa	54.105	37.691	2014/15	Active	TH
		King	RIGHb	54.105	37.691	2014/15	Active	TH
	Rookery Bay	Macaroni	ROOKa				Active	TH
	Salisbury Plain	King	SALIa	54.063	37.329	2010/11	Active	TH
		King	SALIb	54.063	37.329	2014/15	Active	TH
		Gentoo	SALIc	54.063	37.329	2018/19	Active	TH
	Sandebugden	Gentoo	SANDa	54.324	36.340	2014/15	Removed	TH
	St Andrew's Bay	King	STANa	54.442	36.190	2014/15	Active	TH
		King	STANb	54.442	36.190	2014/15	Active	TH
South Orkney Islands	Signy Island	Chinstrap	SIGNa	60.737	45.657	2010/11	Active	MD
South Sandwich Islands	Saunders Island	Chinstrap	SSISa	57.808	26.398	2013/14	Active	TH
		Chinstrap	SSISb	57.808	26.398	2013/14	Active	TH
		Adélie	SSISc	57.805	26.420	2014/15	Active	TH
South Shetland Islands	Aitcho Barrientos Island	Gentoo	AITCa	62.407	59.745	2010/11	Removed	TH
		Gentoo	AITCb	62.406	59.745	2013/14	Active	TH
		Gentoo	AITCc	62.408	59.750	2013/14	Active	TH
		Chinstrap	AITCd	62.406	59.740	2013/14	Active	TH

		Gentoo (F)	AITCe	62.406	59.745	2014/15	Unknown	TH
	Baily Head, Deception Island	Chinstrap	BAILa	62.964	60.504	2012/13	Active	TH
		Chinstrap	BAILb	62.964	60.504	2015/16	Active	TH
	Copacabana, King George Island	Adélie	APN	62.176	58.446	2014/15	Active	JH
		Adélie	BOBO	62.175	58.446	2014/15	Active	JH
		Gentoo	IBC_E	62.174	58.447	2013/14	Active	JH
		Gentoo	IBC_W	62.174	58.447	2013/14	Active	JH
		Gentoo	GH	62.175	58.448	2014/15	Active	JH
		Adélie	CRN	62.176	58.449	2014/15	Active	JH
		Adélie	CRS_E	62.177	58.449	2013/14	Active	JH
		Adélie	CRS_W	62.177	58.449	2013/14	Active	JH
		Gentoo	CB4	62.177	58.446	2013/14	Active	JH
		Gentoo	CB4A	62.177	58.446	2013/14	Active	JH
	Half Moon Island	Chinstrap	HALFa	62.595	59.899	2010/11	Removed	TH
		Chinstrap	HALFb	62.595	59.899	2012/13	Active	TH
		Chinstrap	HALFc	62.596	59.898	2012/13	Active	TH
	Point Wild, Elephant Island	Chinstrap	WILDa	61.099	54.859	2011/12	Active	TH

Appendix 3: Fieldwork Permit Details

Area	Primary Permit Holder	Permit Number/Ethical Review Board
2009/2010 Season		
Antarctica	Tom Hart	S3-5/2009 & S7-2/2009
Ethics		Zoological Society of London (ZSL)
2010/2011 Season		
Antarctica (including the South Shetland and South Orkney Islands)	Tom Hart	S3-01/2010 &S7-5/2010
Falkland Islands		R10/ 2011
South Georgia and the South Sandwich Islands		Letter dated 14th December 2010
Ethics		ZSL
2011/2012 Season		
Antarctica (including the South Shetland and South Orkney Islands)	Tom Hart	S3-05/2011 &S7-08/2011
Falkland Islands		
South Georgia and the South Sandwich Islands		Letter dated 2nd November 2011
Ethics		ZSL
2012/2013 Season		
Antarctica (including the South Shetland and South Orkney Islands)	Tom Hart	S3-04/2012 & S7-06/2012
Falkland Islands		R22/2012
South Georgia and the South Sandwich Islands		None
Ethics		Oxford Zoology Local Ethical Review Committee (LERC)
2013/2014 Season		
Antarctica (including the South Shetland and South Orkney Islands)	Tom Hart	S3-23/2013 & S7 28/2013
Falkland Islands		R22/2013
South Georgia and the South Sandwich Islands		SCI/2013/009
Ethics		Oxford Zoology LERC

2014/2015 Season		
Antarctica (including the South Shetland and South Orkney Islands)	Tom Hart	S3 04/2014 & S7 03-2014
Falkland Islands		R16.2014
South Georgia and the South Sandwich Islands		SCI-2014-017
Ethics		Oxford Zoology LERC
2015/2016 Season		
Antarctica (including the South Shetland and South Orkney Islands)	Tom Hart	S3 04/2014 & S7 03-2014
Antarctica (Danger Islands and Elephant Island)	Stephanie Jenouvrier	ACA 2016-011
Falkland Islands	Tom Hart	R15 2014
South Georgia and the South Sandwich Islands		RAP-2015-018_ammended
Ethics		Oxford Animal Welfare and Ethical Review Body (AWERB)
2016-2017 Season		
Antarctica (including the South Shetland and South Orkney Islands)	Tom Hart	33/2016 & 34/2016
Falkland Islands		
South Georgia and the South Sandwich Islands		RAP-2016-035
Ethics		Oxford AWERB
2017-2018 Season		
Antarctica (including the South Shetland and South Orkney Islands)	Tom Hart	37/2017
Falkland Islands		R04 2017
South Georgia and the South Sandwich Islands		RAP-2017-034
Ethics		Oxford AWERB
2018-2019 Season		
Antarctica (including the South Shetland and South Orkney Islands)	Tom Hart	8-2018
Falkland Islands		
South Georgia and the South Sandwich Islands		RAP-2018-055 RAP-2018-038_amended
Ethics		Oxford AWERB

Appendix 4: Chapter 2 – Metadata Tables

Site (camera location)	Geographic coordinates	Camera name	Image set (name in repository)	File information and camera settings
Damoy Point, Weincke Island, Antarctic Peninsula	64.82° S, 63.49° W	DAMOa	DAMOa2014a	No. of files: 407 Time range: 26/12/2013 1300 – 22/01/2014 1400 Settings: 0700 – 2100; hourly
George's Point, Rongé Island on the Errera Channel, Antarctic Peninsula	64.67° S, 62.67° W	GEORa	GEORa2013	No. of files: 4403 Time range: 22/12/2012 1000 – 24/12/2013 0800 Settings: 0800 – 1900; hourly
Half Moon Island, South Shetland Islands	62.60° S, 59.90° W	HALFb	HALFb2013a	No. of files: 306 Time range: 15/12/2012 1700 – 08/01/2013 1000 Settings: 0700 – 1900; hourly
		HALFc	HALFc2013a	No. of files: 283 Time range: 21/12/2012 1700 – 09/01/2013 1400 Settings: 0700 – 2100; hourly
Port Lockroy, Wiencke Island, Antarctic Peninsula	64.82° S, 63.49° W	LOCKb	LOCKb2013	No. of files: 2372 Time range: 13/12/2012 1100 – 31/05/2013 1600 Settings: 0700 – 2000; hourly
Maiviken, South Georgia	54.24° S, 36.50° W	MAIVb	MAIVb2012a	No. of files: 660 Time range: 13/10/2012 0900 – 03/01/2013 1300 Settings: 0900 – 1100, 1300 – 1700; hourly
			MAIVb2013a	No. of files: 3878 Time range: 03/01/2013 1400 – 28/10/2013 1700

				Settings: 0800 – 2000; hourly
			MAIVb2013c	No. of files: 137 Time range: 29/10/2013 0800 – 08/11/2013 1400 Settings: 0800 – 2000; hourly
		MAIVc	MAIVc2013	No. of files: 4661 Time range: 15/10/2012 0900 – 08/11/2013 1300 Settings: 0900 – 1100, 1300 – 1700; hourly
Neko Harbour, Andvord Bay, Antarctic Peninsula	64.86° S, 62.52° W	NEKOa	NEKOa2012a	No. of files: 1608 Time range: 03/03/2012 1000 – 25/11/2012 1500 Settings: 1000 – 1500; hourly
			NEKOa2013	No. of files: 3552 Time range: 26/11/2012 1000 – 25/12/2013 1500 Settings: 0800 – 1600; hourly
			NEKOa2014a	No. of files: 239 Time range: 25/12/2013 1600 – 13/01/2014 0700 Settings: 0700 – 1900; hourly
		NEKOb	NEKOb2013	No. of files: 6449 Time range: 02/03/2012 1600 – 26/11/2012 0800 Settings: 0000 – 1200; hourly
		NEKOc	NEKOc2013	No. of files: 4415 Time range: 14/12/2012 1600 – 19/11/2013 1000 Settings: 0700 – 1900; hourly
			NEKOc2014b	No. of files: 296

				Time range: 25/12/2013 1600 – 13/01/2014 0700 Settings: 0600-2100; hourly
Petermann Island, Antarctic Peninsula	65.17° S, 64.14° W	PETEc	PETEc2013	No. of files: 506 Time range: 03/12/2012 1600 – 11/01/2013 1400 Settings: 0800 – 1100, 1300 – 2100; hourly
			PETEc2014	No. of files: 10,732 Time range: 11/01/2013 1430 – 04/01/2014 1000 Settings: 0700 – 2130; half hourly
		PETEe	PETEe2013	No. of files: 3874 Time range: 12/12/2012 1600 – 04/01/2014 0900 Settings: 0900 – 1800; hourly
		PETEf	PETEf2014a	No. of files: 8271 Time range: 24/12/2012 1300 – 04/01/2014 1000 Settings: 0100 – 1100, 1300 – 2300; hourly
Spigot Peak, Orne Harbour, Antarctic Peninsula	64.63° S, 62.57° W	SPIGa	SPIGa2012a	No. of files: 390 Time range: 25/11/2012 1700 – 21/12/2012 1600 Settings: 0700 -2100; hourly
			SPIGa2013b	No. of files: 5221 Time range: 09/01/2013 1500 – 23/12/2013 1500 Settings: 0700 – 2100; hourly
			SPIGa2014	No. of files: 410 Time range: 23/12/2013 1600 – 21/01/2014 2000 Settings: 0700 – 2000; hourly

Table 4: Raw images – metadata and file information. Metadata and file information associated with each of the 63,070 'raw images' included in the Dryad Digital Repository (Data Citation 1). These images are stored under the folders: DAMO-MAIV, NEKO, PETE.1, PETE.2 and SPIG. Images retain their original dimensions (either 1920x1080 pixels or 2048x1536 pixels), but have been renamed. See Chapter 2 – 'Code Availability' for a link to the script (R v 3.5.3; R Core Team, 2016) used to rename and resize images.

Camera name	Corresponding image set (see Table 4)	File name (in repository)
DAMOa	DAMOa2014a	DAMOa2014a_concl.csv
GEORa	GEORa2013	GEORa2013a_concl.csv
		GEORa2013b_concl.csv
HALFb	HALFb2013a	HALFb2013a_concl.csv
HALFc	HALFc2013a	HALFc2013a_concl.csv
LOCKb	LOCKb2013	LOCKb2013a_concl.csv
		LOCKb2013b_concl.csv
MAIVb	MAIVb2012a	MAIVb2012a_concl.csv
	MAIVb2013a	MAIVb2013a_concl.csv
	MAIVb2013c	MAIVb2013c_concl.csv
MAIVc	MAIVc2013	MAIVc2013_concl.csv
NEKOa	NEKOa2012a	NEKOa2012a_concl.csv
	NEKOa2013	NEKOa2013a_concl.csv
		NEKOa2013b_concl.csv
		NEKOa2013c_concl.csv
	NEKOa2014a	NEKOa2014a_concl.csv
NEKOb	NEKOb2013	NEKOb2013_concl.csv
NEKOc	NEKOc2013	NEKOc2013a_concl.csv
		NEKOc2013b_concl.csv
		NEKOc2013c_concl.csv
	NEKOc2014b	NEKOc2014b_concl.csv
PETEc	PETEc2013	PETEc2013a_concl.csv
		PETEc2013b_concl.csv
	PETEc2014	PETEc2014a_concl.csv
		PETEc2014b_concl.csv
PETEe	PETEe2013	PETEe2013a_concl.csv
		PETEe2013b_concl.csv
PETEf	PETEf2014a	PETEf2014a_concl.csv
SPIGa	SPIGa2012a	SPIGa2012a_concl.csv
	SPIGa2013b	SPIGa2013b_concl.csv
	SPIGa2014	SPIGa2014a_concl.csv
		SPIGa2014b_concl.csv

Table 5: *Penguin Watch* 'consensus click' data. List of 'PW Consensus Click Data' files stored in the Dryad Digital Repository (Data Citation 1), and the cameras and image sets they are associated with. 'Consensus clicks' are produced via a clustering algorithm, as discussed in 'Data and Analytical Methods'. The clustering algorithm code (Python v 2.7; Python Software Foundation, 2019) is publically available – see Chapter 2, 'Code Availability'.

Camera name	Corresponding image set (see Table 4)	File name (in repository)	Number of unique annotators
DAMOa	DAMOa2014a	DAMOa2014a_metadata.csv	3042
GEORa	GEORa2013	GEORa2013a_metadata.csv	1810
		GEORa2013b_metadata.csv	13,353
HALFb	HALFb2013a	HALFb2013a_metadata.csv	2433
HALFc	HALFc2013a	HALFc2013a_metadata.csv	2211
LOCKb	LOCKb2013	LOCKb2013a_metadata.csv	1286
		LOCKb2013b_metadata.csv	8959
MAIVb	MAIVb2012a	MAIVb2012a_metadata.csv	3977
	MAIVb2013a	MAIVb2013a_metadata.csv	11,367
	MAIVb2013c	MAIVb2013c_metadata.csv	1177
MAIVc	MAIVc2013	MAIVc2013_metadata.csv	13,543
NEKOa	NEKOa2012a	NEKOa2012a_metadata.csv	6874
	NEKOa2013	NEKOa2013a_metadata.csv	705
		NEKOa2013b_metadata.csv	11,068
		NEKOa2013c_metadata.csv	2411
	NEKOa2014a	NEKOa2014a_metadata.csv	1831
NEKOb	NEKOb2013	NEKOb2013_metadata.csv	15,789
NEKOc	NEKOc2013	NEKOc2013a_metadata.csv	28
		NEKOc2013b_metadata.csv	1019
		NEKOc2013c_metadata.csv	13,523
	NEKOc2014b	NEKOc2014b_metadata.csv	2212
PETEc	PETEc2013	PETEc2013a_metadata.csv	2033
		PETEc2013b_metadata.csv	1926
	PETEc2014	PETEc2014a_metadata.csv	21,009
		PETEc2014b_metadata.csv	4492
PETEe	PETEe2013	PETEe2013a_metadata.csv	981
		PETEe2013b_metadata.csv	12,469
PETEf	PETEf2014a	PETEf2014a_metadata.csv	18,461
SPIGa	SPIGa2012a	SPIGa2012a_metadata.csv	2419
	SPIGa2013b	SPIGa2013b_metadata.csv	12,435
	SPIGa2014	SPIGa2014a_metadata.csv	1658
		SPIGa2014b_metadata.csv	1235

Table 6: *Penguin Watch* anonymised raw classifications and metadata. List of 'PW Anonymised Raw Classifications and Metadata' files stored in the Dryad Digital Repository (Data Citation 1), and their associated cameras/image files. The number of unique *Penguin Watch* volunteers that contributed classifications to each file are also shown. See Chapter 2, 'Code Availability' for the script (R v 3.5.3; R Core Team, 2016) used to extract metadata.

References

Python Software Foundation. (2019). *Python Language Reference*. Retrieved from http://www.python.org

R Core Team. (2016). *R: A language and environment for statistical computing.* Retrieved from https://www.r-project.org/

Data Citation 1:

Jones, F.M., Allen, C., Arteta, C., Arthur, J., Black, C., Emmerson, L.M., Freeman, R., Hines, G., Lintott, C.J., Macháčková, Z., Miller, G., Simpson, R., Southwell, C. Torsey, H.R., Zisserman, A. & Hart, T. (2018). Data from: Time-lapse imagery and volunteer classifications from the Zooniverse Penguin Watch project. Dryad Digital Repository. DOI:10.5061/dryad.vv36g

Appendix 5: Chapter 3 – Metadata Tables

Site (camera location)	Camera name	Image set (name in Data Citation 1)	Time-range	*Kraken File* (.csv)	*Narwhal File* (.csv)
Damoy Point, Weincke Island, Antarctic Peninsula; 64.82° S, 63.49° W	DAMOa	DAMOa2014a	26/12/2013 1300 to 22/01/2014 1400 (0700 – 2100; hourly)	DAMOa_kraken	DAMOa_narwhal
George's Point, Rongé Island on the Errera Channel, Antarctic Peninsula; 64.67° S, 62.67° W	GEORa	GEORa2013	22/12/2012 1000 to 24/12/2013 0800 (0800 – 1900; hourly)	GEORa_kraken	GEORa_narwhal
Half Moon Island, South Shetland Islands; 62.60° S, 59.90° W	HALFb	HALFb2013a	15/12/2012 1700 to 08/01/2013 1000 (0700 – 1900; hourly)	HALFb_kraken	HALFb_narwhal
	HALFc	HALFc2013a	21/12/2012 1700 to 09/01/2013 1400 (0700 – 2100; hourly)	HALFc_kraken	HALFc_narwhal
Port Lockroy, Wiencke Island, Antarctic Peninsula; 64.82° S, 63.49° W	LOCKb	LOCKb2013	13/12/2012 1100 to 31/05/2013 1600 (0700 – 2000; hourly)	LOCKb_kraken	LOCKb_narwhal
Maiviken, South Georgia; 54.24° S, 36.50° W	MAIVb	MAIVb2012a	13/10/2012 0900 to 03/01/2013 1300	MAIVb_kraken	MAIVb_narwhal

			(0900 – 1100, 1300 – 1700; hourly)		
		MAIVb2013a	03/01/2013 1400 to 28/10/2013 1700 (0800 – 2000; hourly)		
		MAIVb2013c	29/10/2013 0800 to 08/11/2013 1400 (0800 – 2000; hourly)		
	MAIVc	MAIVc2013	15/10/2012 0900 to 08/11/2013 1300 (0900 – 1100, 1300 – 1700; hourly)	MAIVc_kraken	MAIVc_narwhal
Neko Harbour, Andvord Bay, Antarctic Peninsula; 64.86° S, 62.52° W	NEKOa	NEKOa2012a	03/03/2012 1000 to 25/11/2012 1500 (1000 – 1500; hourly)	NEKOa_kraken	NEKOa_narwhal
		NEKOa2013	26/11/2012 1000 to 25/12/2013 1500 (0800 – 1600; hourly)		
		NEKOa2014a	25/12/2013 1600 to 13/01/2014 0700 (0700 – 1900; hourly)		
	NEKOb	NEKOb2013	02/03/2012 1600 to 26/11/2012 0800	NEKOb_kraken	NEKOb_narwhal

			(0000 – 1200; hourly)		
	NEKOc	NEKOc2013	14/12/2012 1600 to 19/11/2013 1000 (0700 – 1900; hourly)	NEKOc_kraken	NEKOc_narwhal
		NEKOc2014b	(25/12/2013 1600 – 13/01/2014 0700) (0600-2100; hourly)		
Petermann Island, Antarctic Peninsula; 65.17° S, 64.14° W	PETEc	PETEc2013	03/12/2012 1600 to 11/01/2013 1400 (0800 – 1100, 1300 – 2100; hourly)	PETEc_kraken	PETEc_narwhal
		PETEc2014	11/01/2013 1430 to 04/01/2014 1000 (0700 – 2130; half hourly)		
	PETEe	PETEe2013	12/12/2012 1600 to 04/01/2014 0900 (0900 – 1800; hourly)	PETEe_kraken	PETEe_narwhal
	PETEf	PETEf2014a	24/12/2012 1300 to 04/01/2014 1000 (0100 – 1100, 1300 – 2300; hourly)	PETEf_kraken	PETEf_narwhal
Spigot Peak, Orne Harbour, Antarctic Peninsula; 64.63° S, 62.57° W	SPIGa	SPIGa2012a	25/11/2012 1700 to 21/12/2012 1600 (0700 -2100; hourly)	SPIGa_kraken	SPIGa_narwhal

		SPIGa2013b	09/01/2013 1500 to 23/12/2013 1500 (0700 – 2100; hourly)		
		SPIGa2014	23/12/2013 1600 to 21/01/2014 2000 (0700 – 2000; hourly)		

Table 2: List of *Kraken Files* and *Narwhal Files* stored in the Dryad Digital Repository (Data Citation 2), alongside their associated camera information (site name and location, camera name, image set name, and time-lapse camera settings and time-range). These correspond to the image sets described in (Jones et al., 2018 and 2019) – see Appendix 4, Table 4. See Chapter 3 – 'Code Availability' for links to the *Kraken* and *Narwhal* Scripts (R v 3.6.0; R Core Team, 2019) used to create these files.

Camera name	*Narwhal File* (.csv)	*Narwhal Plots*	*Penguin Watch Narwhal Animation*
DAMOa	DAMOa_narwhal	DAMOa2014a_nplots	DAMOa2014a_PW_Animation
GEORa	GEORa_narwhal	GEORa2013_nplots	GEORa2013_PW_Animation
HALFb	HALFb_narwhal	HALFb2013a_nplots	HALFb2013a_PW_ Animation
HALFc	HALFc_narwhal	HALFc2013a_nplots	HALFc2013a_PW_ Animation
LOCKb	LOCKb_narwhal	LOCKb2013_nplots	LOCKb2013_PW_ Animation
MAIVb	MAIVb_narwhal	MAIVb2012a_nplots	MAIVb2012a_PW_ Animation
		MAIVb2013a_nplots*	MAIVb2013a_PW_Animation*
		MAIVb2013c_nplots	MAIVb2013c_PW_ Animation
MAIVc	MAIVc_narwhal	MAIVc2013_nplots	MAIVc2013_PW_ Animation
NEKOa	NEKOa_narwhal	NEKOa2012a_nplots	NEKOa2012a_PW_Animation
		NEKOa2013_nplots	NEKOa2013_PW_Animation
		NEKOa2014a_nplots	NEKOa2014a_PW_Animation
NEKOb	NEKOb_narwhal	NEKOb2013_nplots*	NEKOb2013_PW_Animation*
NEKOc	NEKOc_narwhal	NEKOc2013_nplots	NEKOc2013_PW_Animation
		NEKOc2014b_nplots	NEKOc2014b_PW_Animation
PETEc	PETEc_narwhal	PETEc2013_nplots	PETEc2013_PW_ Animation
		PETEc2014_nplots	PETEc2014_PW_ Animation
PETEe	PETEe_narwhal	PETEe2013_nplots	PETEe2013_PW_ Animation
PETEf	PETEf_narwhal	PETEf2014a_nplots	PETEf2014a_PW_ Animation
SPIGa	SPIGa_narwhal	SPIGa2012a_nplots	SPIGa2012a_PW_ Animation
		SPIGa2013b_nplots	SPIGa2013b_PW_ Animation
		SPIGa2014_nplots	SPIGa2014_PW_ Animation

Table 3: List of *Narwhal Plots* and *Penguin Watch Narwhal Animations* stored in the Dryad Digital Repository (Data Citation 2). See Chapter 3 – 'Code Availability' for a link to the *Narwhal Plotting Script* (R v 3.6.0; R Core Team, 2019). Animations were created using GIMP (v 2.8; *GNU Image Manipulation Program (GIMP)*, 2014) and VirtualDub (v 1.10.4; Lee, 2013).

* Missing one plot owing to corruption of raw time-lapse image.

Camera name	*Pengbot Out Folder* (.matlab)	*Pengbot Density Maps* (.jpg)	*Pengbot Count File* (.csv)
DAMOa	DAMOa_pengbot_out	DAMOa_density_maps	DAMOa_pengbot_count
GEORa	GEORa_pengbot_out	GEORa_density_maps	GEORa_pengbot_count
HALFb	HALFb_pengbot_out	HALFb_density_maps	HALFb_pengbot_count
HALFc	HALFc_pengbot_out	HALFc_density_maps	HALFc_pengbot_count
LOCKb	LOCKb_pengbot_out	LOCKb_density_maps	LOCKb_pengbot_count
MAIVb	MAIVb_pengbot_out	MAIVb_density_maps	MAIVb_pengbot_count
MAIVc	MAIVc_pengbot_out	MAIVc_density_maps	MAIVc_pengbot_count
NEKOa	NEKOa_pengbot_out	NEKOa_density_maps	NEKOa_pengbot_count
NEKOb	NEKOb_pengbot_out	NEKOb_density_maps	NEKOb_pengbot_count
NEKOc	NEKOc_pengbot_out	NEKOc_density_maps	NEKOc_pengbot_count
PETEc	PETEc_pengbot_out	PETEc_density_maps	PETEc_pengbot_count
PETEe	PETEe_pengbot_out	PETEe_density_maps	PETEe_pengbot_count
PETEf	PETEf_pengbot_out	PETEf_density_maps	PETEf_pengbot_count
SPIGa	SPIGa_pengbot_out	SPIGa_density_maps	SPIGa_pengbot_count

Table 4: List of *Pengbot Out Folders*, *Pengbot Density Maps* and *Pengbot Count Files* stored in the Dryad Digital Repository (Data Citation 2). See Chapter 3 – 'Code Availability' for links to the *Pengbot* Model, and the *Pengbot Counting Script* (R v 3.6.0; R Core Team, 2019).

References

GNU Image Manipulation Program (GIMP). (2014). Retrieved from https://www.gimp.org/

Jones, F. M., Allen, C., Arteta, C., Arthur, J., Black, C., Emmerson, L. M., … Hart, T. (2018). Time-lapse imagery and volunteer classifications from the Zooniverse Penguin Watch project. *Scientific Data*, *5*, 180124. https://doi.org/10.1038/sdata.2018.124

Jones, F. M., Allen, C., Arteta, C., Arthur, J., Black, C., Emmerson, L. M., … Hart, T. (2019). Author Correction: Time-lapse imagery and volunteer classifications from the Zooniverse Penguin Watch Project. *Scientific Data*, *6*(1), 1–2.

Lee, A. (2013). *VirtualDub*. Retrieved from http://www.virtualdub.org/

R Core Team. (2019). *R: A language and environment for statistical computing.* Retrieved from https://www.r-project.org/

Data Citation 2:

Jones, F.M., Arteta, C., Lempitsky, V., Zisserman, A., Lintott, C.J. & Hart, T. Data from: Processing citizen science- and machine- annotated time-lapse imagery for biologically meaningful metrics. Dryad Digital Repository. DOI:10.5061/dryad.94sp17b

Appendix 6a: Chapter 4 – Huddle Scores

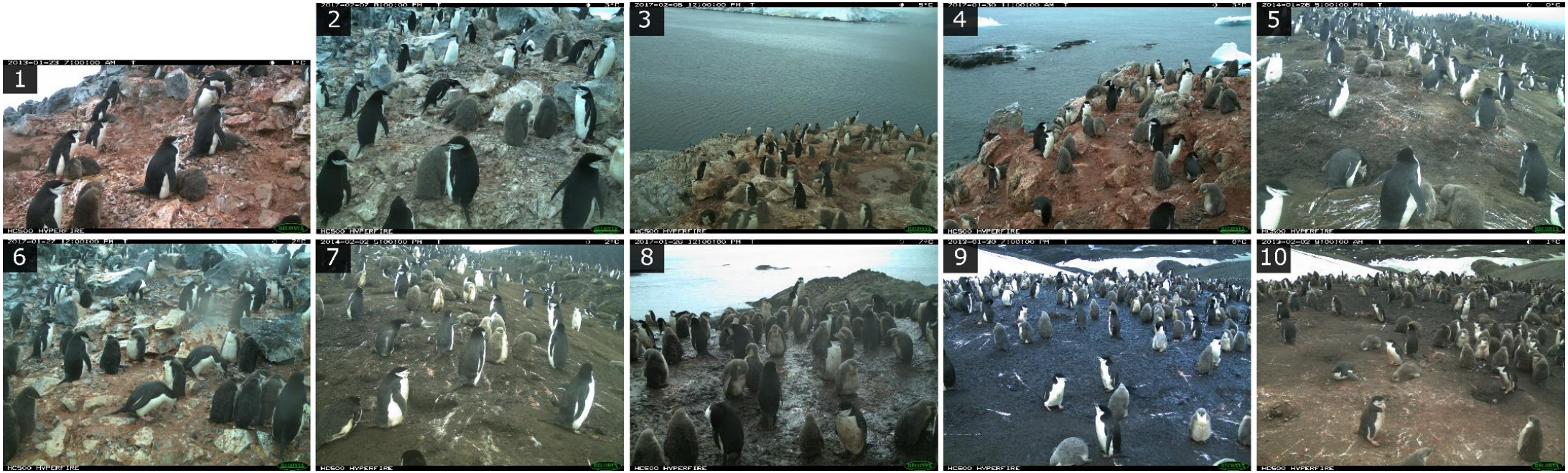


Figure 1: Images representing each 'Huddle Score' (1-10) – a subjective measure of the degree of chick huddling (aggregation) behaviour observed in a colony, within the field of view of a time-lapse image. A Huddle Score of 1 is given when chicks are at the nest, or otherwise show no aggregation behaviour. A Huddle Score of 2-4 shows that some chicks have left the nest and are coming into contact with others, but are not forming distinct groups. A score of 5-7 is used when some individuals are forming small aggregations. A score of 8-10 is given when large, distinct chick aggregations are observed.

Appendix 6b: Chapter 4 – Assessing Temperature as a Driver of Huddling Behaviour – Temperature Data and Analysis

	Site	Season	Date of 'maximum huddle'	Start of two week period	End of two week period	Mean temperature (σ)	Minimum temperature	Maximum temperature	Range
			(Julian Days)			(°C)			
'Maximum huddle' date only (n = 9 for all sites except SSISAa, where n =7)	AITCd	2017	394	N/A	N/A	4.00 (0.50)	3	5	2
	BAILa	2013	395			1.22 (1.09)	0	3	3
	BAILb	2017	372			3.67 (0.87)	2	5	3
	BOOTc	2017	408			2.44 (1.13)	1	4	3
	GASTa	2017	402			5.44 (1.88)	4	10	6
	ORNEb	2013	388			4.89 (1.45)	2	7	5
	ORNEa	2014	403			-1.44 (0.73)	-2	0	2
	ORNEa	2017	392			2.67 (0.71)	2	4	2
	SSISAa	2014	400			0.71 (0.76)	0	2	2
	TETRa	2017	395			4.00 (1.22)	2	6	4
Surrounding two week period (n = 135 for all sites except SSISAa, where n =105)	AITCd	2017	394	387	401	5.01 (2.10)	1	11	10
	BAILa	2013	395	388	402	2.13 (2.08)	-2	9	11
	BAILb	2017	372	365	379	5.73 (3.24)	1	19	18
	BOOTc	2017	408	401	415	4.95 (3.58)	-2	15	17
	GASTa	2017	402	395	409	5.04 (3.11)	0	14	14
	ORNEb	2013	388	381	395	4.44 (4.12)	-1	16	17
	ORNEa	2014	403	396	410	2.61 (4.27)	-4	18	22
	ORNEa	2017	392	385	399	5.70 (4.03)	1	20	19
	SSISAa	2014	400	393	407	2.92 (2.09)	0	10	10
	TETRa	2017	395	388	402	4.56 (3.76)	0	19	19

Table 1: Temperature data (°C) for each site/year, for the 'maximum huddle' date (see main text) and the surrounding two week period.

Site	Season	Difference between 'maximum huddle' date and surrounding two week period (°C)					
		Mean temperature	'Maximum huddle' date is:	Minimum temperature	'Maximum huddle' date is:	Maximum temperature	'Maximum huddle' date is:
AITCd	2017	1.01	Colder	-2	Warmer	6	Colder
BAILa	2013	0.91	Colder	-2	Warmer	6	Colder
BAILb	2017	2.06	Colder	-1	Warmer	14	Colder
BOOTc	2017	2.51	Colder	-3	Warmer	11	Colder
GASTa	2017	-0.40	Warmer	-4	Warmer	4	Colder
ORNEb	2013	-0.45	Warmer	-3	Warmer	9	Colder
ORNEa	2014	4.05	Colder	-2	Warmer	18	Colder
ORNEa	2017	3.03	Colder	-1	Warmer	16	Colder
SSISAa	2014	2.21	Colder	0	Equal	8	Colder
TETRa	2017	0.56	Colder	-2	Warmer	13	Colder

Table 2: Difference (in °C) between the mean, minimum and maximum daytime (0900-1700 hours) temperatures calculated for the 'maximum huddle' date (see main text) and the surrounding two week period, and whether the 'maximum huddle' date was warmer than, colder than, or equal to the surrounding two week period. Differences were calculated by subtracting the 'maximum huddle' temperatures from the temperatures for the surrounding two week period.

Site	Dataset – 1 = 'max. huddle' date; 2 = 2 week period	Mean (std dev) temperature (°C)	n	df	t	95% Confidence Interval (CI)	Lower CI	Upper CI
AITCd2017	1	4.00 (0.5)	9	8	2.306	0.384	3.616	4.384
	2	5.01 (2.10)	135	134	1.9778	0.357	4.653	5.367
BAILa2013	1	1.22 (1.09)	9	8	2.306	0.838	0.382	2.058
	2	2.13 (2.08)	135	134	1.9778	0.354	1.776	2.484
BAILb2017	1	3.67 (0.87)	9	8	2.306	0.669	3.001	4.339
	2	5.73 (3.24)	135	134	1.9778	0.552	5.178	6.282
BOOTc2017	1	2.44 (1.13)	9	8	2.306	0.869	1.571	3.309
	2	4.95 (3.58)	135	134	1.9778	0.609	4.341	5.559
GASTa2017	1	5.44 (1.88)	9	8	2.306	1.445	3.995	6.885
	2	5.04 (3.11)	135	134	1.9778	0.529	4.510	5.569
ORNEb2013	1	4.89 (1.45)	9	8	2.306	1.115	3.775	6.005
	2	4.44 (4.12)	135	134	1.9778	0.701	3.739	5.141
ORNEa2014	1	-1.44 (0.73)	9	8	2.306	0.561	-2.001	-0.879
	2	2.61 (4.27)	135	134	1.9778	0.727	1.883	3.337
ORNEa2017	1	2.67 (0.71)	9	8	2.306	0.546	2.124	3.216
	2	5.70 (4.03)	135	134	1.9778	0.686	5.014	6.386
SSISAa2014	1	0.71 (0.76)	7	6	2.447	0.703	0.007	1.412
	2	2.92 (2.09)	105	104	1.983	0.404	2.516	3.324
TETRa2017	1	4.00 (1.22)	9	8	2.306	0.938	3.062	4.938
	2	4.56 (3.76)	135	134	1.9778	0.640	3.920	5.200

Table 3: Temperature data (see Table 1) with 95% Confidence Interval values.

Site	Season	Mean temperature (°C) over 2 week period	Mean temperature (°C) on 'maximum huddle' date	Difference in temperature (°C)
AITCd	2017	5.01	4.00	1.01
BAILa	2013	2.13	1.22	0.91
BAILb	2017	5.73	3.67	2.06
BOOTc	2017	4.95	2.44	2.51
GASTa	2017	5.04	5.44	-0.40
ORNEb	2013	4.44	4.89	-0.45
ORNEa	2014	2.61	-1.44	4.05
ORNEa	2017	5.70	2.67	3.03
SSISAa	2014	2.92	0.71	2.21
TETRa	2017	4.56	4.00	0.56
			Sample size (n)	10
			Degrees of freedom (df)	9
			Mean difference ($\bar{d}$)	1.549
			Std dev (s_d)	1.474
			Std error ($SE\bar{d}$) = ($\frac{s_d}{\sqrt{n}}$)	0.466
			t-statistic (T) = ($\frac{\bar{d}}{SE\bar{d}}$)	3.324
			p-value (p)	< 0.01

Table 4: Matched pairs t-test – input data and calculation

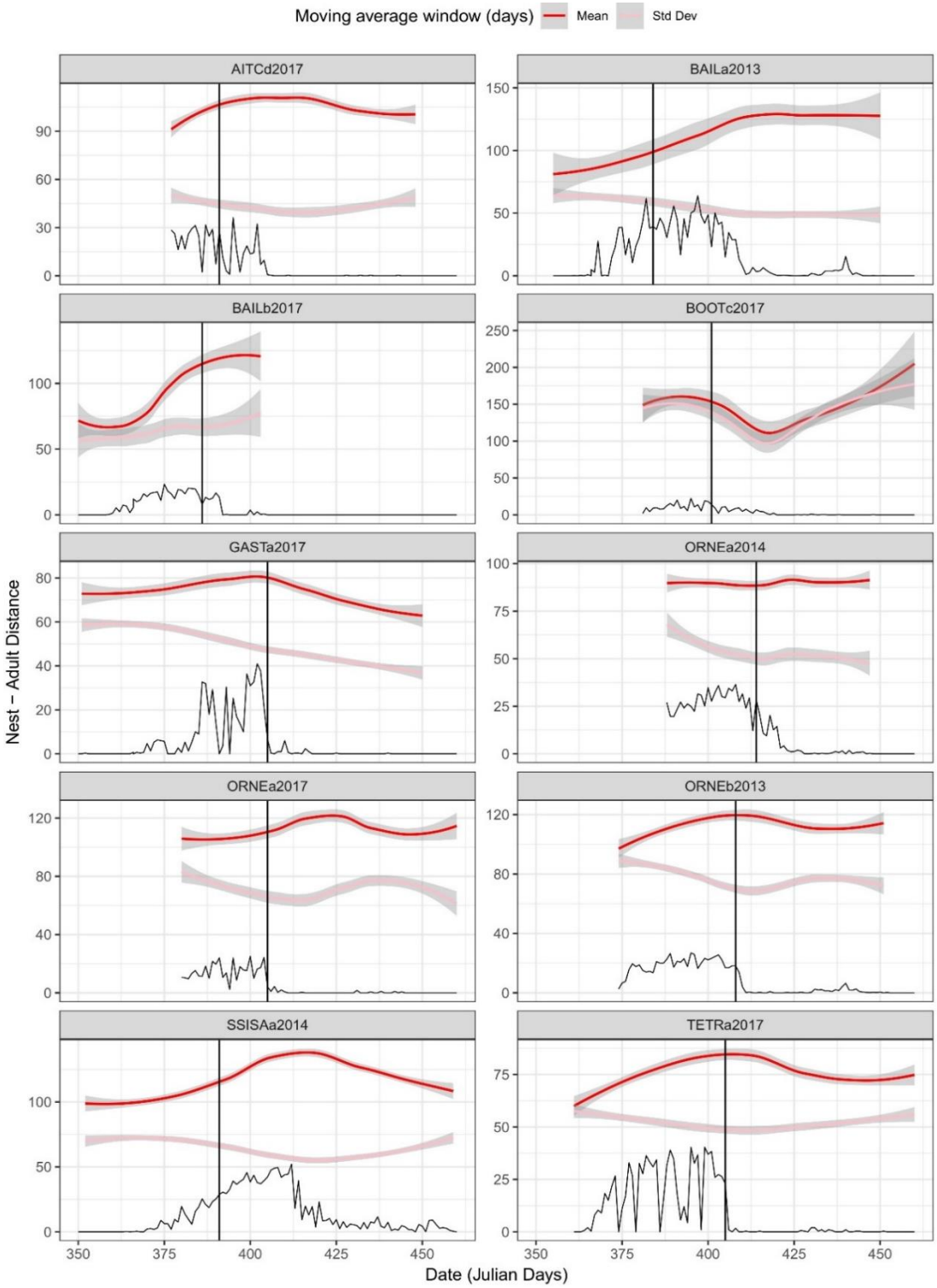


Figure 1: Nest-adult distances across the chick-rearing period for ten Chinstrap penguin (*Pygoscelis antarcticus*) colonies/years. The 50% loess-smoothed mean (red line) and standard deviation (pink line) for nest-adult distances are shown (grey shaded zone = standard error), alongside number of chicks (black trend line) and the crèche start date, as determined by an expert (vertical black line).

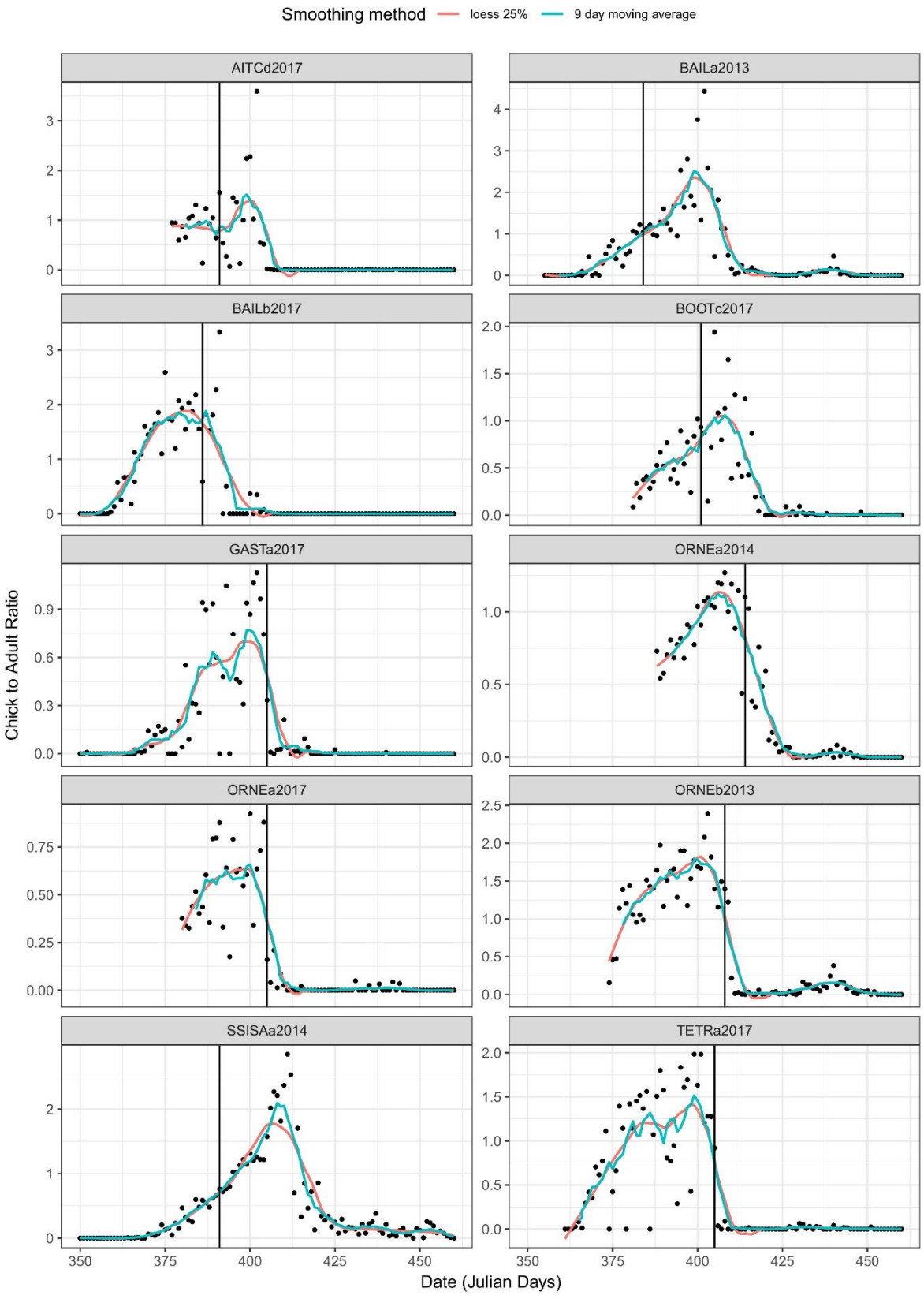


Figure 2: Ratio of chicks to adults (daily mean chicks divided by daily mean adults) for ten Chinstrap penguin (*Pygoscelis antarcticus*) colonies/years which were monitored over the chick rearing period. A 25% loess smoothing curve (pink line) and 9 day moving average (green line) are fitted to the ratios. Black vertical lines show the estimated crèche start dates as determined by an expert.

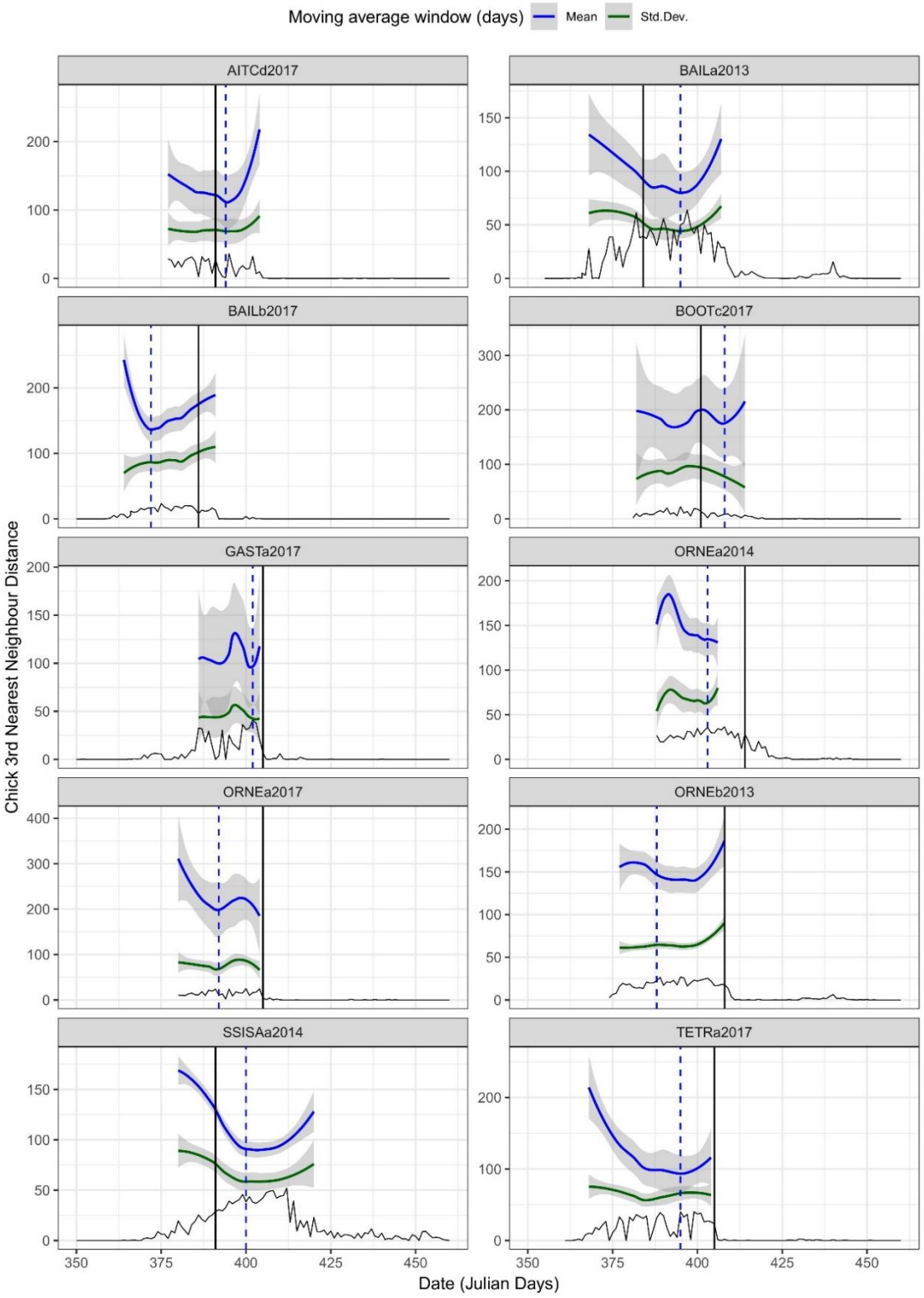


Figure 3: The 50% loess-smoothed mean (blue line) and standard deviation (green line) chick third nearest neighbour distances (nnds) for ten Chinstrap penguin (*Pygoscelis antarcticus*) colonies/years monitored over the chick-rearing period (grey shaded zones = standard error). The number of chicks is also shown (black trend line). The date corresponding to the minimum daily mean chick third nearest neighbour distance is displayed (blue hashed line), alongside the crèche start date, as estimated by an expert (black vertical line).

Appendix 7a: Chapter 5 – Supplementary Materials – Part 1 (S1)

Image annotation, data availability, model checks, and the impact of precipitation

Large-scale assessment of intra- and inter-annual nesting success using a remote camera network

Image Annotation

Images were captured by Reconyx HC500 Hyperfire Trail Cameras, as part of the *Penguin Watch* remote time-lapse camera network (Fig. S1-1 top left). The locations of individual nests were annotated on a representative image from each site/year (Fig. S1-1 top right). The xy coordinates of the nest locations were used to demarcate 'nest zones' through Voronoi tessellation (Fig. S1-1 bottom). These zones remained constant throughout each site/year, as camera position was fixed.

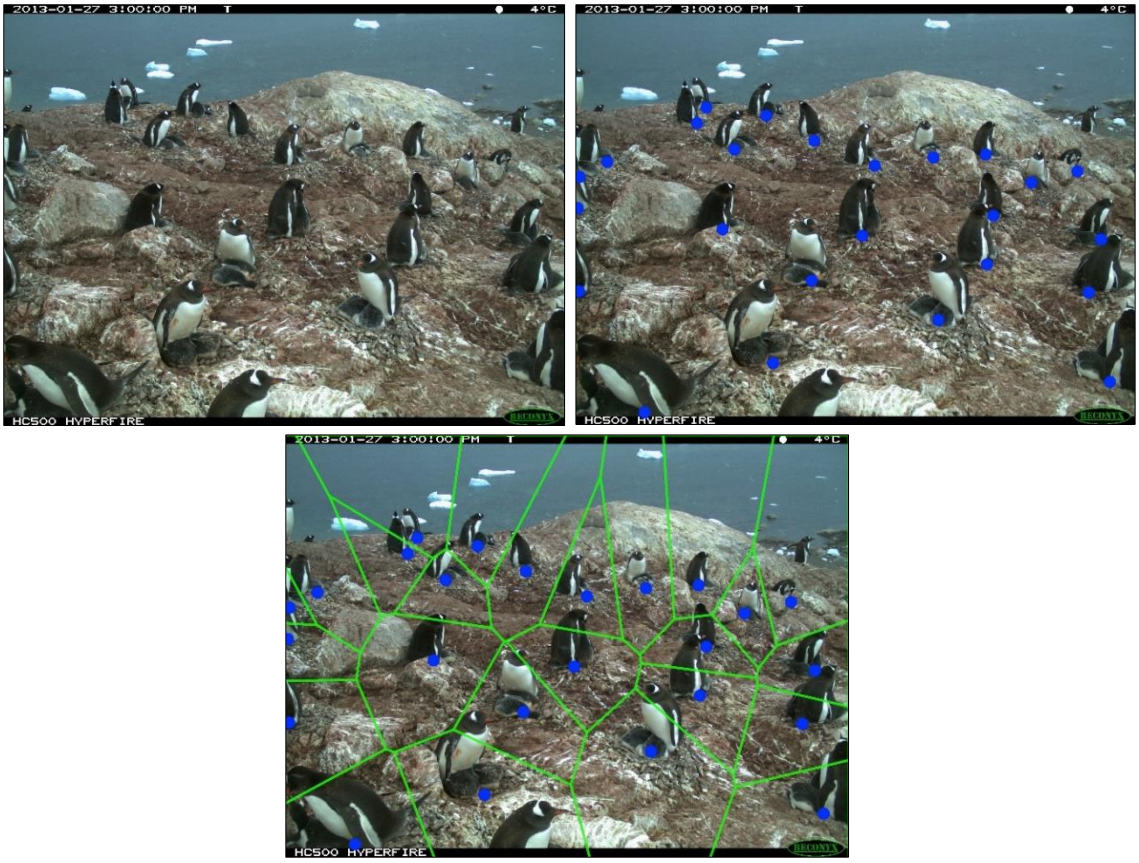


Figure S1-1: Image processing progression – raw image (top left) to image with annotated nest (top right) to image with annotated nest zones (bottom).

Each image captured over the course of the season was uploaded to the Zooniverse platform (Fig. S1-2). We classified the location of penguin chicks in each image obtained from the remote time-lapse camera network from the sighting of the first chick at a site to the crèche stage (the period during which penguin chicks begin to spend less time at the nest and form large chick aggregations). The zones were used as guidelines for users, as chicks from one nest may slightly stray into an adjoining zone. We did not mark any chick where there was any ambiguity regarding which nest it belonged to, as false positives (in the form of incorrect nest membership for a given chick) would bias estimates in this modelling framework. Ambiguous nest membership was rarely an issue in scoring and all classificiations were reviewed for accuracy to avoid potential false positives. Chick body position and proximity to the nest were used to distinguish live from dead chicks (only live chicks were scored), though there was rarely any ambiguity. This method effectively provides a time series of the number of observed chicks for each nest/site/year (Fig. 1b in main text (Chapter 5)).

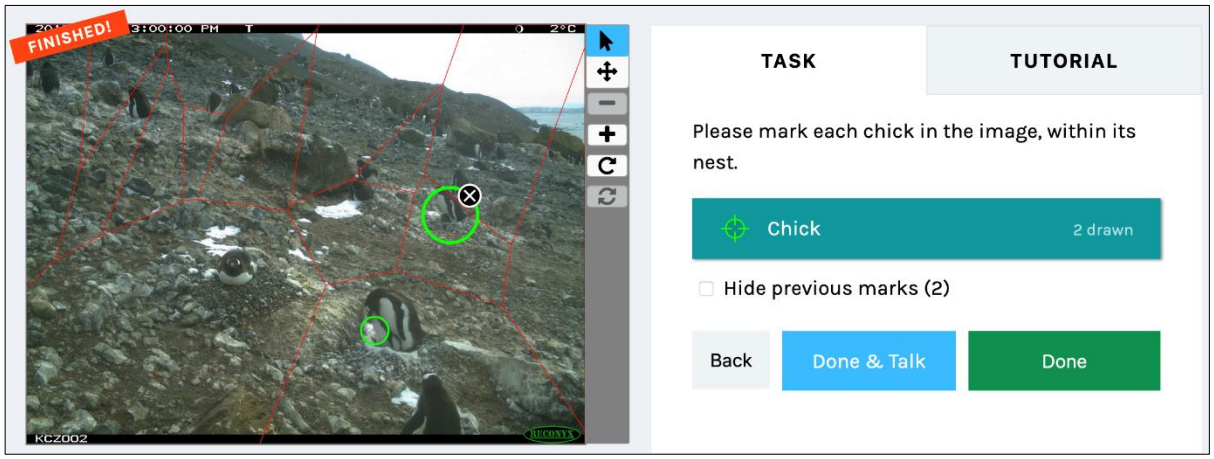


Figure S1-2: Chick classification on the Zooniverse platform. For each image, users classify the location of all visible chicks by clicking on them. The nest zones are represented by the red polygons while user-identified penguin chicks are represented by the green circles.

Data Availability

Site Name	Year	Latitude	Longitude	Chicks/ Pair	Uncertainty (sd)	Source
Brown Bluff	2016	-63.53	-56.89	0.56	<0.01	This study
Brown Bluff	2018	-63.53	-56.89	1.48	0.07	This study
Cooper Bay	2015	-54.78	-35.83	1.90	<0.01	This study
Cooper Bay	2017	-54.78	-35.83	1.66	0.05	This study
Cuverville Island	2014	-64.41	-62.38	1.14	<0.01	This study
Cuverville Island	2015	-64.41	-62.38	1.32	<0.01	This study
Damoy Point	2014	-64.82	-63.49	1.57	<0.01	This study
Danco Island	2013	-64.44	-62.37	1.57	<0.01	This study
Danco Island	2014	-64.44	-62.37	1.31	<0.01	This study
Danco Island	2015	-64.44	-62.37	1.48	<0.01	This study
Georges Point	2013	-64.67	-62.67	1.69	<0.01	This study
Georges Point	2014	-64.67	-62.67	1.50	<0.01	This study
Georges Point	2015	-64.67	-62.67	1.33	<0.01	This study
Georges Point	2017	-64.67	-62.67	1.65	<0.01	This study
Godthul	2018	-54.29	-36.26	1.70	0.03	This study
Jougla Point	2017	-64.50	-63.30	1.53	0.05	This study
Port Lockroy	2013	-64.82	-63.29	1.63	<0.01	This study
Port Lockroy	2014	-64.82	-63.29	1.33	<0.01	This study
Port Lockroy	2015	-64.82	-63.29	1.62	<0.01	This study
Maiviken	2013	-54.24	-36.50	1.62	0.04	This study
Mikkelsen Harbor	2017	-62.54	-60.47	1.08	<0.01	This study
Mikkelsen Harbor	2018	-62.54	-60.47	1.08	0.04	This study
Neko Harbour	2013	-64.86	-62.52	1.87	<0.01	This study
Neko Harbour	2014	-64.86	-62.52	1.57	<0.01	This study
Neko Harbour	2015	-64.86	-62.52	1.13	<0.01	This study

Neko Harbour	2016	-64.86	-62.52	1.57	<0.01	This study
Ocean Harbour	2015	-54.34	-36.27	1.71	<0.01	This study
Petermann Island	2015	-65.17	-64.14	1.41	<0.01	This study
Cape Shirreff	2017	-62.46	-60.79	1.63	-	Hinke et al. 2018
Cierva Cove	2017	-64.14	-60.98	1.47	-	Hinke et al. 2018
Llano Point	2017	-62.18	-58.46	1.53	-	Hinke et al. 2018
Galindez	2017	-65.24	-64.25	1.46	-	Hinke et al. 2018
Lion's Rump	2017	-62.135	-58.13	1.26	-	Hinke et al. 2018
Petermann Island	2017	-65.17	-64.14	1.51	-	Hinke et al. 2018
Petermann Island	2004	-65.17	-64.14	1.52	-	Lynch et al. 2009
Petermann Island	2005	-65.17	-64.14	1.23	-	Lynch et al. 2009
Petermann Island	2006	-65.17	-64.14	1.42	-	Lynch et al. 2009
Petermann Island	2007	-65.17	-64.14	1.46	-	Lynch et al. 2009
Petermann Island	2008	-65.17	-64.14	1.23	-	Lynch et al. 2009

Table S1-1: Breeding success estimates from this study and estimates compiled from two other studies in the literature (providing an additional 11 site/years of information). Estimates from other studies are provided for reference, though these data were not included in any analyses. Note that measures of uncertainty for estimates from this study (given as posterior standard deviations) are not provided for estimates from other studies in the literature.

		Mean daily mortality rate and 95% credible intervals								
Site	Season	'Egg period' mean	'Egg period' 2.5% (lower 95% CI)	'Egg period' 97.5% (upper 95% CI)	'Young chick period' mean	'Young chick period' 2.5% (lower 95% CI)	'Young chick period' 97.5% (upper 95% CI)	'Old chick period' mean	'Old chick period' 2.5% (lower 95% CI)	'Old chick period' 97.5% (upper 95% CI)
BROW	2016	0.616	0.467	0.733	0.257	0.067	0.533	0.008	0	0.067
BROW	2018	0.039	0	0.067	0.042	0	0.133	0.168	0.133	0.267
COOP	2015	0.024	0	0.033	0.016	0	0.067	<0.001	0	0
COOP	2017	0.049	0	0.067	0.025	0	0.133	0.011	0	0.067
CUVE	2014	0.555	0.433	0.667	0.387	0.200	0.600	0.067	0.067	0.067
CUVE	2015	0.317	0.233	0.400	0.343	0.200	0.533	<0.001	0	0
DAMO	2014	0.078	0.033	0.100	0.036	0	0.133	0.001	0	0
DANC	2013	0.381	0.267	0.467	0.209	0	0.400	0.129	0.067	0.200
DANC	2014	0.512	0.400	0.633	0.355	0.133	0.600	0.160	0.067	0.267
DANC	2015	0.286	0.200	0.367	0.192	0.067	0.400	0.235	0.200	0.333
GEOR	2013	0.076	0.033	0.100	0.104	0	0.200	0.068	0.067	0.067
GEOR	2014	0.163	0.100	0.200	0.196	0.133	0.333	<0.001	0	0
GEOR	2015	0.240	0.167	0.300	0.171	0.067	0.333	0.133	0.133	0.133
GEOR	2017	0.167	0.100	0.200	0.056	0	0.200	<0.001	0	0
GODH	2018	0.050	0	0.067	0.066	0	0.200	0.097	0.067	0.200
JOUG	2017	0.057	0.033	0.067	0.078	0	0.133	0.054	0	0.067
LOCK	2013	0.087	0.033	0.100	0.021	0	0.133	0	0	0
LOCK	2014	0.185	0.100	0.233	0.083	0	0.200	0.199	0.200	0.200
LOCK	2015	0.124	0.067	0.167	0.070	0	0.200	0.005	0	0.067
MAIV	2013	0.218	0.133	0.300	0.171	0	0.333	0.180	0.067	0.333
MIKK	2017	0.233	0.133	0.300	0.184	0.067	0.333	0.085	0.067	0.133

MIKK	2018	0.385	0.267	0.467	0.258	0.067	0.467	0.469	0.333	0.600
NEKO	2013	0.076	0.033	0.100	0.043	0	0.133	<0.001	0	0
NEKO	2014	0.187	0.100	0.233	0.146	0.067	0.267	0.067	0.067	0.067
NEKO	2015	0.501	0.400	0.600	0.303	0.133	0.533	0.069	0.067	0.133
NEKO	2016	0.254	0.167	0.333	0.134	0	0.333	0.004	0	0.067
OCEA	2015	0.197	0.100	0.267	0.121	0	0.267	0.002411	0	0.067
PETE	2015	0.229	0.167	0.267	0.128	0.067	0.267	0.066944	0.067	0.067
	Mean	0.225	0.149	0.280	0.150	0.045	0.300	0.081	0.060	0.124

Table S1-2: Posterior mean daily mortality rate, and 95% credible intervals, during the 'egg period' (Day 1 to Day 30), 'young chick period' (Day 31 to Day 45) and the 'old chick period' (Day 46 to Day 60), for each of the 28 site/years included in the study.

		Differences in mean daily mortality rate, with 95% credible intervals								
Site	Season	'Egg period' - 'young chick period' mean difference	'Egg period' - 'young chick period' 2.5% (lower 95% CI)	'Egg period' - 'young chick period' 97.5% (upper 95% CI)	'Egg period' - 'old chick period' mean difference	'Egg period' - 'old chick period' 2.5% (lower 95% CI)	'Egg period' - 'old chick period' 97.5% (upper 95% CI)	'Young chick period' - 'old chick period' mean difference	'Young chick period' - 'old chick period' 2.5% (lower 95% CI)	'Young chick period' - 'old chick period' 97.5% (upper 95% CI)
BROW	2016	0.360	-0.033	0.666	0.608	0.433	0.733	0.249	0	0.533
BROW	2018	-0.002	-0.133	0.067	-0.128	-0.234	-0.066	-0.126	-0.200	0
COOP	2015	0.008	-0.067	0.033	0.024	0	0.033	0.016	0	0.067
COOP	2017	0.023	-0.133	0.067	0.038	-0.034	0.067	0.014	-0.067	0.133
CUVE	2014	0.169	-0.167	0.467	0.488	0.366	0.600	0.320	0.133	0.533
CUVE	2015	-0.025	-0.300	0.200	0.317	0.200	0.400	0.342	0.200	0.533
DAMO	2014	0.042	-0.100	0.100	0.077	0	0.100	0.035	0	0.133
DANC	2013	0.172	-0.167	0.467	0.252	0.133	0.367	0.080	-0.133	0.333
DANC	2014	0.158	-0.200	0.500	0.352	0.167	0.500	0.194	-0.067	0.467
DANC	2015	0.094	-0.200	0.300	0.051	-0.067	0.167	-0.043	-0.200	0.133
GEOR	2013	-0.028	-0.167	0.100	0.008	-0.066	0.033	0.036	-0.067	0.133
GEOR	2014	-0.032	-0.233	0.067	0.163	0.100	0.200	0.196	0.133	0.333
GEOR	2015	0.069	-0.166	0.233	0.107	0.034	0.167	0.037	-0.066	0.200
GEOR	2017	0.111	-0.100	0.200	0.167	0.100	0.200	0.055	0	0.200
GODH	2018	-0.017	-0.200	0.067	-0.047	-0.133	0	-0.030	-0.200	0.133
JOUG	2017	-0.020	-0.100	0.067	0.004	-0.066	0.067	0.024	-0.067	0.133
LOCK	2013	0.067	-0.100	0.100	0.087	0.033	0.100	0.021	0	0.133
LOCK	2014	0.102	-0.100	0.233	-0.014	-0.100	0.033	-0.116	-0.200	0
LOCK	2015	0.053	-0.133	0.167	0.118	0.033	0.167	0.065	-0.067	0.200

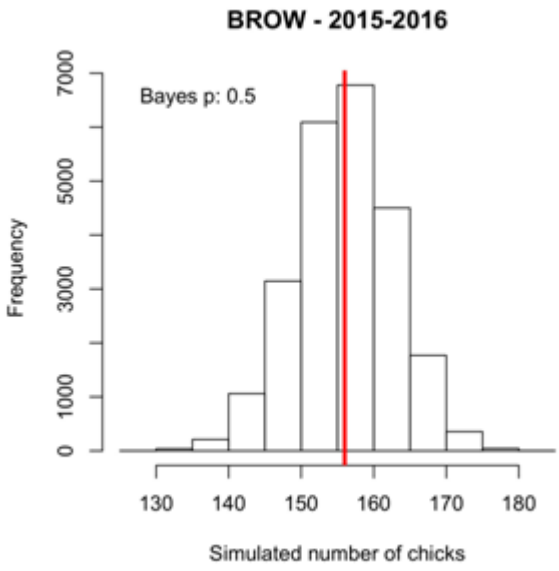
MAIV	2013	0.047	-0.233	0.300	0.038	-0.167	0.200	-0.009	-0.266	0.200
MIKK	2017	0.049	-0.200	0.233	0.148	0.034	0.233	0.099	-0.066	0.266
MIKK	2018	0.128	-0.200	0.400	-0.084	-0.267	0.067	-0.211	-0.466	0.067
NEKO	2013	0.033	-0.100	0.100	0.076	0.033	0.100	0.043	0	0.133
NEKO	2014	0.040	-0.167	0.166	0.119	0.033	0.166	0.079	0	0.200
NEKO	2015	0.198	-0.133	0.467	0.432	0.300	0.533	0.233	0.066	0.466
NEKO	2016	0.120	-0.166	0.333	0.250	0.133	0.333	0.130	0	0.333
OCEA	2015	0.076	-0.167	0.267	0.19	0.100	0.267	0.118	0	0.267
PETE	2015	0.102	-0.100	0.200	0.162	0.100	0.200	0.061	0	0.200
	Mean	0.075	-0.152	0.235	0.143	0.043	0.213	0.068	-0.057	0.231

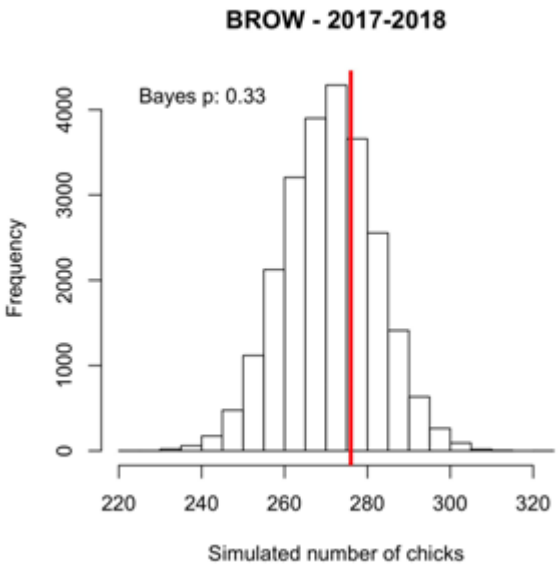
Table S1-3: Differences in posterior mean daily mortality rate, and 95% credible intervals for these differences, between the 'egg period' and the 'young chick period' (Day 1-30, to Day 31-45), 'egg period' and 'old chick period' (Day 1-30 to Day 46-60) and the 'young chick period' and 'old chick period' (Day 31-45 to Day 46-60), for each of the 28 site/years included in the study.

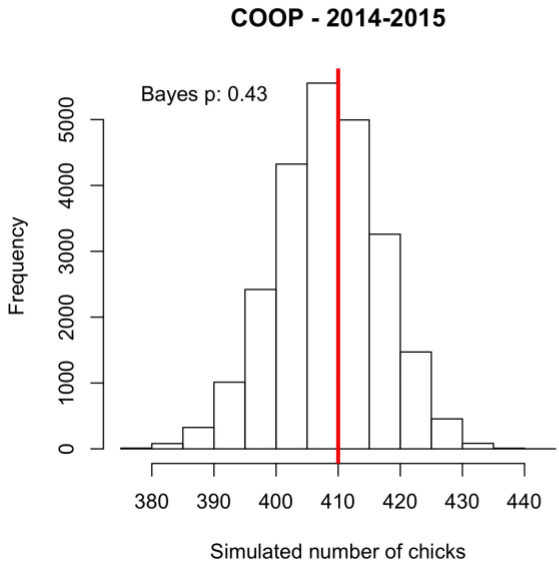
Posterior predictive check - capture-recapture model

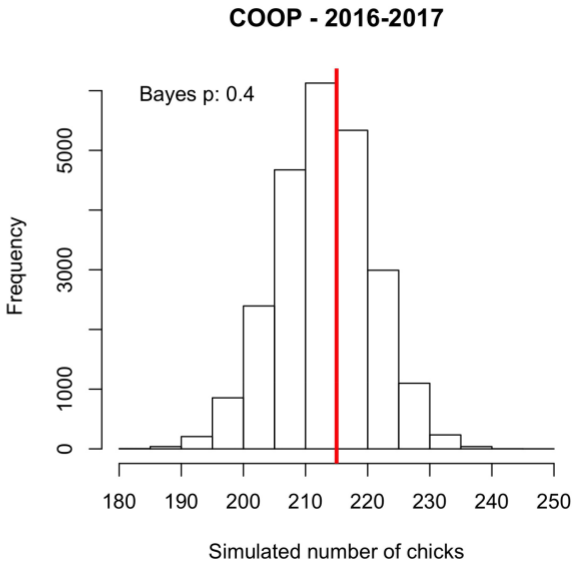
Posterior predictive checks were used to assess the degree to which the model could produce data that resemble the observed data. The test statistic chosen was the total number of observed chicks for each site/year.

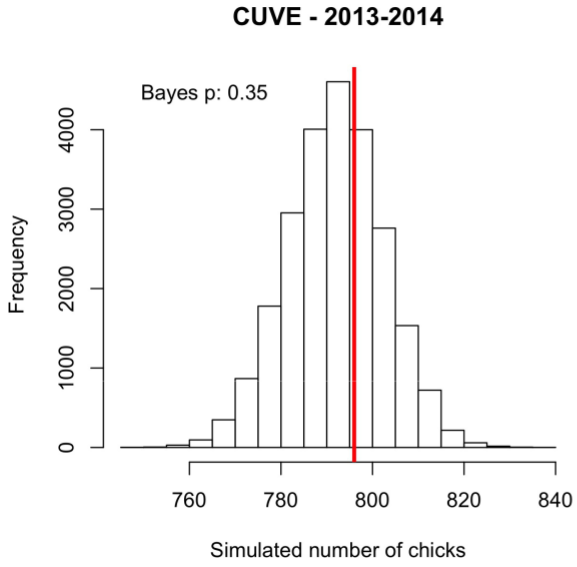
Bayesian p-values represent the proportion of generated datasets with a test statistic (in this case, the total number of chicks 'observed') that is greater than the test statistic derived from the actual data. Bayesian p-values close to zero or one indicate that the model generates data different than those used to fit the model. Results did not indicate any model misfit (as evidenced by Bayesian p-values near 0.5).

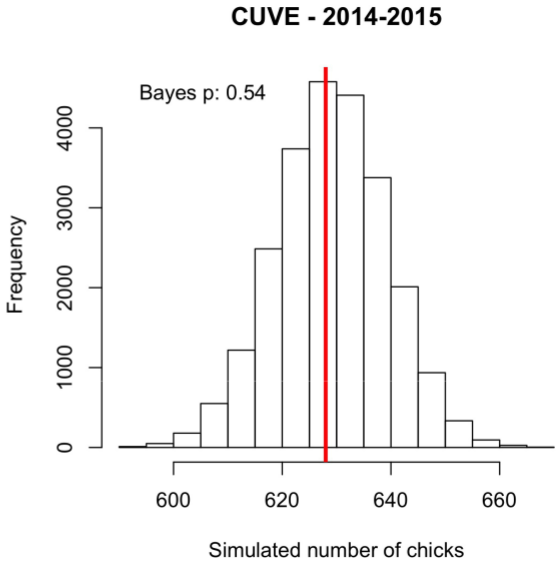


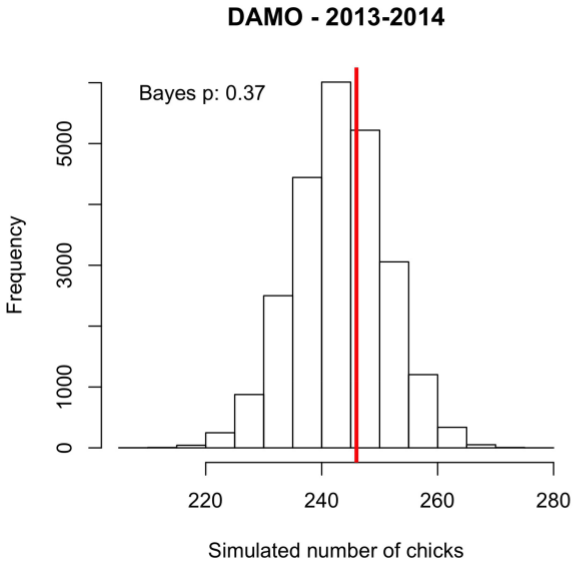


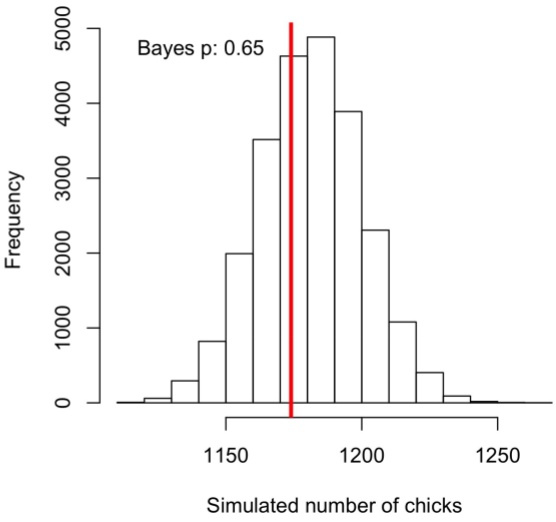

COOP - 2014-2015
Bayes p: 0.43
Frequency
0
1000
2000
3000
4000
5000
380
390
400
410
420
430
440
Simulated number of chicks

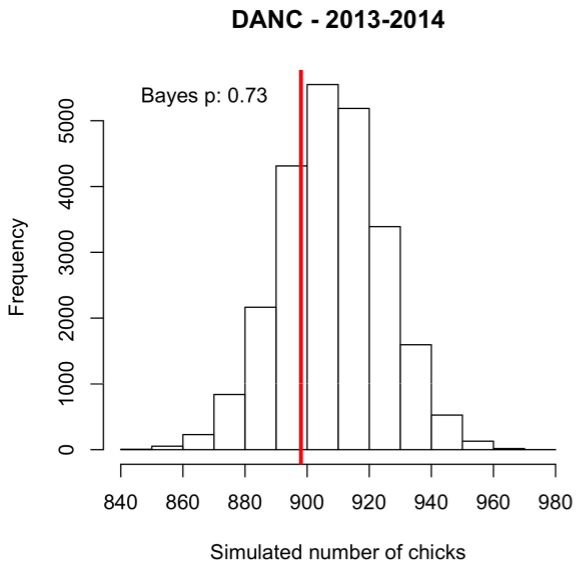

COOP - 2016-2017
Bayes p: 0.4
Frequency
0
1000
3000
5000
180
190
200
210
220
230
240
250
Simulated number of chicks

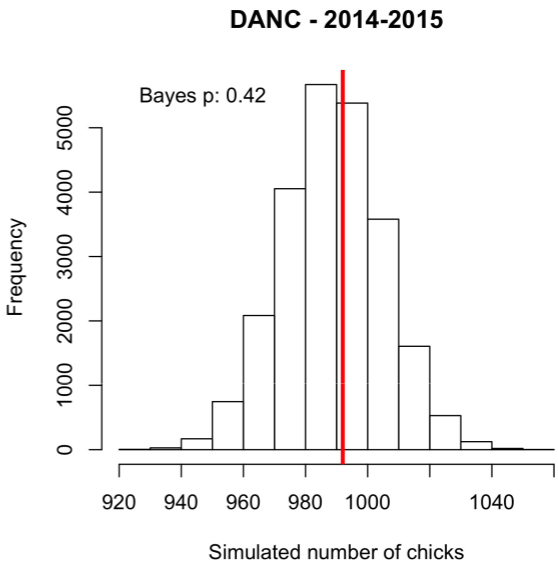

CUVE - 2013-2014
Bayes p: 0.35
Frequency
0
1000
2000
3000
4000
760
780
800
820
840
Simulated number of chicks

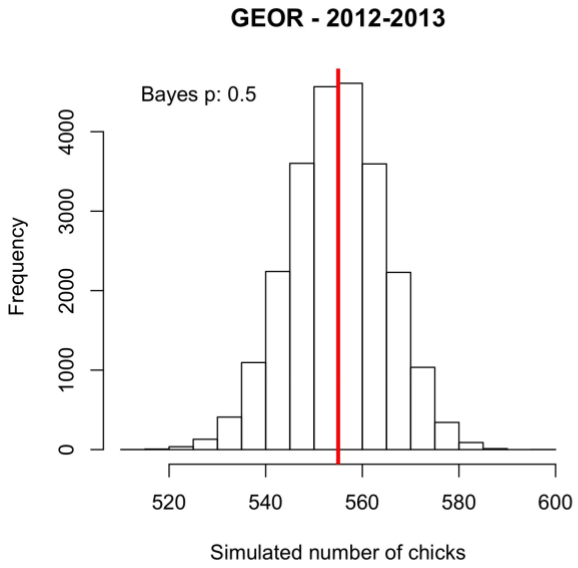

CUVE - 2014-2015
Bayes p: 0.54
Frequency
0
1000
2000
3000
4000
600
620
640
660
Simulated number of chicks

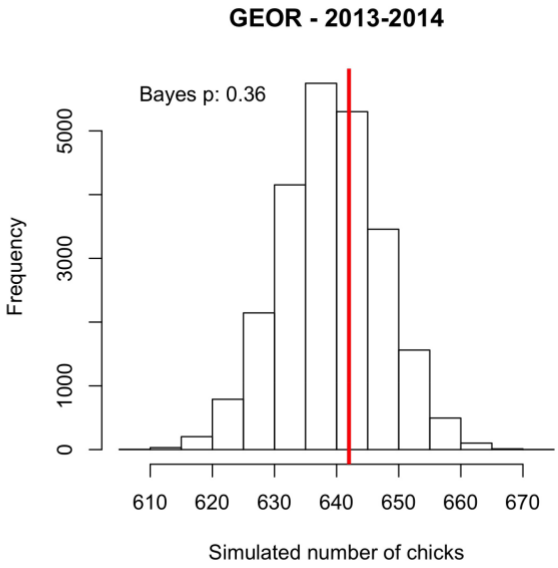

DAMO - 2013-2014
Bayes p: 0.37
Frequency
0
1000
3000
5000
220
240
260
280
Simulated number of chicks

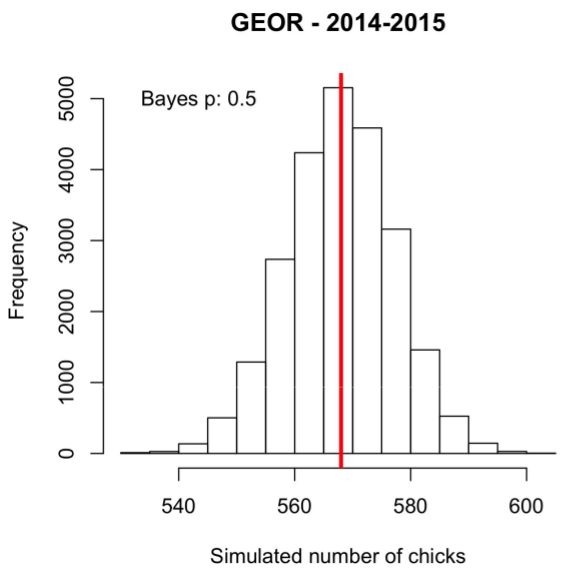

DANC - 2012-2013
Bayes p: 0.65
Frequency
0
1000
2000
3000
4000
5000
1150
1200
1250
Simulated number of chicks

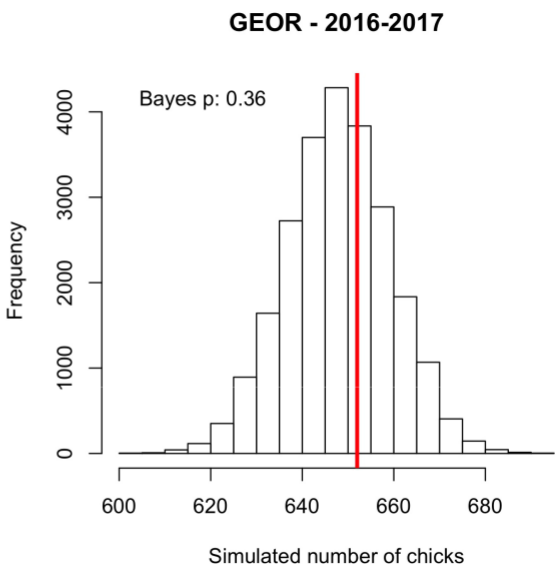

DANC - 2013-2014
Bayes p: 0.73
Frequency
0
1000
2000
3000
4000
5000
840
860
880
900
920
940
960
980
Simulated number of chicks

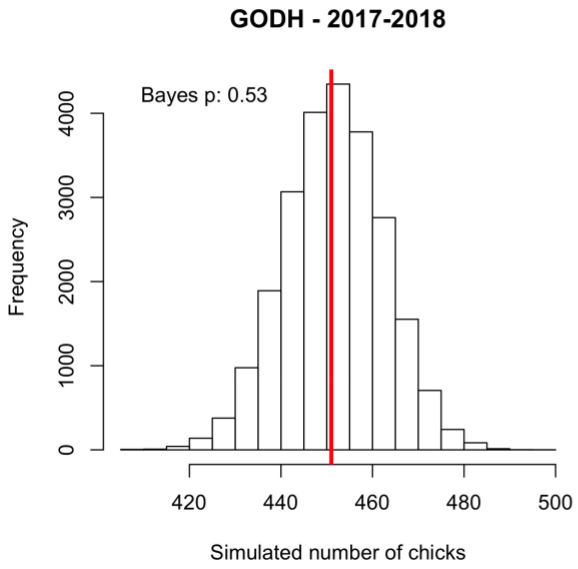

DANC - 2014-2015
Bayes p: 0.42
Frequency
0
1000
2000
3000
4000
5000
920
940
960
980
1000
1040
Simulated number of chicks

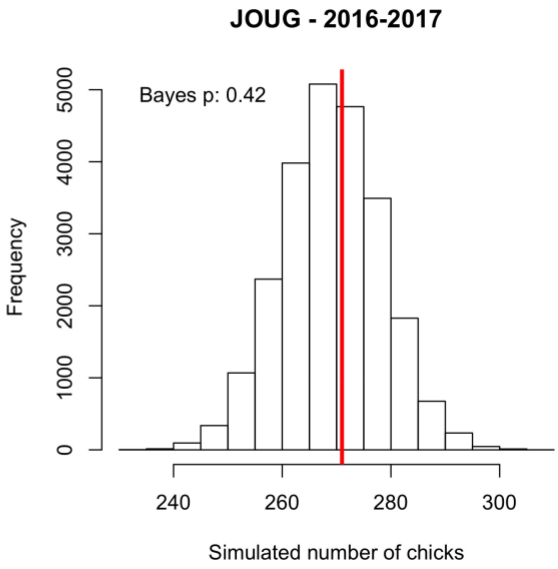

GEOR - 2012-2013
Bayes p: 0.5
Frequency
0
1000
2000
3000
4000
520
540
560
580
600
Simulated number of chicks

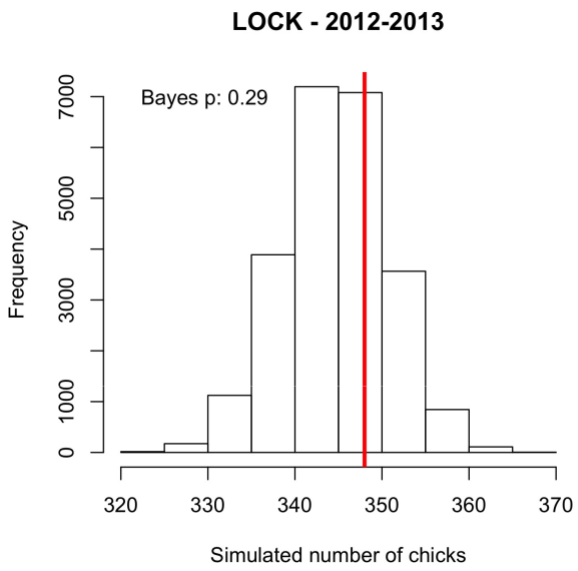

GEOR - 2013-2014
Bayes p: 0.36
Frequency
0
1000
3000
5000
610
620
630
640
650
660
670
Simulated number of chicks

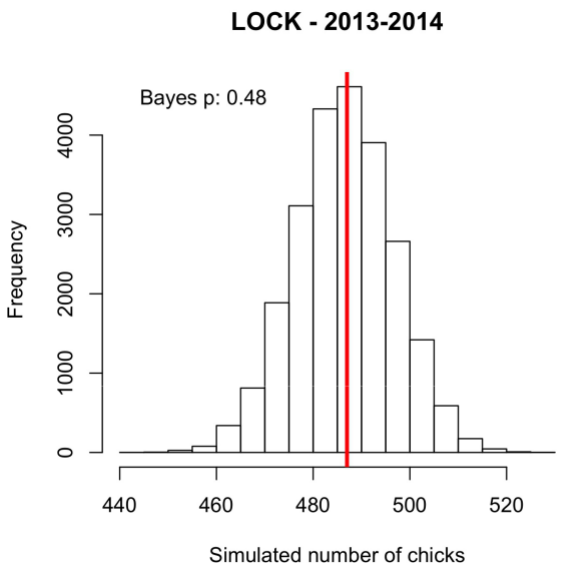

GEOR - 2014-2015
Bayes p: 0.5
Frequency
0
1000
2000
3000
4000
5000
540
560
580
600
Simulated number of chicks

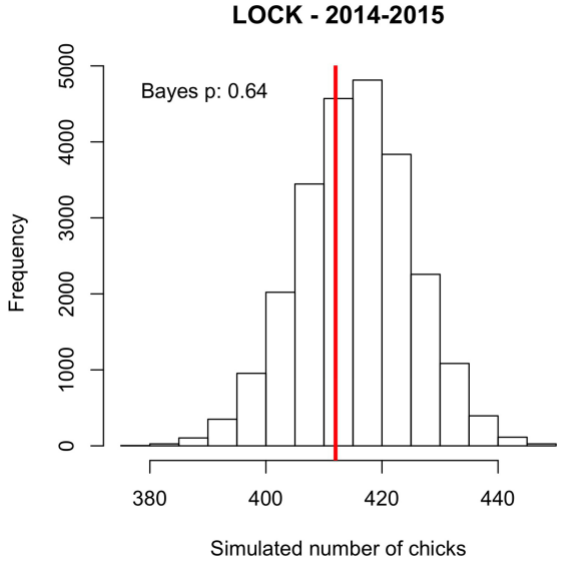

GEOR - 2016-2017
Bayes p: 0.36
Frequency
0
1000
2000
3000
4000
600
620
640
660
680
Simulated number of chicks

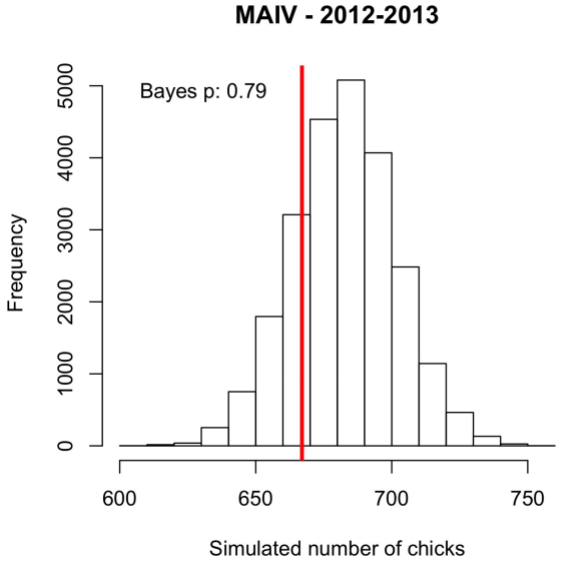

GODH - 2017-2018
Bayes p: 0.53
Frequency
0
1000
2000
3000
4000
420
440
460
480
500
Simulated number of chicks

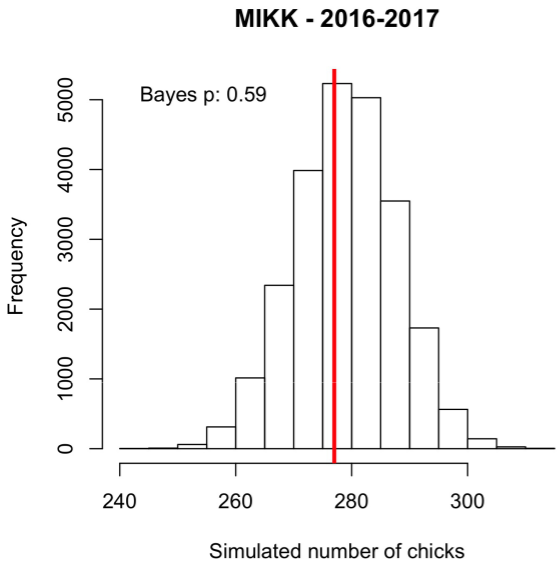

JOUG - 2016-2017
Bayes p: 0.42
Frequency
0
1000
2000
3000
4000
5000
240
260
280
300
Simulated number of chicks

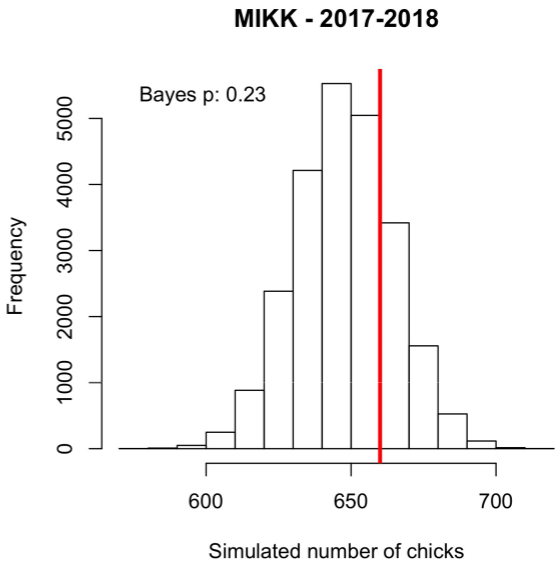

LOCK - 2012-2013
Bayes p: 0.29
Frequency
0
1000
3000
5000
7000
320
330
340
350
360
370
Simulated number of chicks

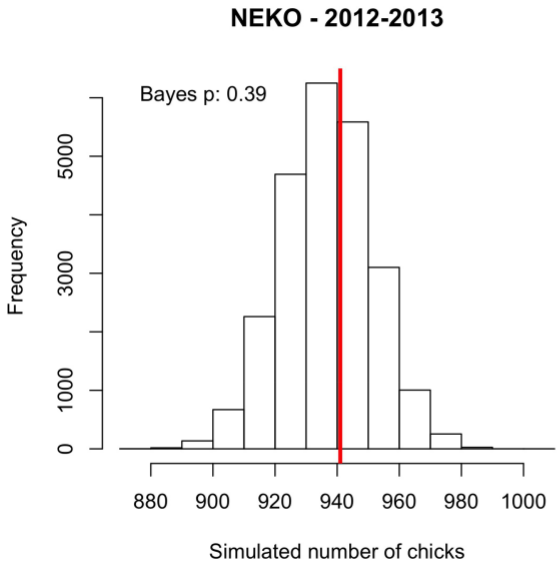

LOCK - 2013-2014
Bayes p: 0.48
Frequency
0
1000
2000
3000
4000
440
460
480
500
520
Simulated number of chicks

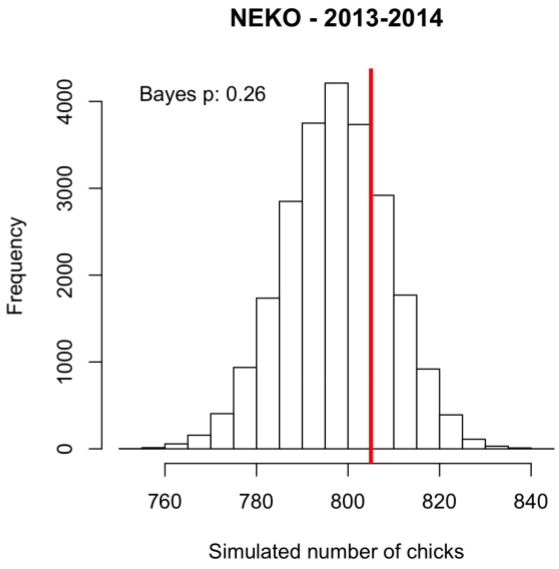

LOCK - 2014-2015
Bayes p: 0.64
Frequency
0
1000
2000
3000
4000
5000
380
400
420
440
Simulated number of chicks

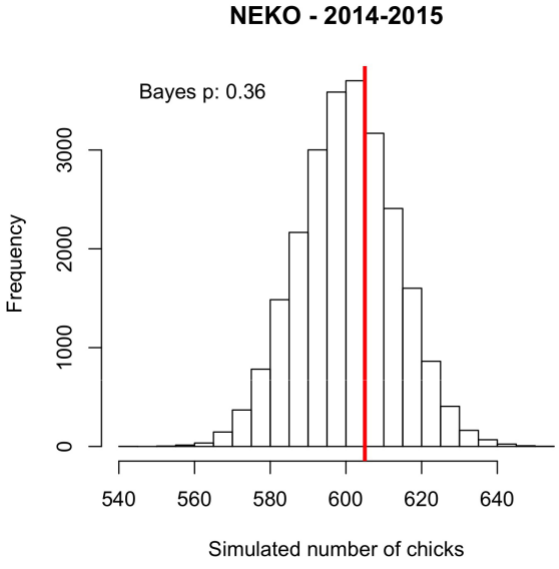

MAIV - 2012-2013
Bayes p: 0.79
Frequency
0
1000
2000
3000
4000
5000
600
650
700
750
Simulated number of chicks

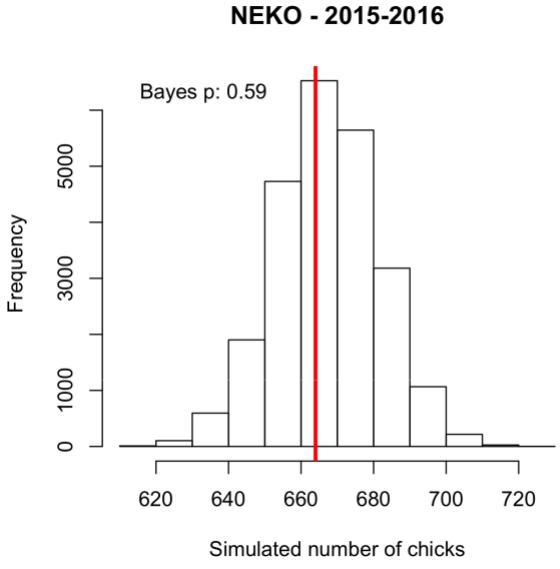

MIKK - 2016-2017
Bayes p: 0.59
Frequency
0
1000
2000
3000
4000
5000
240
260
280
300
Simulated number of chicks


MIKK - 2017-2018
Bayes p: 0.23
Frequency
0
1000
2000
3000
4000
5000
600
650
700
Simulated number of chicks


NEKO - 2012-2013
Bayes p: 0.39
Frequency
0
1000
3000
5000
880
900
920
940
960
980
1000
Simulated number of chicks


NEKO - 2013-2014
Bayes p: 0.26
Frequency
0
1000
2000
3000
4000
760
780
800
820
840
Simulated number of chicks


NEKO - 2014-2015
Bayes p: 0.36
Frequency
0
1000
2000
3000
540
560
580
600
620
640
Simulated number of chicks


NEKO - 2015-2016
Bayes p: 0.59
Frequency
0
1000
3000
5000
620
640
660
680
700
720
Simulated number of chicks

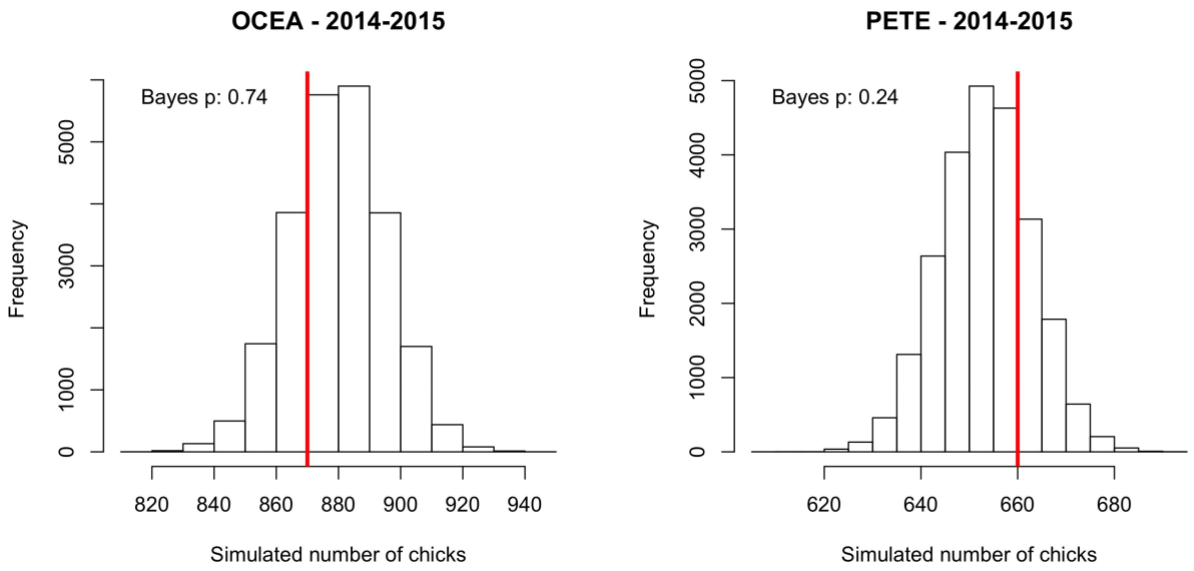


Figure S1-3: Posterior predictive check results. Histograms represent the number of chicks 'observed' in the generated data at each iteration of the model for each site/year, while the vertical red lines represent the actual number of chicks observed for each site/year. Bayesian p-values are given for each site/year. BROW - Brown Bluff, COOP - Cooper Bay, CUVE - Cuverville Island, DAMO - Damoy Point, DANC - Danco Island, GEOR - Georges Point, GODH - Godthul, JOUG - Jougla Point, LOCK - Port Lockroy, MAIV - Maiviken, MIKK - Mikkelsen Harbor, NEKO - Neko Harbor, OCEA - Ocean Harbor, PETE - Petermann Island.

Precipitation Coding

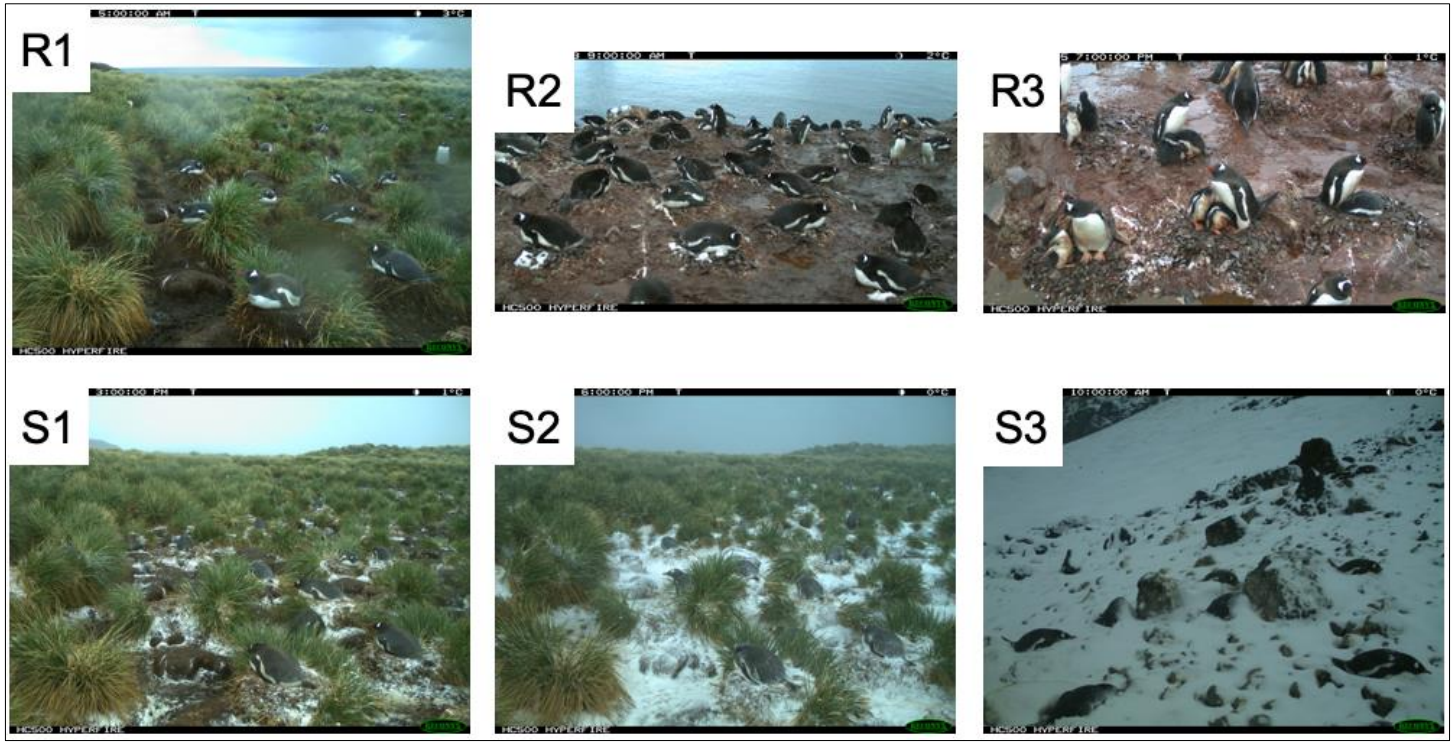


Figure S1-4: Example images used for precipitation coding. Top row represents rain events, while the bottom row represents snow events. For each precipitation type, the magnitude of the events increases from left to right (R1/S1 left, R2/S2 middle, R3/S3 right).

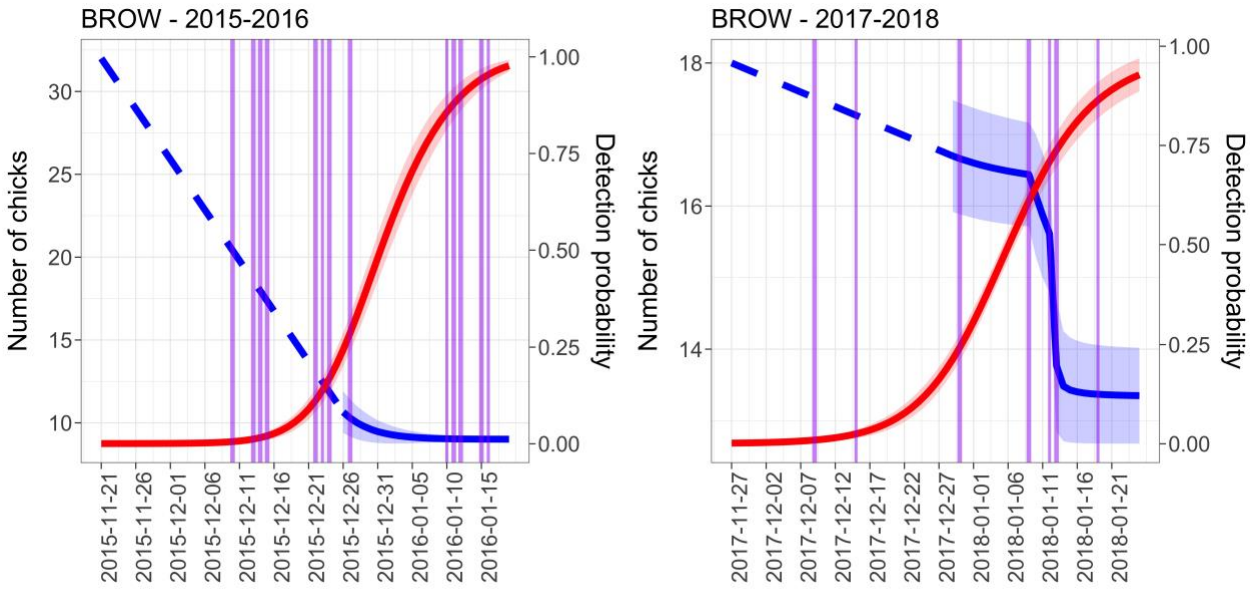
Effect of precipitation events on the timing of nest failure

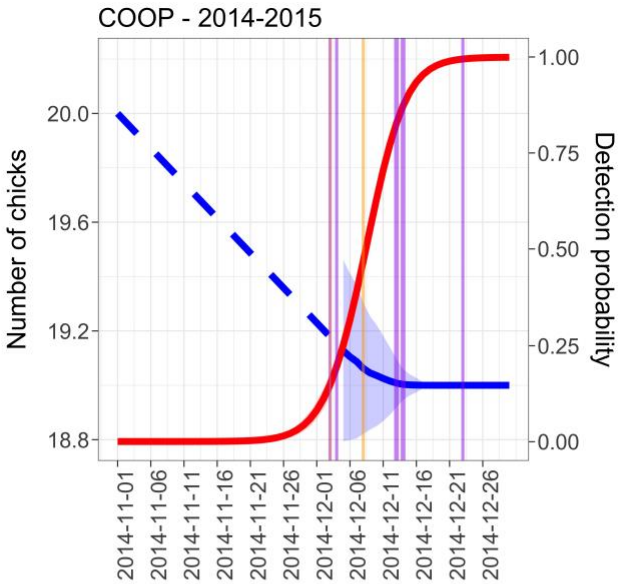
While smaller precipitation events were of little consequence for chick mortality (Fig. S1-5), large precipitation events appear to impact breeding success in a non-linear fashion (main text Fig. 5 (Chapter 5)) although this has not been formally assessed using a statistical test owing to the small sample size of extreme weather events. It appears that extreme weather events can result in increased egg failure or chick mortality.

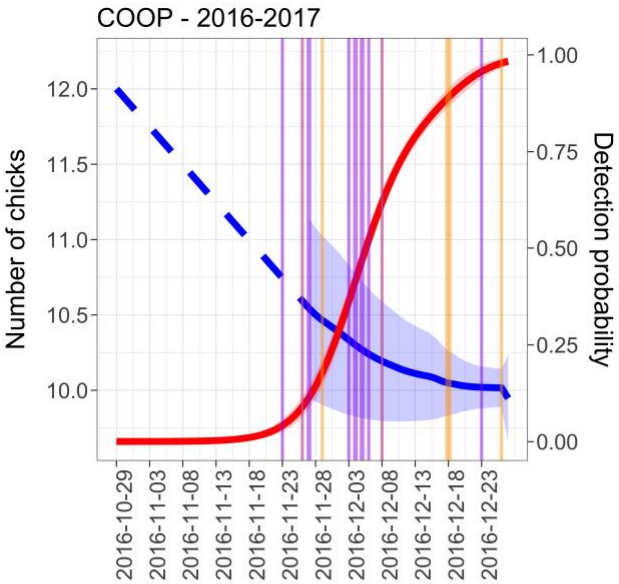
For the plots below, mean detection probability at each site in every year at every time step $\overline{p_{t,J,k}}$ was calculated as a derived quantity by averaging across the detection probabilities $p_{t,i,j,k}$ for all nests i,

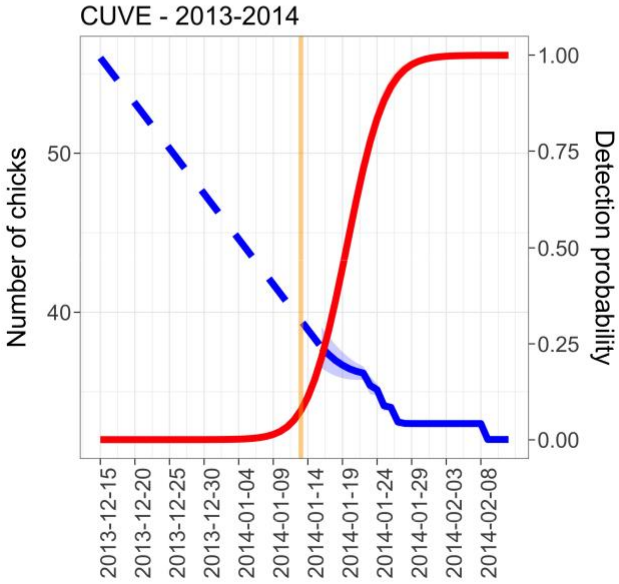
$$\overline{p_{t,J,k}} = \frac{1}{N_{j,k}} \sum p_{t,i,j,k}$$

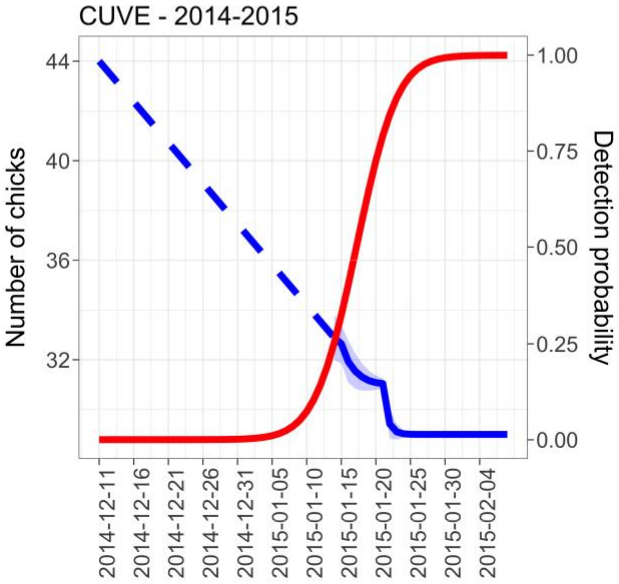
where N is the number of nests for a given site and year.

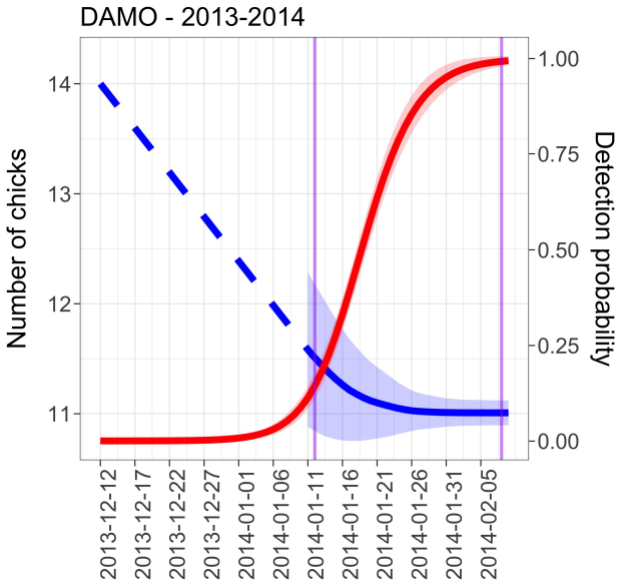


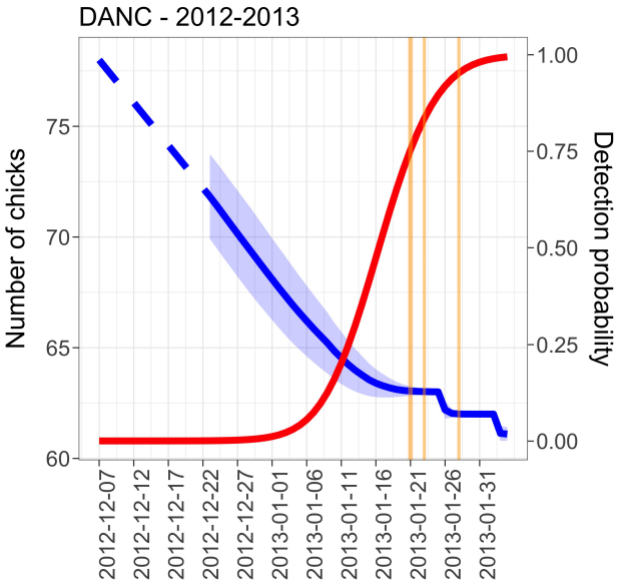

COOP - 2014-2015
Number of chicks
Detection probability
20.0
19.6
19.2
18.8
1.00
0.75
0.50
0.25
0.00
2014-11-01
2014-11-06
2014-11-11
2014-11-16
2014-11-21
2014-11-26
2014-12-01
2014-12-06
2014-12-11
2014-12-16
2014-12-21
2014-12-26

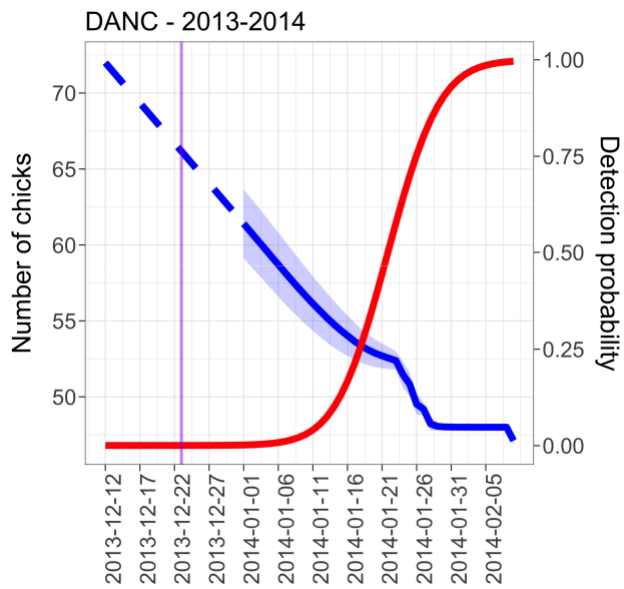

COOP - 2016-2017
Number of chicks
Detection probability
12.0
11.5
11.0
10.5
10.0
1.00
0.75
0.50
0.25
0.00
2016-10-29
2016-11-03
2016-11-08
2016-11-13
2016-11-18
2016-11-23
2016-11-28
2016-12-03
2016-12-08
2016-12-13
2016-12-18
2016-12-23

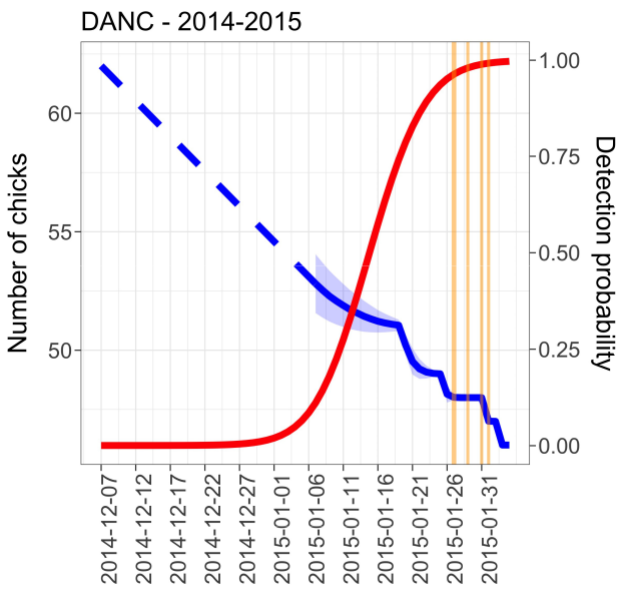

CUVE - 2013-2014
Number of chicks
Detection probability
50
40
1.00
0.75
0.50
0.25
0.00
2013-12-15
2013-12-20
2013-12-25
2013-12-30
2014-01-04
2014-01-09
2014-01-14
2014-01-19
2014-01-24
2014-01-29
2014-02-03
2014-02-08

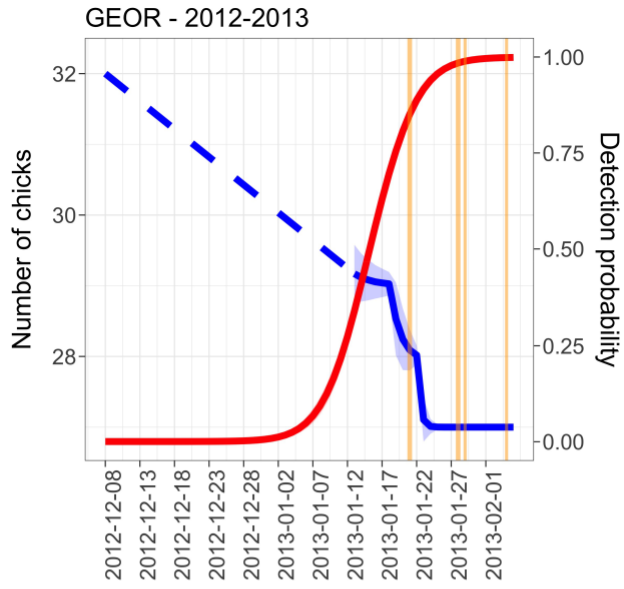

CUVE - 2014-2015
Number of chicks
Detection probability
44
40
36
32
1.00
0.75
0.50
0.25
0.00
2014-12-11
2014-12-16
2014-12-21
2014-12-26
2014-12-31
2015-01-05
2015-01-10
2015-01-15
2015-01-20
2015-01-25
2015-01-30
2015-02-04

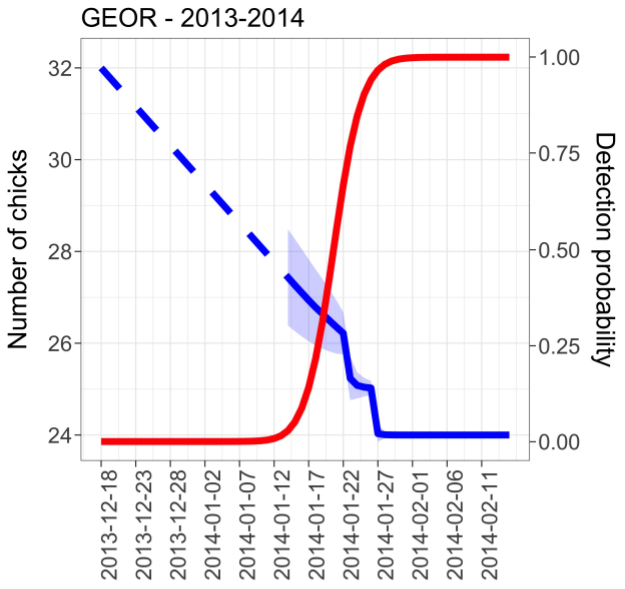

DAMO - 2013-2014
Number of chicks
Detection probability
14
13
12
11
1.00
0.75
0.50
0.25
0.00
2013-12-12
2013-12-17
2013-12-22
2013-12-27
2014-01-01
2014-01-06
2014-01-11
2014-01-16
2014-01-21
2014-01-26
2014-01-31
2014-02-05

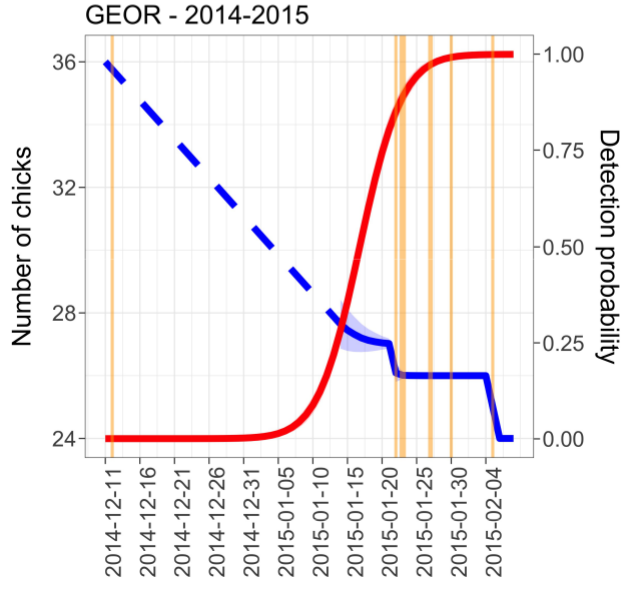

DANC - 2012-2013
Number of chicks
Detection probability
75
70
65
60
1.00
0.75
0.50
0.25
0.00
2012-12-07
2012-12-12
2012-12-17
2012-12-22
2012-12-27
2013-01-01
2013-01-06
2013-01-11
2013-01-16
2013-01-21
2013-01-26
2013-01-31

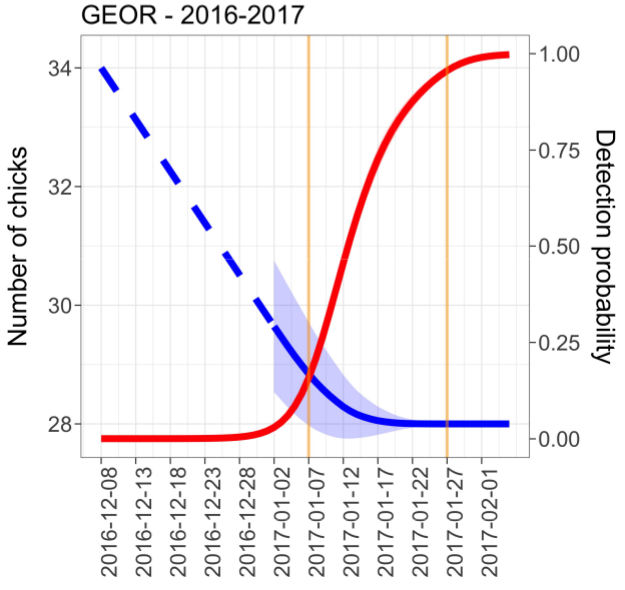

DANC - 2013-2014
Number of chicks
Detection probability
70
65
60
55
50
1.00
0.75
0.50
0.25
0.00
2013-12-12
2013-12-17
2013-12-22
2013-12-27
2014-01-01
2014-01-06
2014-01-11
2014-01-16
2014-01-21
2014-01-26
2014-01-31
2014-02-05

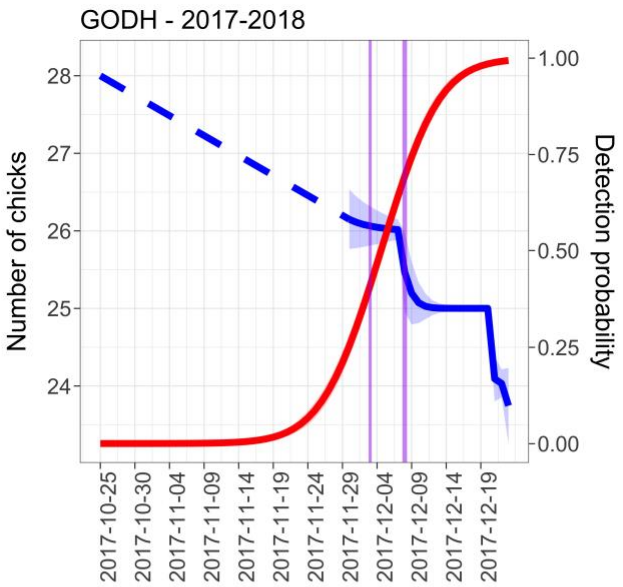

DANC - 2014-2015
Number of chicks
Detection probability
60
55
50
1.00
0.75
0.50
0.25
0.00
2014-12-07
2014-12-12
2014-12-17
2014-12-22
2014-12-27
2015-01-01
2015-01-06
2015-01-11
2015-01-16
2015-01-21
2015-01-26
2015-01-31

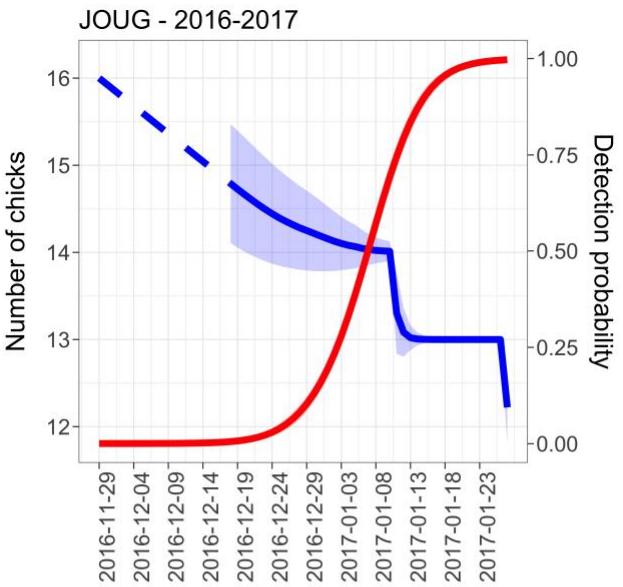

GEOR - 2012-2013
Number of chicks
Detection probability
32
30
28
1.00
0.75
0.50
0.25
0.00
2012-12-08
2012-12-13
2012-12-18
2012-12-23
2012-12-28
2013-01-02
2013-01-07
2013-01-12
2013-01-17
2013-01-22
2013-01-27
2013-02-01

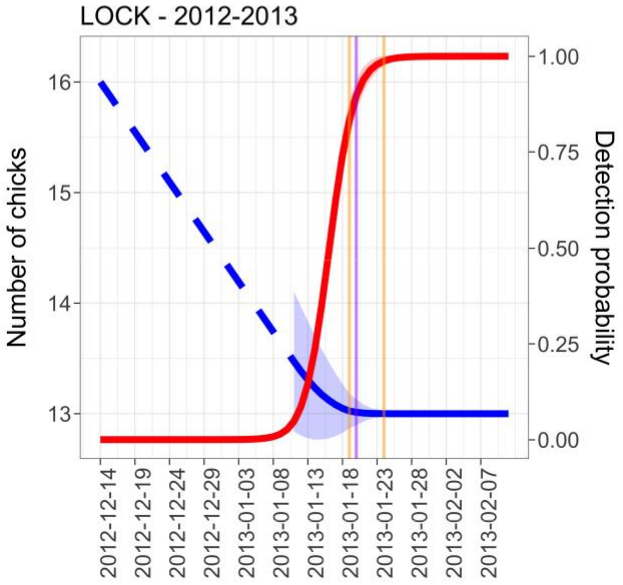

GEOR - 2013-2014
Number of chicks
Detection probability
32
30
28
26
24
1.00
0.75
0.50
0.25
0.00
2013-12-18
2013-12-23
2013-12-28
2014-01-02
2014-01-07
2014-01-12
2014-01-17
2014-01-22
2014-01-27
2014-02-01
2014-02-06
2014-02-11

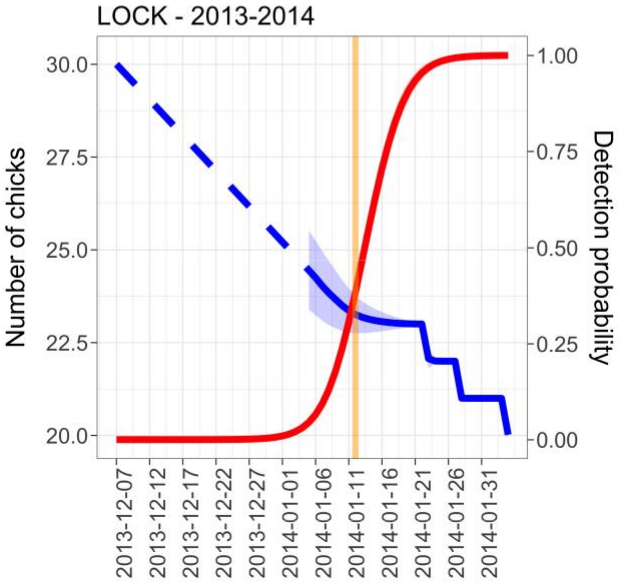

GEOR - 2014-2015
Number of chicks
Detection probability
36
32
28
24
1.00
0.75
0.50
0.25
0.00
2014-12-11
2014-12-16
2014-12-21
2014-12-26
2014-12-31
2015-01-05
2015-01-10
2015-01-15
2015-01-20
2015-01-25
2015-01-30
2015-02-04

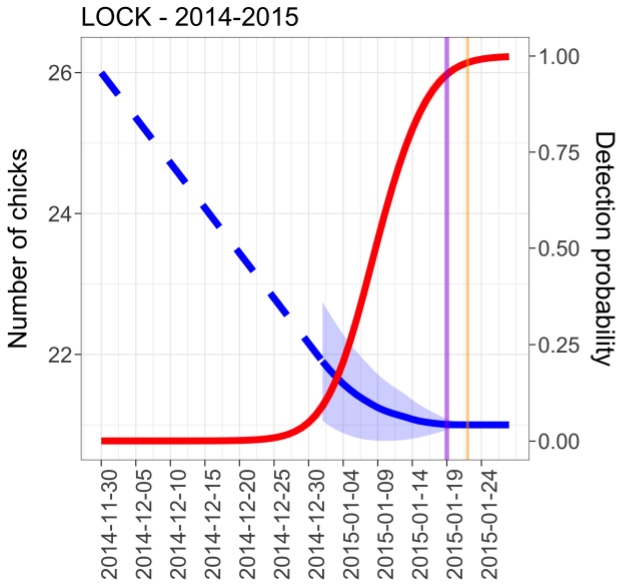

GEOR - 2016-2017
Number of chicks
Detection probability
34
32
30
28
1.00
0.75
0.50
0.25
0.00
2016-12-08
2016-12-13
2016-12-18
2016-12-23
2016-12-28
2017-01-02
2017-01-07
2017-01-12
2017-01-17
2017-01-22
2017-01-27
2017-02-01

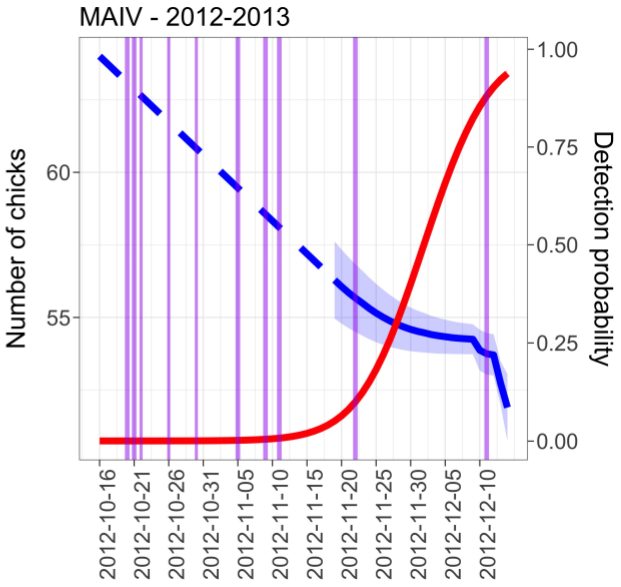

GODH - 2017-2018
Number of chicks
Detection probability
28
27
26
25
24
1.00
0.75
0.50
0.25
0.00
2017-10-25
2017-10-30
2017-11-04
2017-11-09
2017-11-14
2017-11-19
2017-11-24
2017-11-29
2017-12-04
2017-12-09
2017-12-14
2017-12-19

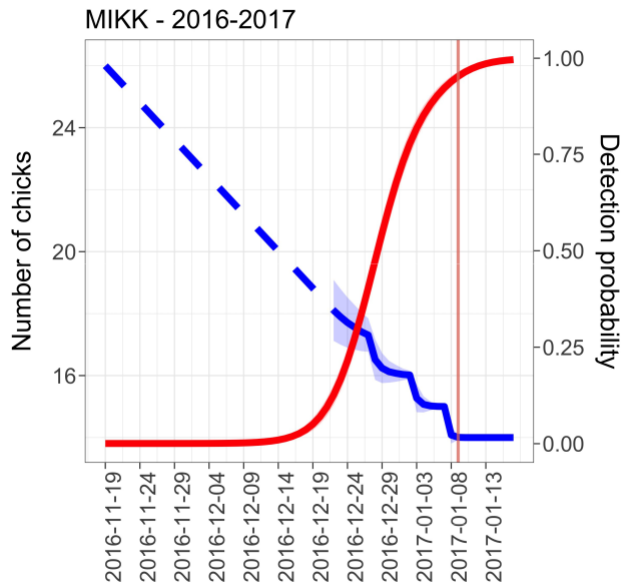

JOUG - 2016-2017
Number of chicks
Detection probability
16
15
14
13
12
1.00
0.75
0.50
0.25
0.00
2016-11-29
2016-12-04
2016-12-09
2016-12-14
2016-12-19
2016-12-24
2016-12-29
2017-01-03
2017-01-08
2017-01-13
2017-01-18
2017-01-23

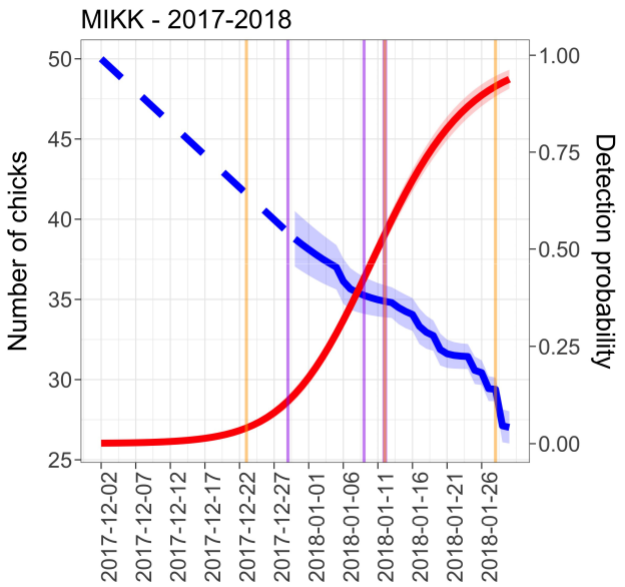

LOCK - 2012-2013
Number of chicks
Detection probability
16
15
14
13
1.00
0.75
0.50
0.25
0.00
2012-12-14
2012-12-19
2012-12-24
2012-12-29
2013-01-03
2013-01-08
2013-01-13
2013-01-18
2013-01-23
2013-01-28
2013-02-02
2013-02-07

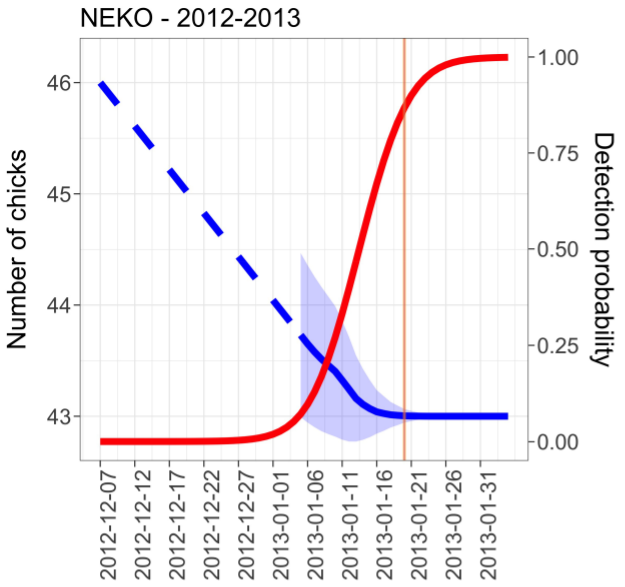

LOCK - 2013-2014
Number of chicks
Detection probability
30.0
27.5
25.0
22.5
20.0
1.00
0.75
0.50
0.25
0.00
2013-12-07
2013-12-12
2013-12-17
2013-12-22
2013-12-27
2014-01-01
2014-01-06
2014-01-11
2014-01-16
2014-01-21
2014-01-26
2014-01-31

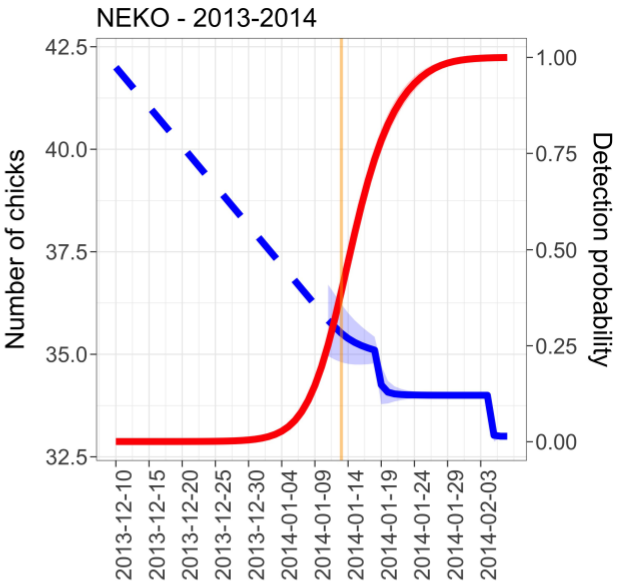

LOCK - 2014-2015
Number of chicks
Detection probability
26
24
22
1.00
0.75
0.50
0.25
0.00
2014-11-30
2014-12-05
2014-12-10
2014-12-15
2014-12-20
2014-12-25
2014-12-30
2015-01-04
2015-01-09
2015-01-14
2015-01-19
2015-01-24

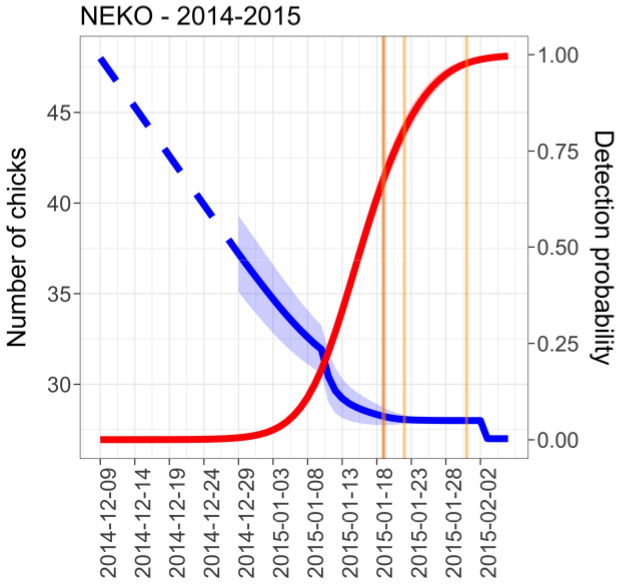

MAIV - 2012-2013
Number of chicks
Detection probability
60
55
1.00
0.75
0.50
0.25
0.00
2012-10-16
2012-10-21
2012-10-26
2012-10-31
2012-11-05
2012-11-10
2012-11-15
2012-11-20
2012-11-25
2012-11-30
2012-12-05
2012-12-10

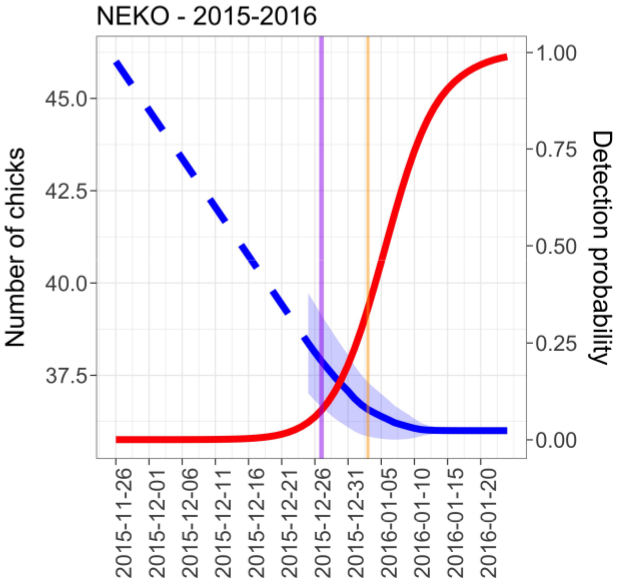

MIKK - 2016-2017
Number of chicks
Detection probability
24
20
16
1.00
0.75
0.50
0.25
0.00
2016-11-19
2016-11-24
2016-11-29
2016-12-04
2016-12-09
2016-12-14
2016-12-19
2016-12-24
2016-12-29
2017-01-03
2017-01-08
2017-01-13


MIKK - 2017-2018
Number of chicks
Detection probability
50
45
40
35
30
25
1.00
0.75
0.50
0.25
0.00
2017-12-02
2017-12-07
2017-12-12
2017-12-17
2017-12-22
2017-12-27
2018-01-01
2018-01-06
2018-01-11
2018-01-16
2018-01-21
2018-01-26


NEKO - 2012-2013
Number of chicks
Detection probability
46
45
44
43
1.00
0.75
0.50
0.25
0.00
2012-12-07
2012-12-12
2012-12-17
2012-12-22
2012-12-27
2013-01-01
2013-01-06
2013-01-11
2013-01-16
2013-01-21
2013-01-26
2013-01-31


NEKO - 2013-2014
Number of chicks
Detection probability
42.5
40.0
37.5
35.0
32.5
1.00
0.75
0.50
0.25
0.00
2013-12-10
2013-12-15
2013-12-20
2013-12-25
2013-12-30
2014-01-04
2014-01-09
2014-01-14
2014-01-19
2014-01-24
2014-01-29
2014-02-03


NEKO - 2014-2015
Number of chicks
Detection probability
45
40
35
30
1.00
0.75
0.50
0.25
0.00
2014-12-09
2014-12-14
2014-12-19
2014-12-24
2014-12-29
2015-01-03
2015-01-08
2015-01-13
2015-01-18
2015-01-23
2015-01-28
2015-02-02


NEKO - 2015-2016
Number of chicks
Detection probability
45.0
42.5
40.0
37.5
1.00
0.75
0.50
0.25
0.00
2015-11-26
2015-12-01
2015-12-06
2015-12-11
2015-12-16
2015-12-21
2015-12-26
2015-12-31
2016-01-05
2016-01-10
2016-01-15
2016-01-20

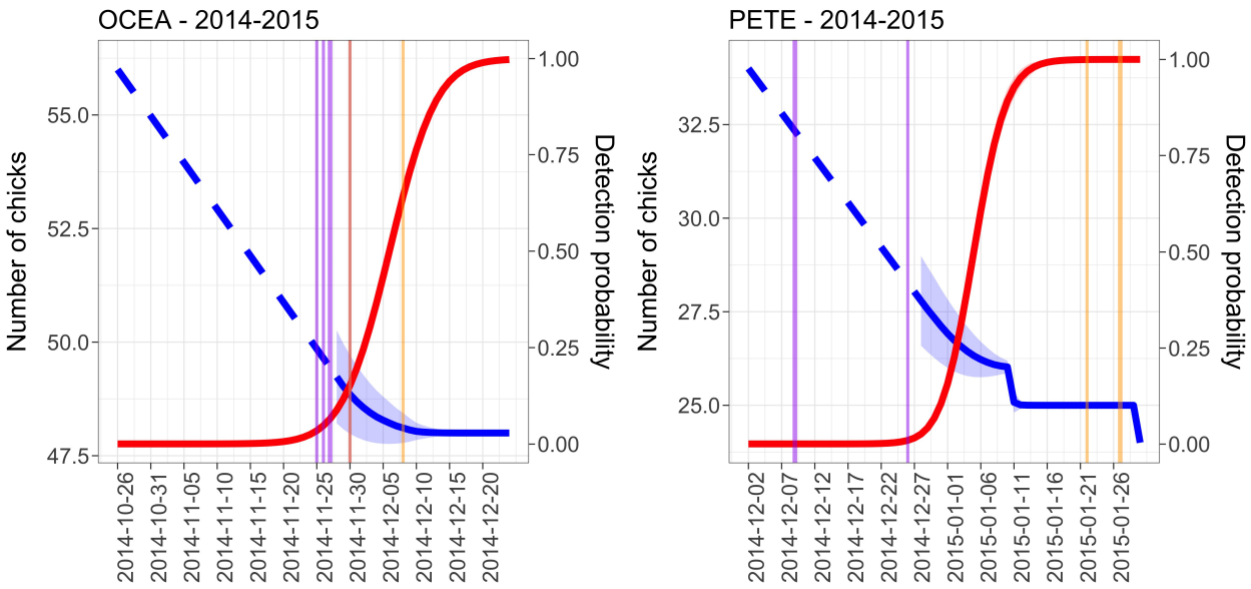
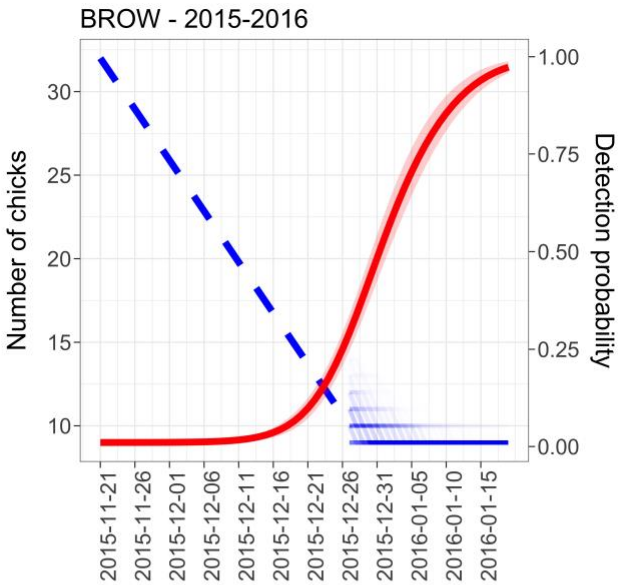
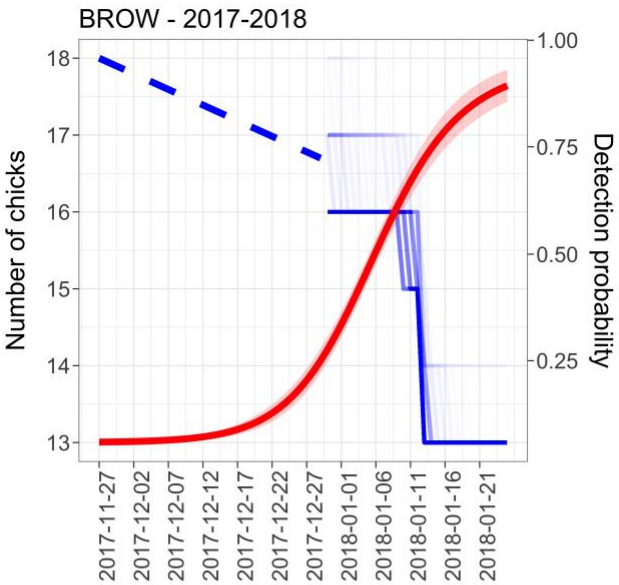
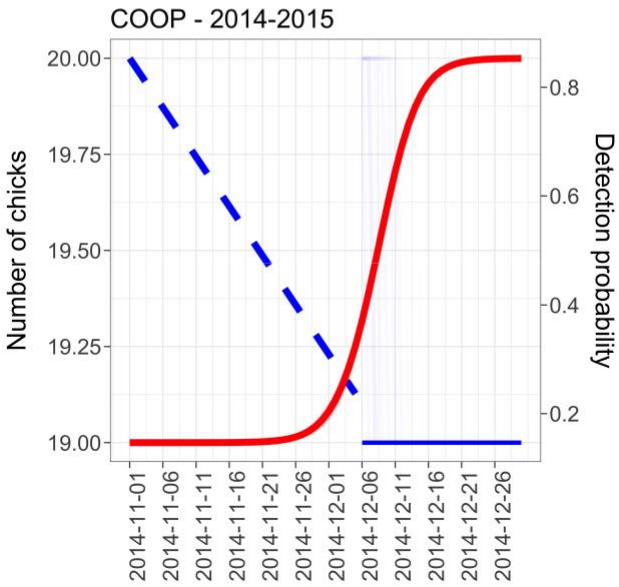


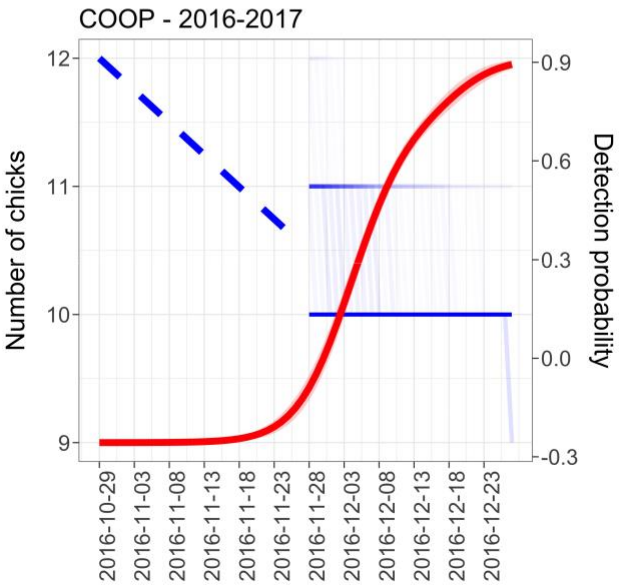
Figure S1-5: Number of chicks alive at each time point in the season for each site/year. The blue dashed lined represents the transition from the extrapolated egg lay date to the date the first penguin chick was recorded in an image. The solid blue line is the latent state *Z*, the total number of chicks at a given site/year at any given time point. The blue ribbon represents one standard deviation around *Z*. The red line is the mean detection probability at each time step across all nests for that site $\bar{p}$. The red ribbon represents one standard deviation around $\bar{p}$ (though the ribbon not visible in most cases due to the small degree of uncertainty for $\bar{p}$). Snow and rain events are represented by purple and orange vertical lines, respectively. BROW - Brown Bluff, COOP - Cooper Bay, CUVE - Cuverville Island, DAMO - Damoy Point, DANC - Danco Island, GEOR - Georges Point, GODH - Godthul, JOUG - Jougla Point, LOCK - Port Lockroy, MAIV - Maiviken, MIKK - Mikkelsen Harbor, NEKO - Neko Harbor, OCEA - Ocean Harbor, PETE - Petermann Island.

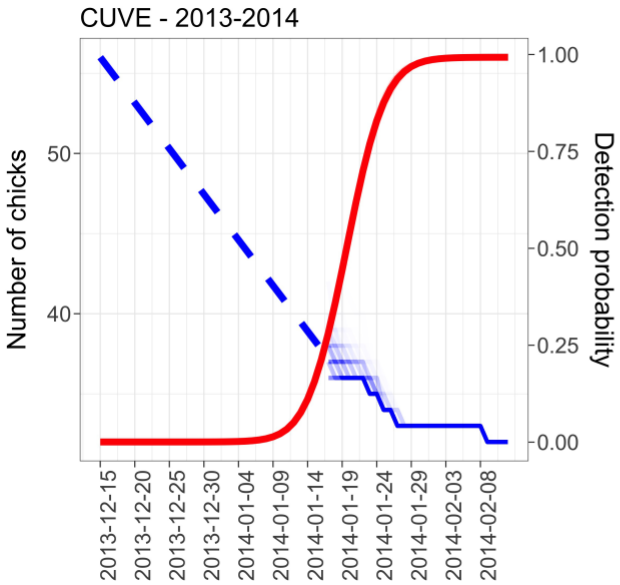
Alternative visualization of changes in chicks over time

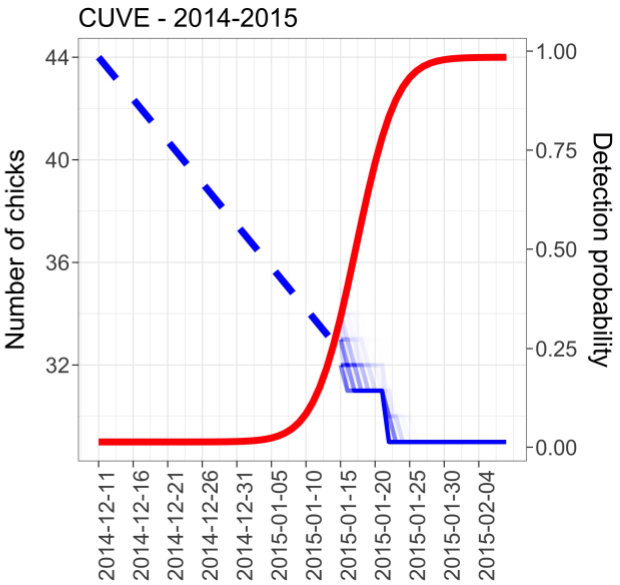


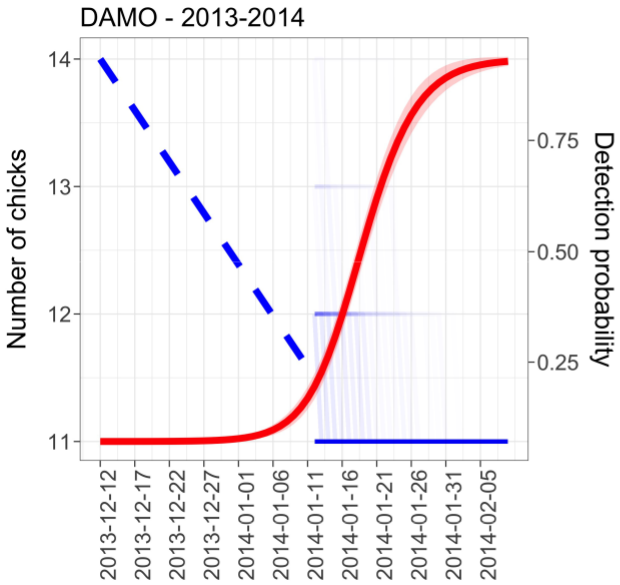


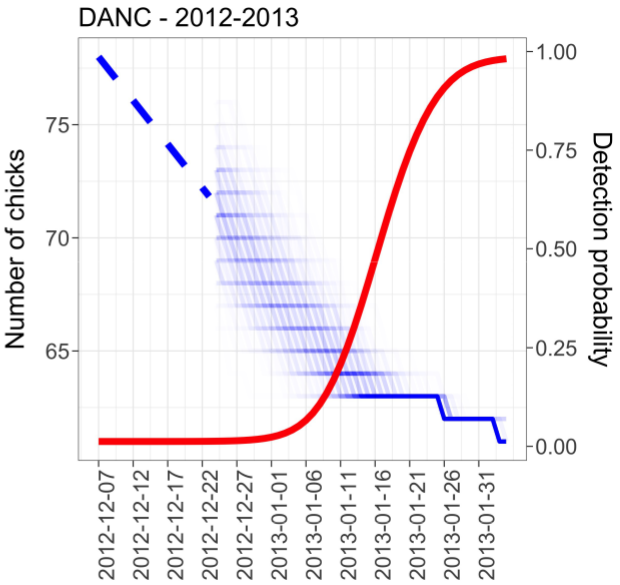


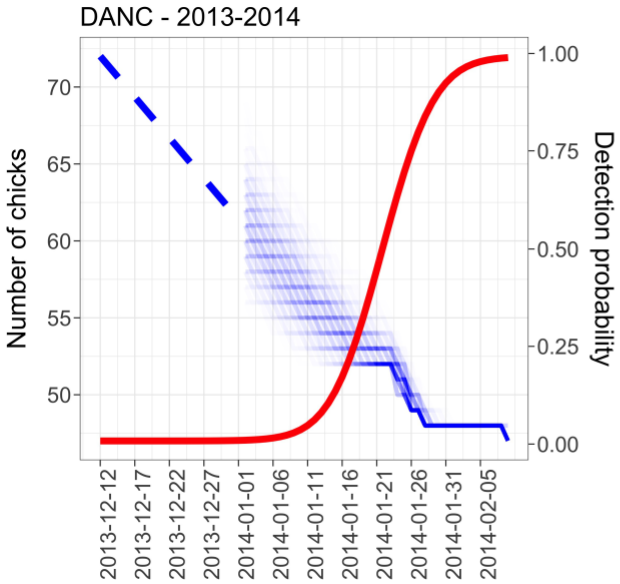


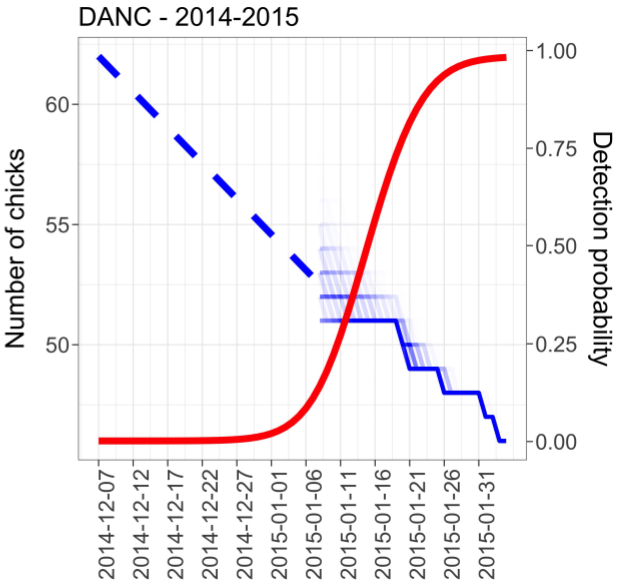

CUVE - 2013-2014
Number of chicks
Detection probability
50
40
1.00
0.75
0.50
0.25
0.00
2013-12-15
2013-12-20
2013-12-25
2013-12-30
2014-01-04
2014-01-09
2014-01-14
2014-01-19
2014-01-24
2014-01-29
2014-02-03
2014-02-08

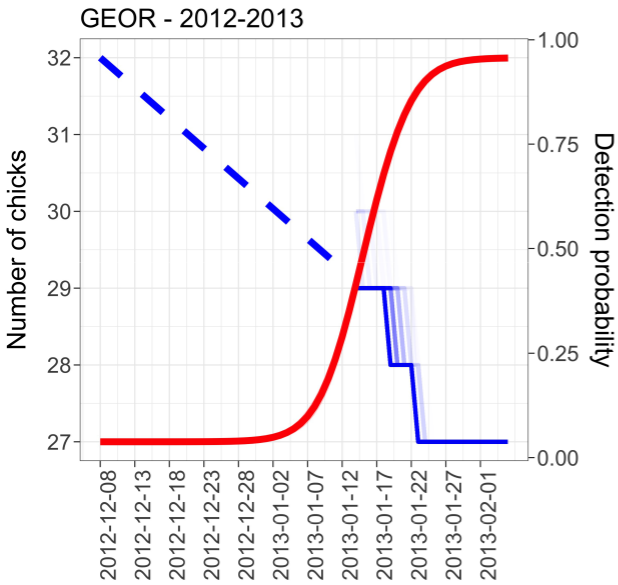

CUVE - 2014-2015
Number of chicks
Detection probability
44
40
36
32
1.00
0.75
0.50
0.25
0.00
2014-12-11
2014-12-16
2014-12-21
2014-12-26
2014-12-31
2015-01-05
2015-01-10
2015-01-15
2015-01-20
2015-01-25
2015-01-30
2015-02-04

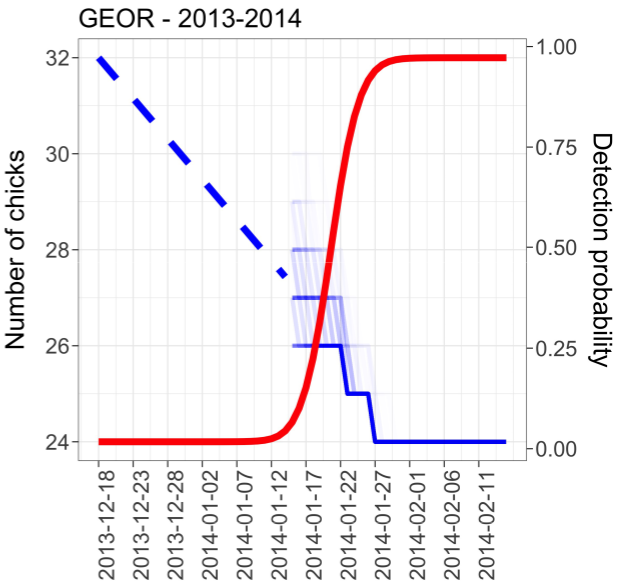

DAMO - 2013-2014
Number of chicks
Detection probability
14
13
12
11
0.75
0.50
0.25
2013-12-12
2013-12-17
2013-12-22
2013-12-27
2014-01-01
2014-01-06
2014-01-11
2014-01-16
2014-01-21
2014-01-26
2014-01-31
2014-02-05

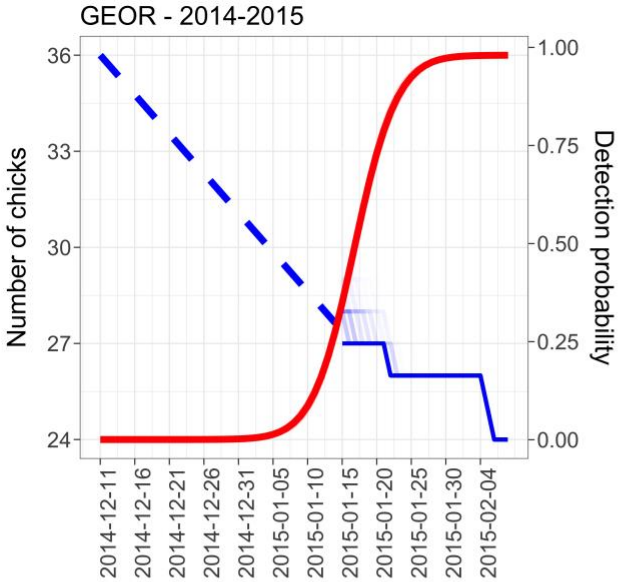

DANC - 2012-2013
Number of chicks
Detection probability
75
70
65
1.00
0.75
0.50
0.25
0.00
2012-12-07
2012-12-12
2012-12-17
2012-12-22
2012-12-27
2013-01-01
2013-01-06
2013-01-11
2013-01-16
2013-01-21
2013-01-26
2013-01-31

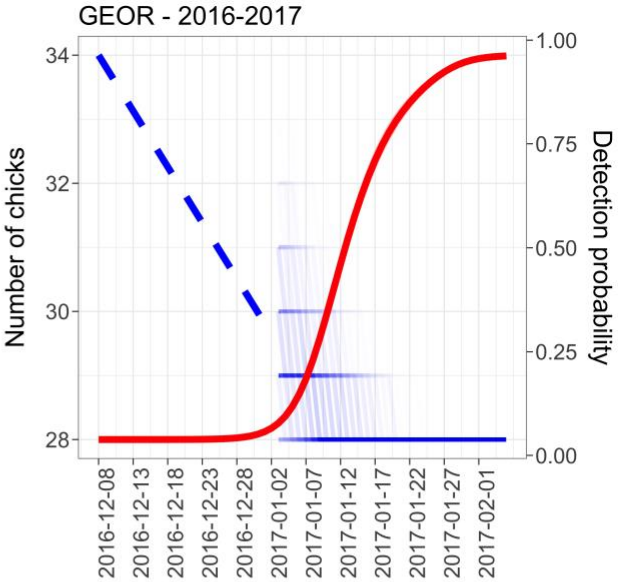

DANC - 2013-2014
Number of chicks
Detection probability
70
65
60
55
50
1.00
0.75
0.50
0.25
0.00
2013-12-12
2013-12-17
2013-12-22
2013-12-27
2014-01-01
2014-01-06
2014-01-11
2014-01-16
2014-01-21
2014-01-26
2014-01-31
2014-02-05

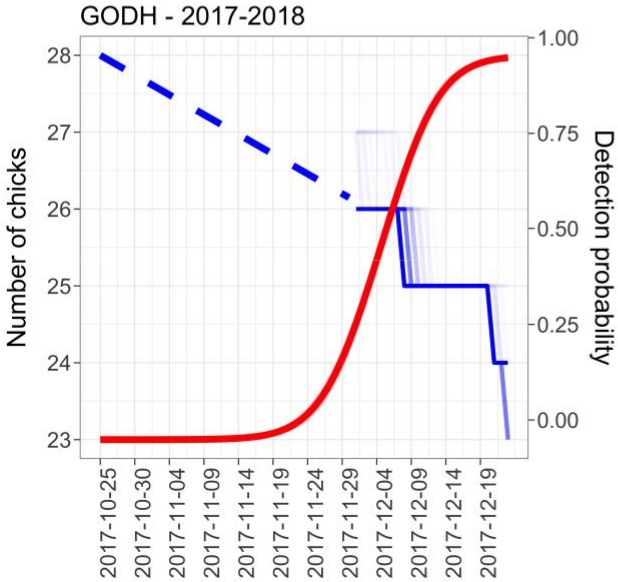

DANC - 2014-2015
Number of chicks
Detection probability
60
55
50
1.00
0.75
0.50
0.25
0.00
2014-12-07
2014-12-12
2014-12-17
2014-12-22
2014-12-27
2015-01-01
2015-01-06
2015-01-11
2015-01-16
2015-01-21
2015-01-26
2015-01-31

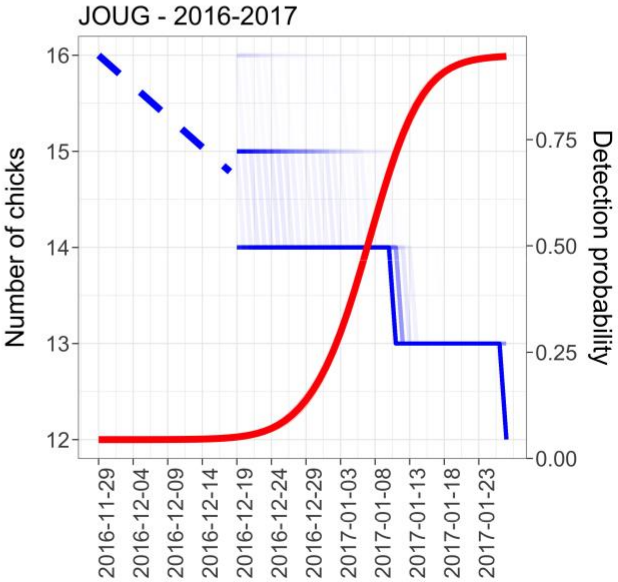

GEOR - 2012-2013
Number of chicks
Detection probability
32
31
30
29
28
27
1.00
0.75
0.50
0.25
0.00
2012-12-08
2012-12-13
2012-12-18
2012-12-23
2012-12-28
2013-01-02
2013-01-07
2013-01-12
2013-01-17
2013-01-22
2013-01-27
2013-02-01

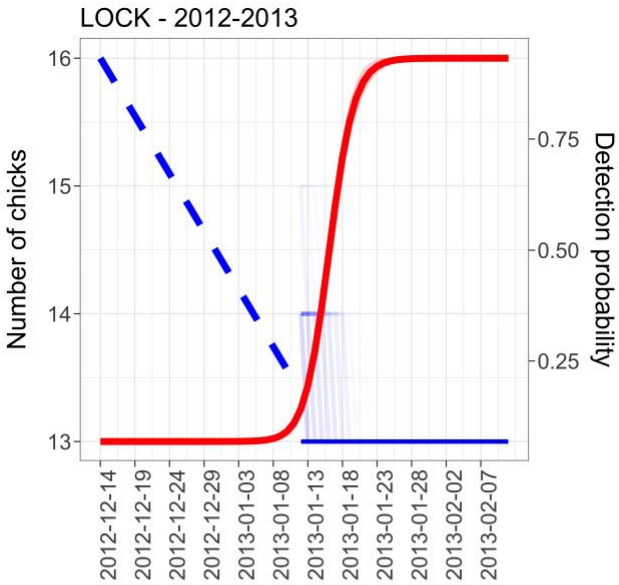

GEOR - 2013-2014
Number of chicks
Detection probability
32
30
28
26
24
1.00
0.75
0.50
0.25
0.00
2013-12-18
2013-12-23
2013-12-28
2014-01-02
2014-01-07
2014-01-12
2014-01-17
2014-01-22
2014-01-27
2014-02-01
2014-02-06
2014-02-11

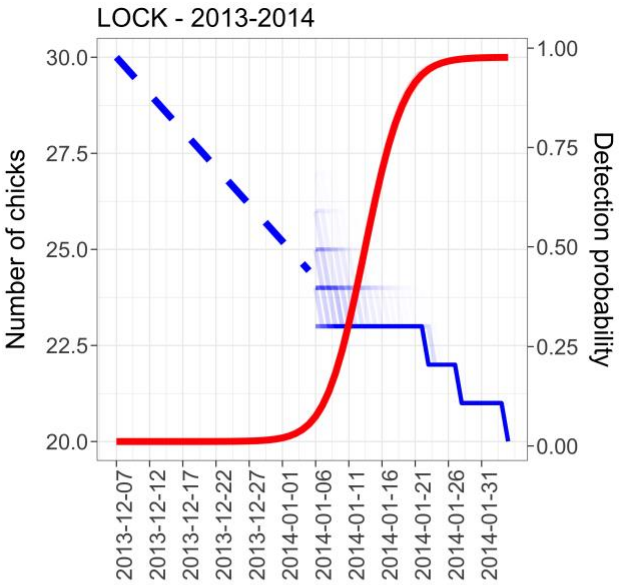

GEOR - 2014-2015
Number of chicks
Detection probability
36
33
30
27
24
1.00
0.75
0.50
0.25
0.00
2014-12-11
2014-12-16
2014-12-21
2014-12-26
2014-12-31
2015-01-05
2015-01-10
2015-01-15
2015-01-20
2015-01-25
2015-01-30
2015-02-04

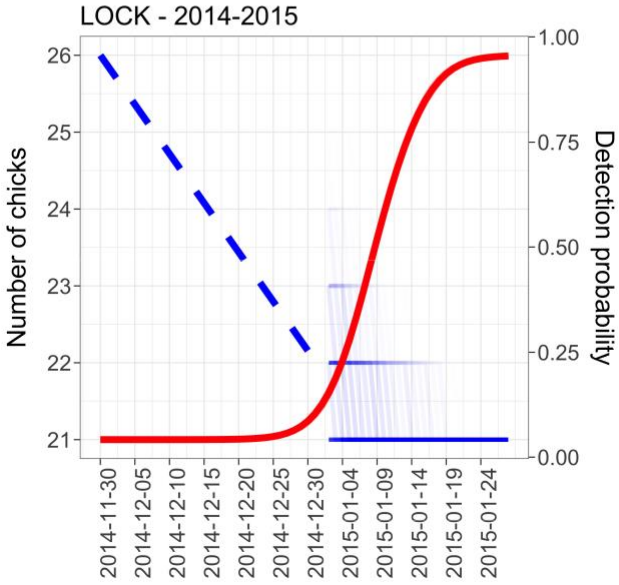

GEOR - 2016-2017
Number of chicks
Detection probability
34
32
30
28
1.00
0.75
0.50
0.25
0.00
2016-12-08
2016-12-13
2016-12-18
2016-12-23
2016-12-28
2017-01-02
2017-01-07
2017-01-12
2017-01-17
2017-01-22
2017-01-27
2017-02-01

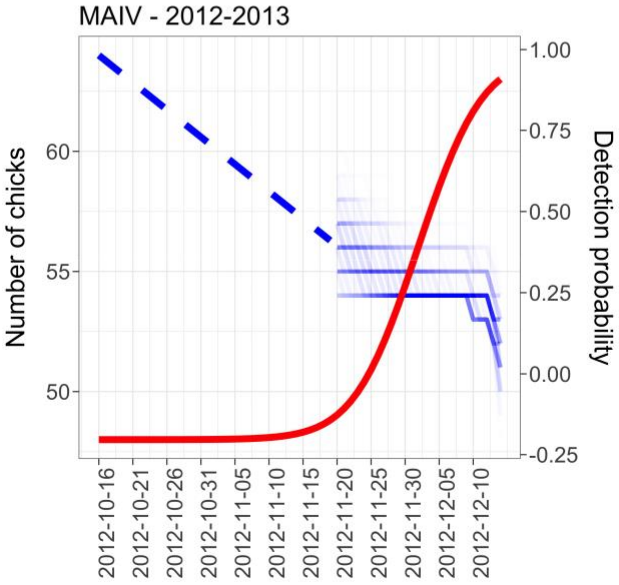

GODH - 2017-2018
Number of chicks
Detection probability
28
27
26
25
24
23
1.00
0.75
0.50
0.25
0.00
2017-10-25
2017-10-30
2017-11-04
2017-11-09
2017-11-14
2017-11-19
2017-11-24
2017-11-29
2017-12-04
2017-12-09
2017-12-14
2017-12-19

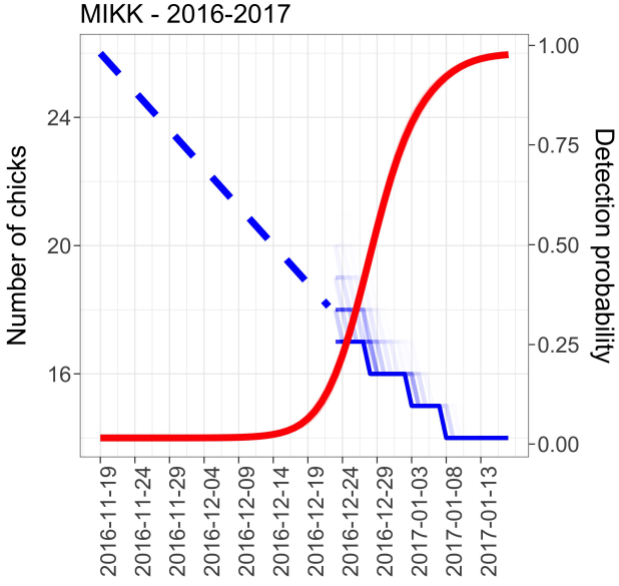

JOUG - 2016-2017
Number of chicks
Detection probability
16
15
14
13
12
0.75
0.50
0.25
0.00
2016-11-29
2016-12-04
2016-12-09
2016-12-14
2016-12-19
2016-12-24
2016-12-29
2017-01-03
2017-01-08
2017-01-13
2017-01-18
2017-01-23

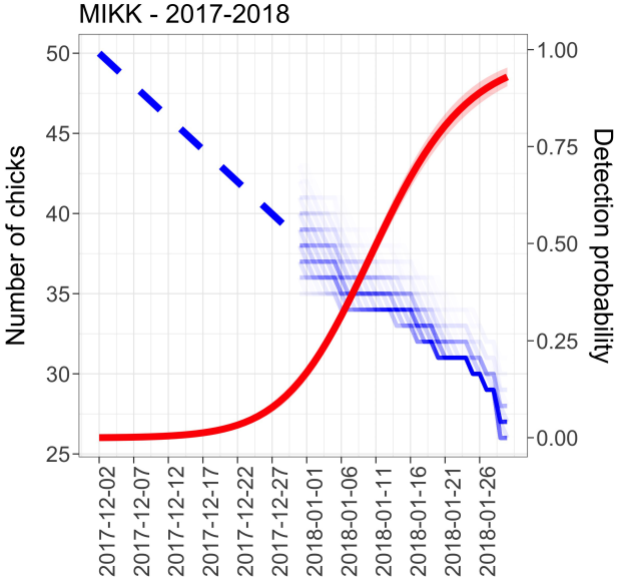

LOCK - 2012-2013
Number of chicks
Detection probability
16
15
14
13
0.75
0.50
0.25
2012-12-14
2012-12-19
2012-12-24
2012-12-29
2013-01-03
2013-01-08
2013-01-13
2013-01-18
2013-01-23
2013-01-28
2013-02-02
2013-02-07

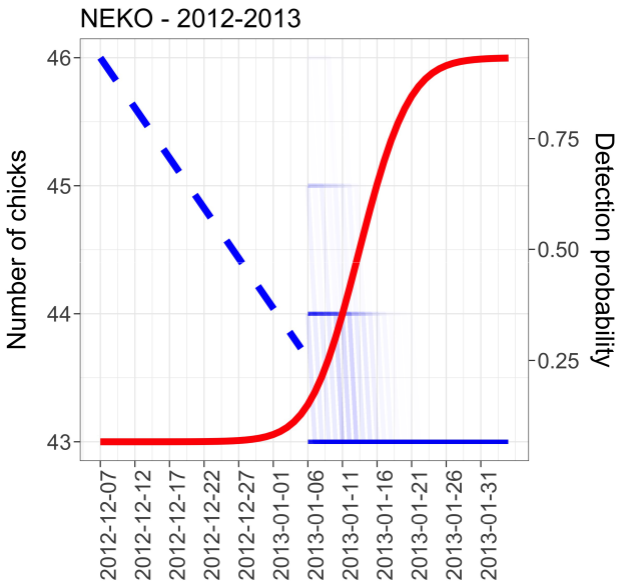

LOCK - 2013-2014
Number of chicks
Detection probability
30.0
27.5
25.0
22.5
20.0
1.00
0.75
0.50
0.25
0.00
2013-12-07
2013-12-12
2013-12-17
2013-12-22
2013-12-27
2014-01-01
2014-01-06
2014-01-11
2014-01-16
2014-01-21
2014-01-26
2014-01-31

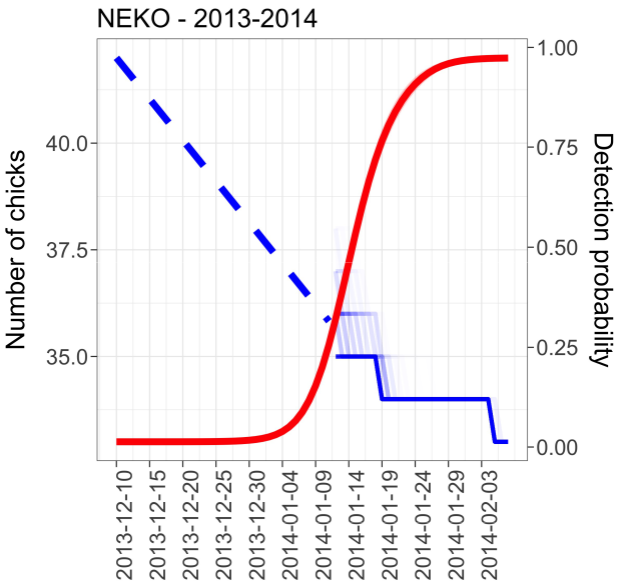

LOCK - 2014-2015
Number of chicks
Detection probability
26
25
24
23
22
21
1.00
0.75
0.50
0.25
0.00
2014-11-30
2014-12-05
2014-12-10
2014-12-15
2014-12-20
2014-12-25
2014-12-30
2015-01-04
2015-01-09
2015-01-14
2015-01-19
2015-01-24

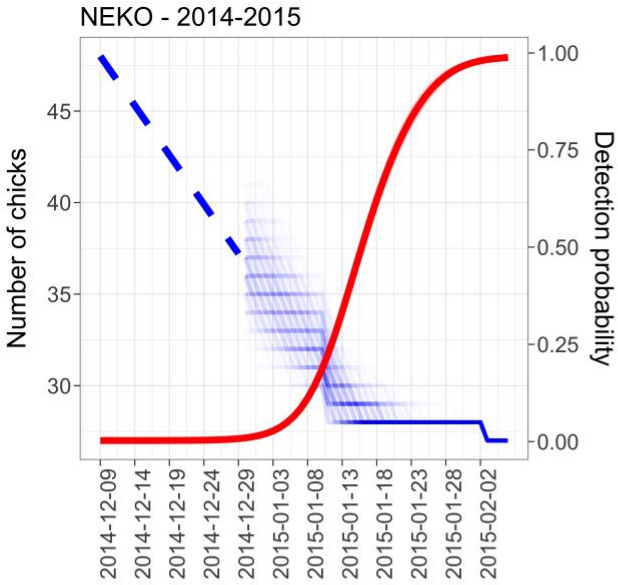

MAIV - 2012-2013
Number of chicks
Detection probability
60
55
50
1.00
0.75
0.50
0.25
0.00
-0.25
2012-10-16
2012-10-21
2012-10-26
2012-10-31
2012-11-05
2012-11-10
2012-11-15
2012-11-20
2012-11-25
2012-11-30
2012-12-05
2012-12-10

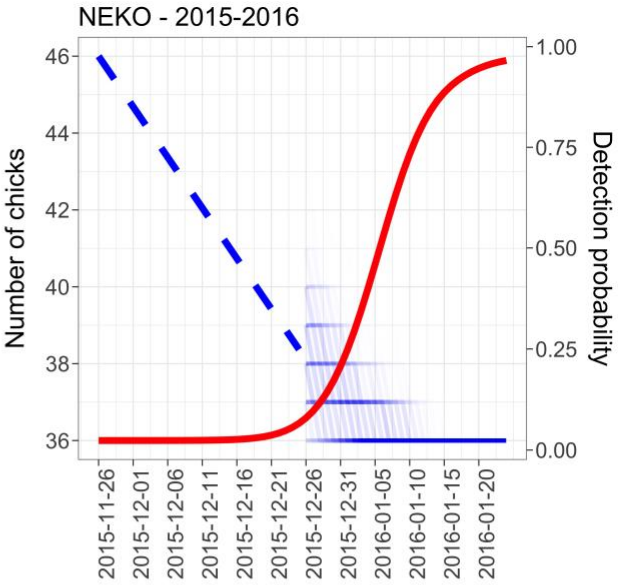

MIKK - 2016-2017
Number of chicks
Detection probability
24
20
16
1.00
0.75
0.50
0.25
0.00
2016-11-19
2016-11-24
2016-11-29
2016-12-04
2016-12-09
2016-12-14
2016-12-19
2016-12-24
2016-12-29
2017-01-03
2017-01-08
2017-01-13

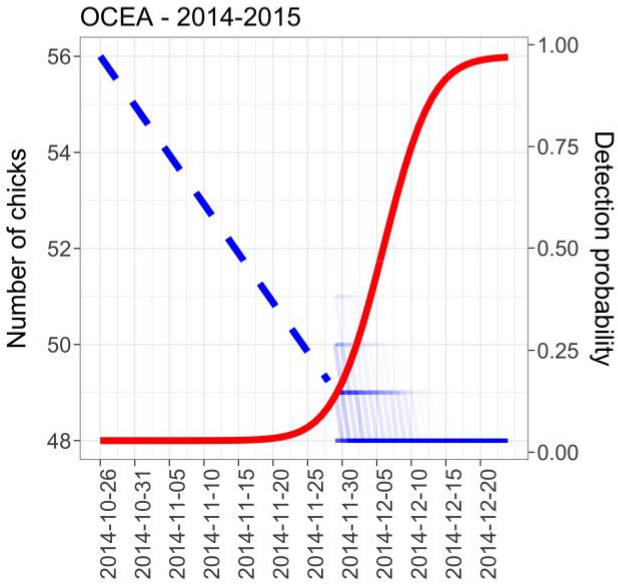

MIKK - 2017-2018
Number of chicks
Detection probability
50
45
40
35
30
25
1.00
0.75
0.50
0.25
0.00
2017-12-02
2017-12-07
2017-12-12
2017-12-17
2017-12-22
2017-12-27
2018-01-01
2018-01-06
2018-01-11
2018-01-16
2018-01-21
2018-01-26


NEKO - 2012-2013
Number of chicks
Detection probability
46
45
44
43
0.75
0.50
0.25
2012-12-07
2012-12-12
2012-12-17
2012-12-22
2012-12-27
2013-01-01
2013-01-06
2013-01-11
2013-01-16
2013-01-21
2013-01-26
2013-01-31


NEKO - 2013-2014
Number of chicks
Detection probability
40.0
37.5
35.0
1.00
0.75
0.50
0.25
0.00
2013-12-10
2013-12-15
2013-12-20
2013-12-25
2013-12-30
2014-01-04
2014-01-09
2014-01-14
2014-01-19
2014-01-24
2014-01-29
2014-02-03


NEKO - 2014-2015
Number of chicks
Detection probability
45
40
35
30
1.00
0.75
0.50
0.25
0.00
2014-12-09
2014-12-14
2014-12-19
2014-12-24
2014-12-29
2015-01-03
2015-01-08
2015-01-13
2015-01-18
2015-01-23
2015-01-28
2015-02-02


NEKO - 2015-2016
Number of chicks
Detection probability
46
44
42
40
38
36
1.00
0.75
0.50
0.25
0.00
2015-11-26
2015-12-01
2015-12-06
2015-12-11
2015-12-16
2015-12-21
2015-12-26
2015-12-31
2016-01-05
2016-01-10
2016-01-15
2016-01-20


OCEA - 2014-2015
Number of chicks
Detection probability
56
54
52
50
48
1.00
0.75
0.50
0.25
0.00
2014-10-26
2014-10-31
2014-11-05
2014-11-10
2014-11-15
2014-11-20
2014-11-25
2014-11-30
2014-12-05
2014-12-10
2014-12-15
2014-12-20

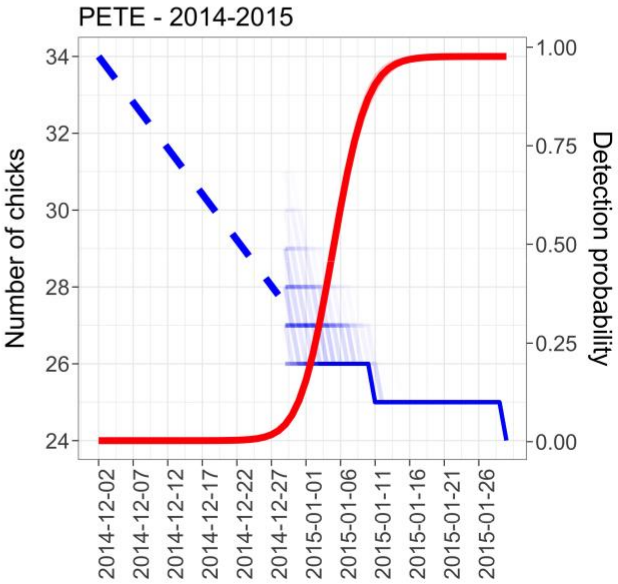

PETE - 2014-2015
Number of chicks
Detection probability
34
32
30
28
26
24
1.00
0.75
0.50
0.25
0.00
2014-12-02
2014-12-07
2014-12-12
2014-12-17
2014-12-22
2014-12-27
2015-01-01
2015-01-06
2015-01-11
2015-01-16
2015-01-21
2015-01-26

Figure S1-6: An alternative visualization (to Fig. S1-5) of the number of chicks alive at each time point in the season for each site/year. The blue dashed lined represents the transition from the extrapolated egg lay date to the date the first penguin chick was recorded in an image. Each solid blue line represents one realization (i.e. iteration of the posterior chain) of the latent state *Z*, the total number of chicks at a given site/year at any given time point. A total of 500 realizations are displayed. The red line is the mean detection probability at each time step across all nests for that site $\bar{p}$. The red ribbon represents one standard deviation around $\bar{p}$ (though the ribbon not visible in most cases due to the small degree of uncertainty for $\bar{p}$). BROW - Brown Bluff, COOP - Cooper Bay, CUVE - Cuverville Island, DAMO - Damoy Point, DANC - Danco Island, GEOR - Georges Point, GODH - Godthul, JOUG - Jougla Point, LOCK - Port Lockroy, MAIV - Maiviken, MIKK - Mikkelsen Harbor, NEKO - Neko Harbor, OCEA - Ocean Harbor, PETE - Petermann Island.

Appendix 7b: Chapter 5 – Supplementary Materials – Part 2 (S2)

Processing and analysis of krill catch data

Large-scale assessment of intra- and inter-annual nesting success using a remote camera network

Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR) Small-Scale Management Units (SSMUs)

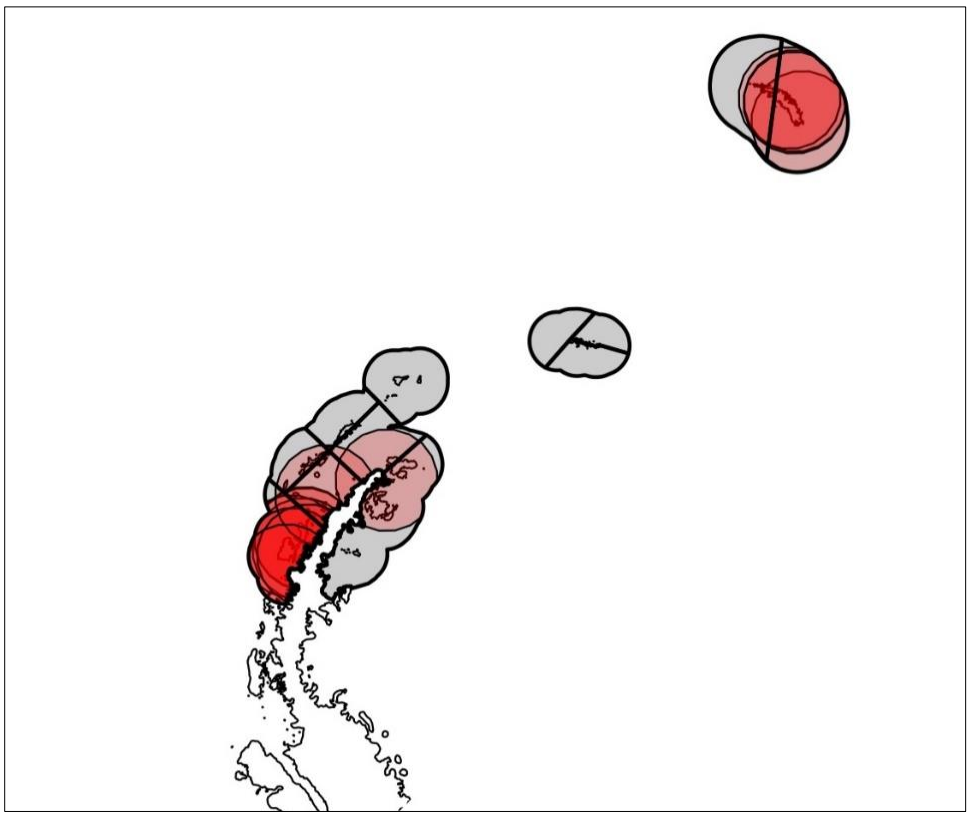


Figure S2-1: Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR) Small-Scale Management Units (SSMUs) are represented as grey shaded regions. The 150 km buffers around each site used in this study are represented by red circles.

Effect of krill fishing on penguin breeding success during the breeding season

Analyses presented in the main text used the sum of the weighted average krill catch from March (of year t-1) through January (year t), representing all krill caught at a site prior to and during the breeding season. Krill catch over just the breeding season (December - January) was also calculated using a 25km buffer to test for the effect of krill extraction during the breeding season in areas where Gentoo penguins may be foraging during this time period.

Breeding success was modelled as a function of krill catch during the breeding season,

$$bs_{obs_i} \sim N\left(bs_{true_i}, \sigma_{bs_i}\right)$$

$$bs_{true_i} \sim N(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta \times x_i$$

where bs_{obs} is the estimated breeding success (posterior mean of the derived quantity from the capture-recapture model) for a given site/year, σ_{obs} is the uncertainty about the estimate of breeding success for each site/year (posterior standard deviation of the derived quantity; this is given rather than estimated by the model), bs_{true} is the latent true state of breeding success, x is the covariate of interest for site/year i, α is the intercept parameter, β is the slope parameter, and σ is the residual error term.

Models were fit using the R package 'rjags' to interface with JAGS in the R statistical environment. Inferences were obtained from 10,000 samples drawn from four chains, following a 'burn-in' period of 10,000 draws and an adaptation phase of 5000 draws. All chains unambiguously converged (Rhat < 1.1) and effective samples sizes were sufficiently

large (ess / chain > 100). Results are similar to those presented in the main text, showing no clear relationship between breeding success and krill catch.

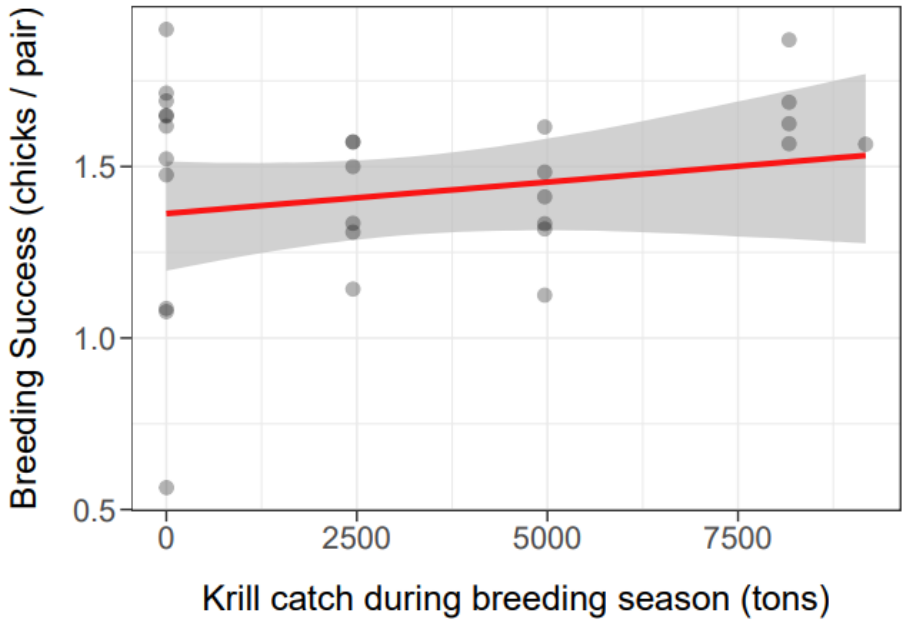


Figure S2-2: Breeding success as a function of krill catch during the penguin breeding season. The red line represents the posterior median. The grey ribbon represents the 95% credible interval. Median slope (β) = 1.8×10^{-5} [95% CI: -1.8×10^{-5} - 5.4×10^{-5}].

Temporal distribution of krill fishing effort

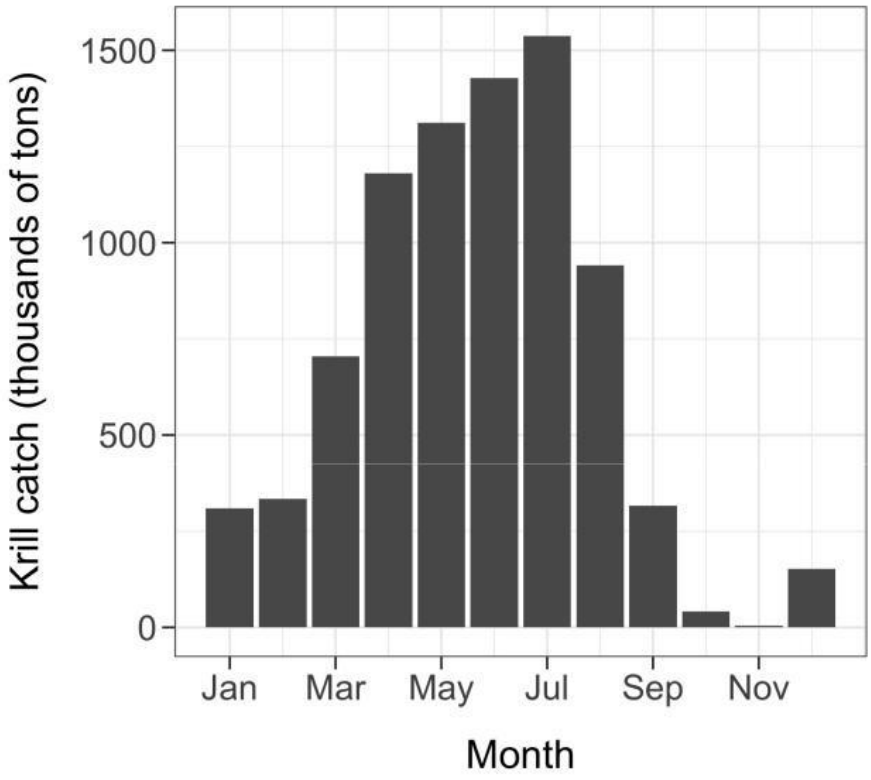


Figure S2-3: Total monthly reported krill catch over the course of the calendar year for all sites in this study. Monthly totals were averaged across all focal years (2013-2018). Krill fishing at these sites (within the 150km buffers) is highest during the overwinter period.

Appendix 8: *Penguin Watch* as a Tool for Public Engagement

Citizen science is an effective tool for outreach and education, actively engaging members of the public who may have no formal scientific training (Bonney et al., 2009; Dickinson et al., 2012).

Since beginning work on the *Penguin Watch* project, I have been contacted by countless school teachers, parents and educators, all of whom either expressed an interest in finding out more about *Penguin Watch*, or wanted to communicate how much their classes or children had enjoyed taking part. Prompted by this, I compiled a digital information pack for schools, comprising *Penguin Watch* images, video, project information and data, which can be accessed from the 'Education' page of the project website:

https://www.zooniverse.org/projects/penguintom79/penguin-watch/about/education

In 2017 we were selected as the featured project for British Science Week, and reached our target of 250,000 online classifications for the week by the Tuesday.

It should be noted that interest and participation is not restricted to young people, and *Penguin Watch* hosts an active 'Talk' forum, where questions and comments about observed behaviours are frequently posted and discussed. Since the launch of *Penguin Watch* on 17th September 2014, there have been 115,664 comments on the 'Talk' forum, by ~6000 different volunteers, and over 50,000 volunteers have taken part in the project.

The *Penguin Watch* team is actively engaged in outreach activities both within Oxford and further afield, and in 2017 we were delighted to receive one of the Vice-Chancellor's Public

Engagement with Research Awards, for ‘Collaboration and Communication with the Public’.

References

Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009). Citizen science: a developing tool for expanding science knowledge and scientific literacy. *BioScience*, *59*(11), 977–984. https://doi.org/10.1525/bio.2009.59.11.9

Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., … Purcell, K. (2012). The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment*, *10*(6), 291–297. https://doi.org/10.1890/110236