# Prioritising conservation areas in Southeast Asia: highlighting zoological biodiversity hotspots and assessing spatiotemporal patterns of the felids guild.




Luca Chiaverini

Lady Margaret Hall

University of Oxford

A thesis submitted for the degree of

*Doctor of Philosophy*

Michaelmas Term 2022

**Prioritising conservation areas in Southeast Asia: highlighting zoological biodiversity hotspots and assessing spatiotemporal patterns of the felids guild.**


## Abstract

Biodiversity is declining globally, at rates comparable with the five previous mass extinctions of Earth's history. The causes of the current sixth mass extinction are to be found in the effects of human population and economic growth on natural resources. Growth of human population determines increases of the resources required, which are provided by anthropic enterprises. These, in turn, lead to direct exploitation of habitats and species, land use degradation and climate change. Southeast Asia is one of most biodiverse region, but it has also experienced extraordinary human population and economic growths in the last decades, leading to pristine habitat loss and deforestation rates among the highest globally. Conservation actions are therefore urgent but, for them to have effective outcomes, and considering the general paucity of resources for conservation, it is important to prioritise the most suitable areas in order to provide precise and reliable indications about where conservation actions can be most effective. Southeast Asia encompasses four of the biodiversity hotspots identified by Myers on the basis of endemism and natural habitat loss. Specifically, biodiversity hotspots have originally been identified as those global areas containing at least 0.5% or 1,500 of the World's vascular plant species as endemic and that have lost at least 70% of their primary vegetation. However, these hotspots are at a scale too coarse for prompting direct and focused conservation actions. Instead, it is necessary to highlight, within the hotspots, the regions that effectively harbour the highest numbers of species richness, which have been defined as "hotspots within hotspots" or "micro-hotspots".

In Chapter 2 of my thesis, by assessing the multivariate relationships between a vast suite of Southeast Asian terrestrial vertebrates and ecologically meaningful habitat factors, I mapped the most biodiverse areas in mainland Southeast Asia. Similarly, in Chapter 3 I evaluated the multivariate relationships with habitat factors for species sampled in the islands of Borneo and Sumatra. I highlighted that, within this vast area that encompasses four biodiversity hotspots and that extends for almost 5,000,000 $km^2$, only a few smaller regions are characterised by high levels of biodiversity, and even smaller biodiverse areas are scattered throughout disturbed landscapes. Large, biodiverse regions are mostly associated with pristine landscapes, while anthropic factors were the main predictors of low levels of species richness. In the continental extents of Southeast Asia, the most biodiverse areas occurred in the mountainous region of Southwest China and in the Thai-Malay Peninsula, both characterised by an extent ranging between 250 and 300 $km^2$, providing a more precise distribution of the terrestrial vertebrate biodiversity within the original biodiversity hotspots that, otherwise, have a total extent of almost 3,500,000 $km^2$ in the mainland. In Borneo the areas of highest terrestrial vertebrate richness have been mapped mainly in the Malaysian state of Sabah, and in Sumatra along the Barisan Mountains, with an extent, respectively, of approximately 180 $km^2$ and 150 $km^2$, within the Sundaland biodiversity hotspot characterised by an area of 1,500,000 $km^2$. Then, focusing solely on the medium- and small-sized felid species, of which Southeast Asia represents a centre of endemism, in Chapter 4 I tested different algorithms to model habitat suitability for this often understudied group of species. Then, I compared the differences and the predictive performances of the varying modelling algorithms, finding that there were large and systematic differences in the modelled probability layers, irrespective of the underlying ecological relationships, but driven by the choice of the algorithm. In Chapter 5, by using the most supported model for each species modelled in the previous chapter, I produced multi-species predictive surfaces for Southeast

Asian small felids. I highlighted that the most suitable habitats for this group of species showed often pronounced differences with the hotspots for terrestrial vertebrate biodiversity modelled in the first two chapters of this thesis, likely due to the fact that small felids are more specialised than the pool of species used to model biodiversity hotspots, and they require larger and more intact habitats. In mainland Southeast Asia, I predicted the hotspots for smaller felids to occur in the Northern Forest Complex of Myanmar, and in the extensive forested area between Myanmar and Laos. In Borneo, the most suitable habitats for small felids occurred in the central highlands of the island, while in Sumatra it occurred along the Barisan Mountains. Together, my findings provide a useful tool to help preserving the unique and endangered Southeast Asia biodiversity. By highlighting the main habitat factors driving the simultaneous presence of terrestrial vertebrate biodiversity, as well as the habitat suitability for individual species of small felids, I provided useful insights for habitat restoration and protection. Additionally, by mapping the areas harbouring the highest numbers of terrestrial vertebrate, as well as the most suitable habitats for small felids, I provided a means for guiding prioritisation of conservation areas. Considering the current policy discussion on protection of habitats and species (e.g., the 30 by 30 target), my thesis could represent an invaluable tool to guide conservation actions towards more efficient and effective strategies aimed at maximising the protected biodiversity. The differences between habitat suitability models for small felids dependent on the algorithm applied that I highlighted, however, suggest conservationists to apply and test different modelling frameworks, finding the one that is more appropriate to the specific context.

## Declaration of authorship

I declare that this thesis is primarily my work. The conception was collaborative and the individual authors are specified at the beginning of each chapter. All the analyses and writing are my own. Since each chapter was collaborative, I use the pronoun *we* to indicate the same.

Luca Chiaverini

## Funding

This DPhil was generously funded by scholarships from Wildlife Conservation Research Unit (WildCRU) and from the Horne Foundation. Additional financial support was provided by the Robertson Foundation.

**Table of Contents**

Abstract ..... i
Declaration of authorship ..... iv
Funding ..... iv
Chapter 1
General introduction ..... 1
Overview of the thesis structure ..... 9
References ..... 12
Chapter 2
Predicting biodiversity richness in rapidly changing landscapes: climate, low human pressure or protection as salvation? ..... 20
Abstract ..... 21
Introduction ..... 22
Methods ..... 25
Results ..... 32
Discussion ..... 45
Conclusions ..... 54
References ..... 55
Supplementary material for Chapter 2 ..... 65
Chapter 3
Multi-scale, multivariate community models improve designation of biodiversity hotspots in the Sunda Islands. ..... 122
Abstract ..... 123
Introduction ..... 125
Materials and methods ..... 127
Results ..... 148
Discussion ..... 160
References ..... 173
Supplementary material for Chapter 3 ..... 183
Chapter 4
Not seeing the forest for the trees: generalised linear model out-performs random forest in species distribution modelling for Southeast Asian felids. ..... 217
Abstract ..... 218
Introduction ..... 220
Materials and methods ..... 223

Results ..... 229
Discussion ..... 238
Conclusions ..... 243
References ..... 244
Supplementary material for Chapter 4 ..... 252
Chapter 5
Identifying gaps in the conservation of small wild cats of Southeast Asia. ..... 365
Abstract ..... 366
Introduction ..... 368
Materials and methods ..... 370
Results ..... 379
Discussion ..... 401
References ..... 407
Supplementary material for Chapter 5 ..... 416
Chapter 6
General discussion ..... 464
Implications for conservation in Southeast Asia ..... 479
Scope and limitation ..... 484
Conclusions ..... 486
References ..... 486

# Chapter 1

# General introduction

A major objective for conservation biology is to highlight the areas of highest biodiversity in order to prioritise and direct conservation actions. To this purpose, Myers et al. (2000) identified global regions, defined as biodiversity hotspots, characterised by exceptional concentration of endemic species and, at the same time, facing worrying loss of habitat. Specifically, to be considered a biodiversity hotspot, a biogeographic region must have contained at least 0.5% or 1,500 of the world's vascular plant species as endemic and must have lost at least 70% of its primary vegetation (Myers et al. 2000). The first quantitative assessment of the biogeographic regions respecting these criteria led to the identification of 25 biodiversity hotspots globally that, overall, hosted 44% of the global endemic vascular plants and 35% of the endemic terrestrial vertebrates, in an areas comprising only 1.4% of the Earth's surface, and 11.8% of the land surface (Myers et al. 2000). Further updates and revisions of this first assessment, following availability of additional data for other global regions, coupled with better-defined criteria, led to an increase in the number of biodiversity hotspots to 35 (Mittermeier et al. 2004, Mittermeier et al. 2011), increasing their extent to 15.7% of the land surface, and sustaining about 50% of the world's plant biodiversity and 77% of global mammals, birds, amphibians and reptiles (Brooks et al. 2002, Mittermeier et al. 2004). More recently, an additional hotspot has been identified and designed (Noss et al. 2015), increasing the number of the global biodiversity hotspots to 36.

The appeal of identifying the global regions that harbour most of the species richness is straightforward: although, from a purely ideological point of view, it would be desirable to preserve the biodiversity in its entirety, this target is unrealistic and, to maximise practical and economic efforts, conservation actions have to be carefully focused on those regions where the major outcomes can be achieved, and tailored to have the greatest effectiveness (Myers 2003, Tulloch et al. 2015, Brum et al. 2017). However, the concept of biodiversity

hotspots as proposed by Myers et al. (2000) has often raised a number of criticisms for several reasons. Jepson and Canney (2001), for example, warned that a simplistic application of the concept of biodiversity hotspot, as the protection of a few areas where most of the endemic species occur, might give a false sense of security, and it might be seen, especially from decision-makers, as a panacea. Similarly, Kareiva and Marvier (2003) argued that the over-simplistic appeal of the biodiversity hotspots, as defined by Myers et al. (2000), would take the attention of investors and conservationists away from several other regions, that they defined as coldspots, that, although less rich in biodiversity, represent large natural areas fundamental for the ecosystem services that they provide. Likewise, Smith et al. (2001) argued that the process of identification of the biodiversity hotspots neglects the regions of ecological transitions, and advocated for conservation actions to be focused not only towards the areas that harbour the highest number of species, but also towards areas that are more comprehensively important for the generation and for the maintenance of biodiversity. Additionally, biodiversity hotspots have been identified based on criteria concerning solely habitat loss and occurrence of endemic vascular plants and there is no proof that they can be considered also a proxy for invertebrate biodiversity, which accounts for the vast majority of the Earth's total species (Stork and Habel 2014, Titley et al. 2017). Finally, biodiversity hotspots evaluated only terrestrial environments, neglecting marine habitats (Roberts et al. 2002, Fan et al. 2022).

It can also be argued that the biodiversity hotspots, as defined by Myers et al. (2000), are characterised by a spatial scale too coarse to allow conservation actions to be focused and effective, and several authors have advocated for the need of identifying smaller areas within the biodiversity hotspots, where the majority of the species occur (Murray-Smith et al. 2009, Canadas et al. 2014, Noroozi et al. 2018). The strategy of identifying these smaller areas, defined as "hotspots within hotspots" (Canadas et al. 2014) or "micro-hotspots" (Fenu et al.

2010, Grant and Samways 2011, Zampíglia et al. 2019), would allow both a better availability of data on species distribution for smaller regions and a more effective utilisation of the practical and economic resources for conservation.

Among the global regions where the highest number of biodiversity hotspots have been identified is Southeast Asia, which encompasses four of them, namely the Indo-Burma, the Mountains of Southwest China and the Himalaya in the mainland, and the Sundaland comprising the island of Borneo, Sumatra and Java. Altogether, these hotspots extend over an areas of almost 5,000,000 $Km^2$, encompassing most of the Southeast Asian subcontinent. This area is home to a uniquely rich biodiversity, such that Myers et al. (2000) ranked the Sundaland and the Indo-Burma respectively as the 3$^{rd}$ and 6$^{th}$ "hottest hotspot", including more than 25,000 endemic plant species and almost 1,500 endemic terrestrial vertebrates, respectively representing 8.5% and 5.2% of the global vascular plants and vertebrate species (Myers et al. 2000, Mittermeier et al. 2004, Mittermeier et al. 2011). Nonetheless, it is clear that ensuring an equal conservation effort over such an immensely vast area is an impossible task.

Southeast Asia, however, in addition of being one of the most biodiverse region globally, is also one of the most critically endangered by a suite of anthropogenic threats that pose a serious risk of extinction to regional biodiversity (Sodhi et al. 2004, Wilcove et al. 2013, Hughes 2017, Ng et al. 2020). Southeast Asia has experienced exceptionally high rates of human population growth over the last 60 years, at well 2% per annum from the 1950s to the beginning to the 1990s, and still above 1% per annum. More than 500 million people currently live in the region, at densities from twice to six time the world's mean, depending on the country. Even though fertility rates have now fallen below the replacement levels in most of the region, the population is projected to continue growing for the next 30 years at least (United Nations 2019).

Additionally, human population growth in Southeast Asia has coincided with a positive economic development in the region since the 1980s. Southeast Asia's average growth of Gross Domestic Product (GDP) has been 5.5% over the period 2000-2018 (Nawaz et al. 2019), and several countries have moved from an agriculture-based economy to a different, more advanced economic transformation stage (Hartley and Medlock 2017). Southeast Asian countries, however, strongly rely on their natural resources such as soil, forests and water for the livelihood and economic opportunities of their inhabitants, and increased population and economic growth, along with social inequalities, have resulted in an increased pressure over environments, with consequent depletion of natural resources, land degradation and water and air pollution (Khan et al. 2019). Therefore, the levels of population and economic growth that Southeast Asia has experienced, pose a serious threat to regional biodiversity.

Human population correlates with species endangerment through several mechanisms, ranging from direct threats to species to more hidden and understudied ones. The growth of human population is directly linked with an increase of the resources needed which, in turn, is maintained by a suite of human enterprises such as agriculture, industry, fishing and logging. These enterprises transform the landscapes, alter bio-geochemical cycles and add or remove species from the ecosystems. These changes drive further alterations to the functions of the ecological systems, by driving global climate change and causing direct loss of biodiversity (Vitousek et al. 1997, Pettorelli et al. 2021, Storch et al. 2022). Among the most impelling causes of loss of biodiversity in Southeast Asia, however, deforestation and habitat loss and degradation represent the primary and most direct driving forces (Sala et al. 2000, Newbold et al. 2015, Hughes 2017). Indonesia, for example, shows the largest recent increase in the rates of forest loss globally, and Malaysia shows the highest level of forest loss in relation to land area (Hansen et al. 2013). In 2011, Conservation International ranked the

Indo-Burma and the Mountains of Southwest China biodiversity hotspots as the 1st and 6th most endangered hotspots globally, given that they retain only 5% and 8% of their original natural habitats, respectively. Additionally, very high rates of forest loss in the region are expected to continue during the next decade at least (Cushman et al. 2017), and it has been modelled that, if current rates of deforestation persist, Southeast Asia would lose three-quarter of its original forest cover by 2100 (Achard et al. 2002). Habitat loss and degradation would, in turn, reduce the quality and the accessibility of suitable habitat for a wide range of species (Kaszta et al. 2019), with dramatic consequences for biodiversity, which would lose between 24-63% of its terrestrial endemic species (Sodhi and Brook 2006).

In light of the critical threats that Southeast Asian biodiversity is facing, it is evident that timely and robust information about the spatial distribution of biodiversity across the region is urgently needed to direct conservation actions on the most important "hotspots within hotspots" sensu Canadas et al. (2014). Additionally, prioritising areas for conservation in Southeast Asia is of utmost importance also in light of the current policy discussions, including the 30/30 target. The 30/30 target was agreed at the COP15 meeting of the Convention on Biological Diversity to [*Ensure and enable that by 2030 at least 30 per cent of terrestrial, inland water, and of coastal and marine areas, especially areas of particular importance for biodiversity and ecosystem functions and services, are effectively conserved and managed through ecologically representative, well-connected and equitably governed systems of protected areas and other effective area-based conservation measures…*]. Therefore, to achieve this target, countries and regions are expected to create networks of protected areas that represent a wide and comprehensive range of ecosystem, habitats and species, in order to maintain key ecological processes and services, with benefits to both nature and humanity.

Although extinction risks are ubiquitous over taxa, carnivores are particularly vulnerable to anthropic pressure as their ecological role, atop the food pyramid, often dictates low population density and large home ranges, leading to conflict with humans (Cardillo et al. 2004). Additionally, habitat loss and fragmentation further reduce the availability of large and intact patches of suitable habitat for carnivores (Cushman et al. 2018, Kaszta et al. 2020), which are inevitably pushed towards smaller, less suitable areas, often outside protected areas and embedded within anthropic matrices (Linnell et al. 2001, Crooks et al. 2017, Loveridge et al. 2022).

Southeast Asia is the richest region globally for number of sympatric felids, as well as the region with the highest number of endemic felids (Macdonald and Loveridge 2010, Luo et al. 2014). Specifically, considering the recent reclassification of the leopard cat in two distinct species: mainland leopard cat *Prionailurus bengalensis* and Sunda leopard cat *Prionailurus javanensis* (Kitchener et al. 2017), 13 felid species occur in Southeast Asia, of which 11 are endemic to the region. Of the 13 felids species, only tiger *Panthera tigris* and leopard *Panthera pardus* are classified as big carnivores, while mainland clouded leopard *Neofelis nebulosa* and Sunda clouded leopard *Neofelis diardi* are at the mass boundary identified by Carbone et al. (1999) to define the large carnivores energetically. The remaining felid species are all classified as small felids.

In Southeast Asia, large and charismatic species, including felids, have been the focus for most conservation actions, and have attracted the majority of the economic investments for conservation (Brodie 2009, Macdonald et al. 2015, Colléony et al. 2017). Unsurprisingly, Macdonald et al. (2015) and Macdonald et al. (2017) highlighted that large species with a high threat level were more likely to catalyse the attention of the general population, making large felids the species group with the greatest ambassadorial potential. Consequently, despite the prevalence of small felids in Southeast Asia, these species have received very limited

attention from conservationists, and information about their ecology and conservation status is often lacking or limited to local contexts (Marneweck et al. 2021).

In addition to their intrinsic conservation value, small felids have important ecological roles: as secondary consumers they can exercise top-down control on herbivorous mammals that can, in turn, affect apex predators (Roemer et al. 2009), and it has also been observed that they can take on the role of apex predator through ecological release when larger carnivores are extirpated (de Oliveira et al. 2010). Additionally, the morphological similarities that small felids share, coupled with their correspondingly similar behaviours and ecological needs, inevitably predispose these species for interspecific competition. Therefore Southeast Asia, with its high number of sympatric felids, represents a promising testing ground to explore niche separation and guild structure among small felids.

Thus, the main goals of this thesis are: (1) to highlight the areas harbouring the highest number of terrestrial species, as well as the habitat factors driving the species richness, within the mainland Southeast Asian biodiversity hotspots, as well as (2) within the offshore Southeast Asian biodiversity hotspot. Then, focusing on the small felids occurring in the region, (3) I first carried out an evaluation of the best methodological approaches to model habitat suitability individually for sampled small felids, and then (4) I highlighted the most suitable areas, as well as the most effective protected areas, for Southeast Asian small felids.

## Overview of the thesis structure

Chapters 1 and 6 provide, respectively, a general introduction to contextualise the background of my thesis and an overarching discussion in which I summarised the main findings, as well as the scope and limitations, and the final conclusions. Chapters 2 – 5 represent the data chapters of this thesis, of which I provide a brief overview of the main objectives below.

### Chapter 2

In this chapter, I addressed four interrelated topics associated with geographic distribution of terrestrial vertebrate biodiversity in mainland Southeast Asia. First, by implementing multivariate modelling techniques such as Canonical Correspondence Analysis and variance partitioning, I investigated the main habitat factors driving assemblages of terrestrial vertebrate biodiversity. Second, by using these habitat factors, I produced habitat suitability models for the sampled species by implementing Generalised Linear Models and Poisson distribution, and then aggregated the predictive surfaces of the single species to map and predict the geographic distribution of overall terrestrial species richness in mainland Southeast Asia. By overlaying layers of the protected areas, I also evaluated the effectiveness of protected areas in preserving biodiversity, and highlighted the most suitable areas currently lacking protection. Third, by training linear models using the modelled species richness as predictor variable and the modelled single-species habitat suitability models as response variables, I assessed the species whose presence was most strongly associated with overall species richness and could therefore have an important role as indicator species. Fourth, I evaluated an independent set of response variables to determine to what degree climate,

human pressure or protection explained the modelled pattern of terrestrial vertebrate biodiversity.

**Chapter 3**

Similar to the previous chapter, in the third chapter my first goal was to highlight the multi-scale, multivariate habitat factors driving assemblages of terrestrial vertebrate biodiversity in the Sunda Islands of Borneo and Sumatra. I achieved this first goal by implementing multivariate modelling techniques in order to simultaneously assess the main drivers of sampled species, while still being able to identify each species' individual contribution. I applied multivariate techniques such as Canonical Correspondence Analysis and variance partitioning. Second, I used the habitat factors highlighted in the first step to produce habitat suitability models for each sampled species, by training Generalised Linear Models with quasi-Poisson distribution. Then, I aggregated the single-species habitat suitability models to map the areas harbouring the highest values of species richness in Borneo and Sumatra, independently and, by overlapping layers of the protected areas in the Sunda Islands, I also assessed the effectiveness of the protected areas systems in preserving biodiversity, and highlighted the unprotected, most biodiverse areas. Then, by sampling the number of species from the aggregated species richness predictive layers, and the occurrence of species from the single-species habitat suitability models I evaluated, independently for Borneo and Sumatra, the role of sampled species in being a proxy of the overall vertebrate terrestrial biodiversity by performing linear models. Finally, I defined which factor was more strongly associated with the modelled levels of species richness among climatic, anthropic and protection factors.

**Chapter 4**

In the fourth chapter, I explored the varying results of different modelling frameworks to produce habitat suitability models, using small felids occurring in Southeast Asia as a case study. Specifically, I modelled habitat suitability independently for small felids sampled in mainland Southeast Asia, Borneo and Sumatra by adopting two algorithms (i.e., Generalised Linear Model and Random Forest) and two data resampling techniques (i.e., bootstrapping and not bootstrapping the training data), for a total of four modelling frameworks. Then, I first investigated the effects of the different algorithms and of the resampling techniques on the predictive performances (i.e., Area Under the ROC Curve, Percent Correctly Classified, sensitivity, specificity and True Skill Statistics) of the models by applying ANOVA tests. Second, I assessed the effects of algorithms and resampling techniques on the Pearson's correlation indexes calculated between each of the four competing models for each species, by implementing Mantel tests, in order to evaluate the effect of algorithm and resampling technique on the correlation between habitat suitability models. Finally, I performed simple and partial Mantel tests between the difference in the ordinal scales among pairs of covariates across models and the difference in each of the factors I hypothesised to affected spatial scale selection (i.e., body size, species, genus, study area and algorithm), in order to evaluate whether the most representative spatial scales of the habitat factors for small felids were associated with bio-ecological traits of the species, or whether they were due to the different algorithms adopted to assess them.

**Chapter 5**

In this chapter, building on the results previously obtained in Chapter 4 on the statistically most supported model for each small felid species sampled in Southeast Asia, I

aggregated the individual habitat suitability models with the primary objective to define the geographic distribution of small felids in mainland Southeast Asia, Borneo and Sumatra. Then, by reclassifying the aggregated predictive surfaces in low, medium and high suitability on the basis of their quantile distribution, I evaluated the effectiveness of the current systems of protected areas in preserving suitable habitats for small felid species by calculating the ratio between the extent of each suitability class occurring within protected areas, and the extent occurring outside the protected areas networks. My third goal was then to highlight the highly suitable habitat for small felids still lacking protection. Finally, I evaluated whether small felids can be considered a good proxy of terrestrial vertebrate biodiversity and, among the sampled small felids, which species was the best indicator of biodiversity, by calculating Pearson's correlation indexes between the terrestrial biodiversity distribution layers modelled in Chapters 2 and 3, and the single-species habitat suitability models produced for sampled felids, as well as the aggregated multispecies predictive surfaces for all sampled felids.

**References**

Achard, F., H. D. Eva, H. J. Stibig, P. Mayaux, J. Gallego, T. Richards, and J. P. Malingreau. 2002. Determination of deforestation rates of the world's humid tropical forests. Science **297**:999-1002.

Brodie, J. F. 2009. Is research effort allocated efficiently for conservation? Felidae as a global case study. Biodiversity and Conservation **18**:2927-2939.

Brooks, T. M., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, A. B. Rylands, W. R. Konstant, P. Flick, J. Pilgrim, S. Oldfield, G. Magin, and C. Hilton-Taylor. 2002. Habitat loss and extinction in the hotspots of biodiversity. Conservation Biology **16**:909-923.

Brum, F. T., C. H. Graham, G. C. Costa, S. B. Hedges, C. Penone, V. C. Radeloff, C. Rondinini, R. Loyola, and A. D. Davidson. 2017. Global priorities for conservation across multiple dimensions of mammalian diversity. Proceedings of the National Academy of Sciences of the United States of America **114**:7641-7646.

Canadas, E. M., G. Fenu, J. Penas, J. Lorite, E. Mattana, and G. Bacchetta. 2014. Hotspots within hotspots: Endemic plant richness, environmental drivers, and implications for conservation. Biological Conservation **170**:282-291.

Carbone, C., G. M. Mace, S. C. Roberts, and D. W. Macdonald. 1999. Energetic constraints on the diet of terrestrial carnivores. Nature **402**:286-288.

Cardillo, M., A. Purvis, W. Sechrest, J. L. Gittleman, J. Bielby, and G. M. Mace. 2004. Human population density and extinction risk in the world's carnivores. Plos Biology **2**:909-914.

Colléony, A., S. Clayton, D. Couvet, M. Saint Jalme, and A. C. Prévot. 2017. Human preferences for species conservation: Animal charisma trumps endangered status. Biological Conservation **206**:263-269.

Crooks, K. R., C. L. Burdett, D. M. Theobald, S. R. B. King, M. Di Marco, C. Rondinini, and L. Boitani. 2017. Quantification of habitat fragmentation reveals extinction risk in terrestrial mammals. Proceedings of the National Academy of Sciences of the United States of America **114**:7635-7640.

Cushman, S. A., N. B. Elliott, D. Bauer, K. Kesch, L. Bahaa-el-din, H. Bothwell, M. Flyman, G. Mtare, D. W. Macdonald, and A. J. Loveridge. 2018. Prioritizing core areas, corridors and conflict hotspots for lion conservation in southern Africa. Plos One **13**.

Cushman, S. A., E. Macdonald, E. Landguth, Y. Malhi, and D. Macdonald. 2017. Multiple-scale prediction of forest loss risk across Borneo. Landscape Ecology **32**:1581-1598.

de Oliveira, T. G., M. A. Tortato, L. Silveira, C. B. Kasper, F. D. Mazim, M. Lucherini, A. T. Jácomo, J. B. G. Soares, R. V. Marques, and M. Sunquist. 2010. Ocelot ecology and its effect on the small-felid guild in the lowland Neotropics. Pages 559-580 *in* D. W. Macdonald and A. J. Loveridge, editors. The Biology and Conservation of Wild Felids. Oxford University Press, Oxford, UK.

Fan, H. Z., M. P. Huang, Y. H. Chen, W. L. Zhou, Y. B. Hu, and F. W. Wei. 2022. Conservation priorities for global marine biodiversity across multiple dimensions. National Science Review.

Fenu, G., E. Mattana, A. Congiu, and G. Bacchetta. 2010. The endemic vascular flora of Supramontes (Sardinia), a priority plant conservation area. Candollea **65**:347-358.

Grant, P. B. C., and M. J. Samways. 2011. Micro-hotspot determination and buffer zone value for Odonata in a globally significant biosphere reserve. Biological Conservation **144**:772-781.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science **342**:850-853.

Hartley, P. R., and K. B. Medlock. 2017. The Valley of Death for New Energy Technologies. Energy Journal **38**:33-61.

Hughes, A. C. 2017. Understanding the drivers of Southeast Asian biodiversity loss. Ecosphere **8**.

Jepson, P., and S. Canney. 2001. Biodiversity hotspots: hot for what? Global Ecology and Biogeography **10**:225-227.

Kareiva, P., and M. Marvier. 2003. Conserving biodiversity coldspots - Recent calls to direct conservation funding to the world's biodiversity hotspots may be bad investment advise. American Scientist **91**:344-351.

Kaszta, Z., S. A. Cushman, A. J. Hearn, D. Burnham, E. A. Macdonald, B. Goossens, S. K. S. S. Nathan, and D. W. Macdonald. 2019. Integrating Sunda clouded leopard (Neofelis diardi) conservation into development and restoration planning in Sabah (Borneo). Biological Conservation **235**:63-76.

Kaszta, Z., S. A. Cushman, and D. W. Macdonald. 2020. Prioritizing habitat core areas and corridors for a large carnivore across its range. Animal Conservation.

Khan, K. A., K. Zaman, A. M. Shoukry, A. Sharkawy, S. Gani, Sasmoko, J. Ahmad, A. Khan, and S. S. Hishan. 2019. Natural disasters and economic losses: controlling external migration, energy and environmental resources, water demand, and financial development for global prosperity. Environmental Science and Pollution Research **26**:14287-14299.

Kitchener, A. C., C. Breitenmoser-Würsten, E. Eizirik, A. Gentry, L. Werdelin, A. Wilting, N. Yamaguchi, A. V. Abramov, P. Christiansen, C. Driscoll, J. W. Duckworth, W. Johnson, S. J. Luo, E. Meijaard, P. O'Donoghue, J. Sanderson, K. Seymour, M. Bruford, C. Groves, M. Hoffmann, K. Nowell, Z. Timmons, and S. Tobe. 2017. A revised taxonomy of the Felidae. The final report of the Cat Classification Task Force of the IUCN/SSC Cat Specialist Group.

Linnell, J. D. C., J. E. Swenson, and R. Andersen. 2001. Predators and people: conservation of large carnivores is possible at high human densities if management policy is favourable. Animal Conservation **4**:345-349.

Loveridge, A. J., L. L. Sousa, S. Cushman, Z. Kaszta, and D. W. Macdonald. 2022. Where have all the lions gone? Establishing realistic baselines to assess decline and recovery of African lions. Diversity and Distributions **28**:2388-2402.

Luo, S. J., Y. Zhang, W. E. Johnson, L. Miao, P. Martelli, A. Antunes, J. L. D. Smith, and S. J. O'Brien. 2014. Sympatric Asian felid phylogeography reveals a major Indochinese-Sundaic divergence. Molecular Ecology **23**:2072-2092.

Macdonald, D. W., and A. J. Loveridge. 2010. Biology and conservation of wild felids. Oxford University Press, Oxford ; New York.

Macdonald, E. A., D. Burnham, A. E. Hinks, A. J. Dickman, Y. Malhi, and D. W. Macdonald. 2015. Conservation inequality and the charismatic cat: Felis felicis. Global Ecology and Conservation **3**:851-866.

Macdonald, E. A., A. Hinks, D. J. Weiss, A. Dickman, D. Burnham, C. J. Sandom, Y. Malhi, and D. W. Macdonald. 2017. Identifying ambassador species for conservation marketing. Global Ecology and Conservation **12**:204-214.

Marneweck, C., A. R. Butler, L. C. Gigliotti, S. N. Harris, A. J. Jensen, M. Muthersbaugh, B. A. Newman, E. A. Saldo, K. Shute, K. L. Titus, S. W. Yu, and D. S. Jachowski. 2021. Shining the spotlight on small mammalian carnivores: Global status and threats. Biological Conservation **255**.

Mittermeier, R. A., P. R. Gil, M. Hoffman, J. Pilgrim, T. Brooks, C. G. Mittermeier, J. Lamoreux, and G. A. B. da Fonseconda. 2004. Hotspots revisited: Earth's biologically richest and most endangered terrestrial ecoregions. Cemex, Mexico City.

Mittermeier, R. A., W. R. Turner, F. W. Larsen, T. M. Brooks, and C. Gascon. 2011. Global Biodiversity Conservation: The Critical Role of Hotspots. Pages 3-22 *in* F. E. Zachos and J. C. Habel, editors. Biodiversity Hotspots. Distribution and Protection of Conservation Priority Areas. Springer, Berlin, Heidelberg.

Murray-Smith, C., N. A. Brummitt, A. T. Oliveira-Filho, S. Bachman, J. Moat, E. M. N. Lughadha, and E. J. Lucas. 2009. Plant Diversity Hotspots in the Atlantic Coastal Forests of Brazil. Conservation Biology **23**:151-163.

Myers, N. 2003. Biodiversity hotspots revisited. Bioscience **53**:916-917.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. Nature **403**:853-858.

Nawaz, M. A., A. Azam, and M. A. Bhatti. 2019. Natural resources depletion and economic growth: evidence from ASEAN countries. Pakistan Journal of Economic Studies **2**:155-172.

Newbold, T., L. N. Hudson, S. L. L. Hill, S. Contu, I. Lysenko, R. A. Senior, L. Borger, D. J. Bennett, A. Choimes, B. Collen, J. Day, A. De Palma, S. Diaz, S. Echeverria-Londono, M. J. Edgar, A. Feldman, M. Garon, M. L. K. Harrison, T. Alhusseini, D. J. Ingram, Y. Itescu, J. Kattge, V. Kemp, L. Kirkpatrick, M. Kleyer, D. L. P. Correia, C. D. Martin, S. Meiri, M. Novosolov, Y. Pan, H. R. P. Phillips, D. W. Purves, A. Robinson, J. Simpson, S. L. Tuck, E. Weiher, H. J. White, R. M. Ewers, G. M. Mace, J. P. W. Scharlemann, and A. Purvis. 2015. Global effects of land use on local terrestrial biodiversity. Nature **520**:45-+.

Ng, L., A. Campos-Arceiz, S. Sloan, A. C. Hughes, D. C. F. Tiang, B. V. Li, and A. M. Lechner. 2020. The scale of biodiversity impacts of the Belt and Road Initiative in Southeast Asia. Biological Conservation **248**.

Noroozi, J., A. Talebi, M. Doostmohammadi, S. B. Rumpf, H. P. Linder, and G. M. Schneeweiss. 2018. Hotspots within a global biodiversity hotspot - areas of endemism are associated with high mountain ranges. Scientific Reports **8**.

Noss, R. F., W. J. Platt, B. A. Sorrie, A. S. Weakley, D. B. Means, J. Costanza, and R. K. Peet. 2015. How global biodiversity hotspots may go unrecognized: lessons from the North American Coastal Plain. Diversity and Distributions **21**:236-244.

Pettorelli, N., N. A. J. Graham, N. Seddon, M. M. D. Bustamante, M. J. Lowton, W. J. Sutherland, H. J. Koldewey, H. C. Prentice, and J. Barlow. 2021. Time to integrate global climate change and biodiversity science-policy agendas. Journal of Applied Ecology **58**:2384-2393.

Roberts, C. M., C. J. McClean, J. E. N. Veron, J. P. Hawkins, G. R. Allen, D. E. McAllister, C. G. Mittermeier, F. W. Schueler, M. Spalding, F. Wells, C. Vynne, and T. B. Werner. 2002. Marine biodiversity hotspots and conservation priorities for tropical reefs. Science **295**:1280-1284.

Roemer, G. W., M. E. Gompper, and B. Van Valkenburgh. 2009. The Ecological Role of the Mammalian Mesocarnivore. Bioscience **59**:165-173.

Sala, O. E., F. S. Chapin, J. J. Armesto, E. Berlow, J. Bloomfield, R. Dirzo, E. Huber-Sanwald, L. F. Huenneke, R. B. Jackson, A. Kinzig, R. Leemans, D. M. Lodge, H. A. Mooney, M. Oesterheld, N. L. Poff, M. T. Sykes, B. H. Walker, M. Walker, and D. H. Wall. 2000. Biodiversity - Global biodiversity scenarios for the year 2100. Science **287**:1770-1774.

Smith, T. B., S. Kark, C. J. Schneider, R. K. Wayne, and C. Moritz. 2001. Biodiversity hotspots and beyond: the need for preserving environmental transitions. Trends in Ecology & Evolution **16**:431-431.

Sodhi, N. S., and B. W. Brook. 2006. Southeast Asian biodiversity in crisis. Cambridge University Press, Cambridge.

Sodhi, N. S., L. P. Koh, B. W. Brook, and P. K. L. Ng. 2004. Southeast Asian biodiversity: an impending disaster. Trends in Ecology & Evolution **19**:654-660.

Storch, D., I. Símová, J. Smycka, E. Bohdalková, A. Toszogyova, and J. G. Okie. 2022. Biodiversity dynamics in the Anthropocene: how human activities change equilibria of species richness. Ecography **2022**.

Stork, N. E., and J. C. Habel. 2014. Can biodiversity hotspots protect more than tropical forest plants and vertebrates? Journal of Biogeography **41**:421-428.

Titley, M. A., J. L. Snaddon, and E. C. Turner. 2017. Scientific research on animal biodiversity is systematically biased towards vertebrates and temperate regions. Plos One **12**.

Tulloch, V. J. D., A. I. T. Tulloch, P. Visconti, B. S. Halpern, J. E. M. Watson, M. C. Evans, N. A. Auerbach, M. Barnes, M. Beger, I. Chadès, S. Giakoumi, E. McDonald-Madden, N. J. Murray, J. Ringma, and H. P. Possingham. 2015. Why do we map threats? Linking threat mapping with actions to make better conservation decisions. Frontiers in Ecology and the Environment **13**:91-99.

United Nations, D. o. E. a. S. A., Population Division. 2019. World Populations Prospects 2019: Highlights.

Vitousek, P. M., H. A. Mooney, J. Lubchenco, and J. M. Melillo. 1997. Human domination of Earth's ecosystems. Science **277**:494-499.

Wilcove, D. S., X. Giam, D. P. Edwards, B. Fisher, and L. P. Koh. 2013. Navjot's nightmare revisited: logging, agriculture, and biodiversity in Southeast Asia. Trends in Ecology & Evolution **28**:531-540.

Zampíglia, M., R. Bisconti, L. Maiorano, G. Aloise, A. Siclari, F. Pellegrino, G. Martino, A. Pezzarossa, A. Chiocchio, C. Martino, G. Nascetti, and D. Canestrelli. 2019. Drilling Down Hotspots of Intraspecific Diversity to Bring Them Into On-Ground Conservation of Threatened Species. Frontiers in Ecology and Evolution **7**.

# Chapter 2

# Predicting biodiversity richness in rapidly changing landscapes: climate, low human pressure or protection as salvation?

David W. Macdonald*, Luca Chiaverini*, Helen M. Bothwell, Żaneta Kaszta, Eric Ash, Gilmoore Bolongon, Özgün Emre Can, Ahimsa Campos-Arceiz, Phan Channa, Gopalasamy Reuben Clements, Andrew J. Hearn, Laurie Hedges, Saw Htun, Jan F. Kamler, Ewan A. Macdonald, Jonathan Moore, Hla Naing, Manabu Onuma, Akchousanh Rasphone, D. Mark Rayan, Joanna Ross, Priya Singh, Cedric Kai Wei Tan, Jamie Wadey, Bhupendra P. Yadav, Samuel A. Cushman

* Joint first authors

Published in Biodiversity and Conservation. (2020) 29: 4035-4057

https://doi.org/10.1007/s10531-020-02062-x

**Abstract**

Rates of biodiversity loss in Southeast Asia are among the highest in the world, and the Indo-Burma and South-Central China Biodiversity Hotspots rank among the world's most threatened. Developing robust multi-species conservation models is critical for stemming biodiversity loss both here and globally. We used a large and geographically extensive remote-camera survey and multi-scale, multivariate optimization species distribution modelling to investigate the factors driving biodiversity across these two adjoining biodiversity hotspots. Four major findings emerged from the work. (i) We identified clear spatial patterns of species richness, with two main biodiverse centres in the Thai-Malay Peninsula and in the mountainous region of Southwest China. (ii) Carnivores in particular, and large ungulates to a lesser degree, were the strongest indicators of species richness. (iii) Climate had the largest effect on biodiversity, followed by protected status and human footprint. (iv) Gap analysis between the biodiversity model and the current system of protected areas revealed that the majority of areas supporting the highest predicted biodiversity are not protected. Our results highlighted several key locations that should be prioritized for expanding the protected area network to maximize conservation effectiveness. We demonstrated the importance of switching from single-species to multi-species approaches to highlight areas of high priority for biodiversity conservation. In addition, since these areas mostly occur over multiple countries, we also advocate for a paradigmatic focus on transboundary conservation planning.

## 1. Introduction

Originally, 25 biodiversity hotspots were identified globally, collectively supporting 44% and 35% of the world's vascular plants and terrestrial vertebrates, respectively, in an area equal to 1.4% of the Earth's land surface (Myers et al. 2000). Among these hotspots, two important ones are the Indo-Burma and South-Central China Biodiversity Hotspots (Figure 1). Together, they form a continuous area of almost 3,000,000km$^2$, encompassing much of Southeast Asia. This region is home to 10,500 endemic plants and 706 endemic vertebrates, representing 3.5% and 2.6% of global vascular plants and vertebrates, respectively (Myers et al. 2000). To these are added also the Sundaland Biodiversity Hotspot, only partially included in mainland Southeast Asia, and the Himalaya Biodiversity Hotspot, only later highlighted as a hotspot (Mittermeier et al. 2004).

In addition to being the most biodiverse hotspots, these four are also among the most threatened. The biological richness of Southeast Asia is endangered by a suite of anthropogenic threats, including land conversion and over-exploitation of natural resources, which are drastically modifying regional landscapes (Cushman et al. 2017, Hughes 2017b), driving some of the highest rates of habitat loss globally (Miettinen et al. 2011, Gaveau et al. 2016). Among the most impactful causes of land conversion in the region are deforestation for monocultures (Azhar et al. 2017), urbanisation (Zhao et al. 2018) and ancillary enterprises related to human activities, such as road development (Kaszta et al. 2020a). Habitat loss is reducing the extent, quality and accessibility of suitable habitats (Fahrig 1997, Hearn et al. 2018), and 24-63% of the regional terrestrial endemic species will likely become extinct by 2100 if current rates of habitat loss continue (Sodhi and Brook 2006).

Traditionally, conservation strategies have focused on single-species in local geographic areas, resulting in a limited ability to optimize multi-species conservation across ecosystems

and management boundaries (Ruter et al. 2014). The extent of biodiversity loss throughout Southeast Asia stresses the importance of shifting conservation focus towards multi-species, transboundary scope (Lim 2016).

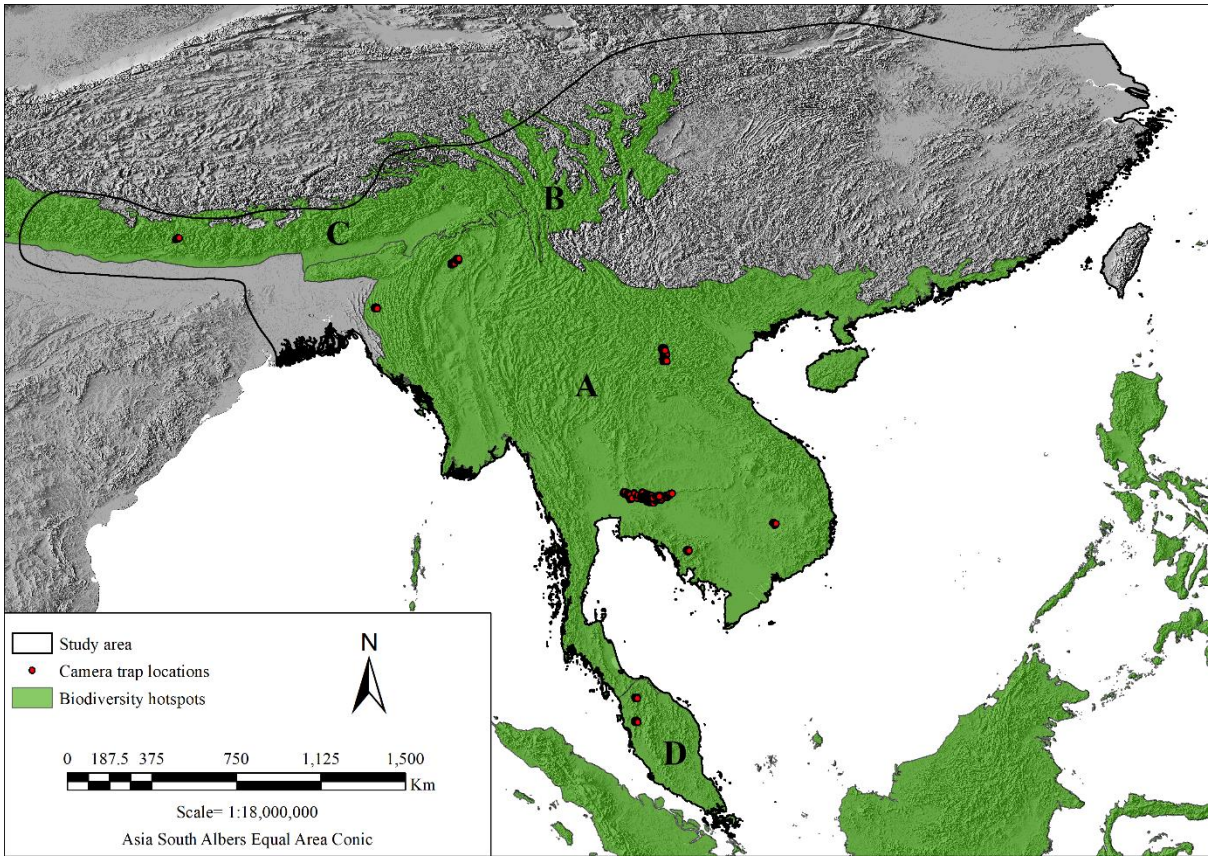



**Figure 1.** Map of the study area, showing the camera trap locations and the Biodiversity Hotspots encompassed: Indo-Burma (A) and South-Central China (B). In addition, the Himalaya (C) is shown, an area recognised subsequently as a hotspot (Mittermeier et al. 2004), and the Sundaland (D), only partially included in the study area. The polygons delineating the biodiversity hotspots have been downloaded from the Critical Ecosystem Partnership Fund (CEPF)'s website (Hoffman et al. 2016).

We present a framework for characterizing broad-scale, multi-species community assemblage drivers and predicting spatial patterns in biodiversity. Broad-scale biodiversity patterns are influenced by anthropogenic, environmental and management factors: biodiversity might be highest (a) where climatic factors create favourable environmental conditions for diverse species (Pearson and Dawson 2003), (b) where low levels of human pressure occur (Di Marco et al. 2013), (c) where higher levels of protection create safe havens (Bruner et al. 2001), or (d) where interactions between the aforementioned factors occur. Disentangling the drivers of biodiversity richness is essential for developing efficient regional, multi-species conservation planning. Moreover, describing species distribution is a foundation for prioritising and planning conservation strategies, providing a means to predict future changes in biodiversity and guiding optimization of conservation and development scenarios.

We addressed four interrelated goals. First, we investigated the factors driving species presence using multi-scale, multivariate modelling. Second, we predicted and mapped species richness across the region, and performed a gap analysis between the biodiversity model and the network of protected areas to highlight the hotspots lacking protection. Third, we determined the species whose presence was most strongly correlated to overall biodiversity and have an important role as indicators of species richness. Lastly, we employed statistical approaches to determine to what degree climate (hypothesis a), accessibility to human activities (hypothesis b), protected area status (hypothesis c), or their interactions (hypothesis d) explain modelled patterns of biodiversity. This framework provides a model for understanding patterns and drivers of biodiversity, and for prioritizing conservation areas for multiple species across transboundary extents using consistent and objective methods.

## 2. Methods

### *2.1.*Study area

The study area extends between 0º-30ºN latitude and 80ºE-120ºE longitude, and covers a broad altitudinal range, spanning coastal areas to Himalayan peaks above 8,000masl (Figure 1). The area encompasses all or part of 13 countries: Bangladesh, Bhutan, Cambodia, Laos, Myanmar, Nepal, Singapore, Taiwan, Thailand and Vietnam fall entirely within the study area, while China, India and Malaysia are partially encompassed.

Following the Köppen-Geiger climate classification (Beck et al. 2018), most of the study area falls within the tropical and humid-subtropical climatic zones, although the northernmost areas lie in the subtropical highland climatic zone, and the highest Himalayan regions in the tundra and subarctic climatic zones. The climatic diversity is reflected in the variety of ecosystems of the region, which includes nine of the fourteen biomes highlighted by Olson et al. (2001). In turn, the richness of ecosystems and habitats is associated with the aforementioned uniquely rich biodiversity that characterise the area.

Southeast Asia, however, has also been characterised, for at least the last 70 years, by intense human population growth and anthropogenic landscape change. The total human population of the countries encompassed in the region (excluding China and India) grew from 130 million people in 1950 to more than 500 million presently (United Nations 2019). Furthermore, the population is expected to grow for at least the next 30 years. Associated with this striking human population growth, Southeast Asia has experienced severe rates of forest loss, among the highest globally (Sodhi and Brook 2006). Overall, the countries encompassed in the study area (excluding China and India, but including Malaysia), have lost more than 25Mha of forest from 2001 to 2019, with the greatest losses during this period

being 28% for Malaysia (including the insular states) and 26% for Cambodia (Hansen et al. 2013, Global Forest Watch 2020).

### *2.2.*Data collection

Between 2008 and 2016, our field teams carried out systematic camera trap surveys across Southeast Asia, covering the full mainland range of clouded leopard (*Neofelis nebulosa*). Sampling occurred mainly in national parks and reserves, and spanned a broad altitudinal range, from 45masl in Pang Sida National Park, in Thailand, to 3,901masl in Langtang National Park, in Nepal. Camera traps were set 1.0-2.0km apart, with two cameras per station at ~40cm above the ground, and were deployed along forest trails, natural ridgelines and disused logging roads to maximize detection success of large felids (Macdonald et al. 2018). Nevertheless, the sampling protocol also captured a rich dataset of regional biodiversity. All terrestrial mammals, birds and reptiles whose species was unambiguously identified were included in the analysis. When a species could not be clearly identified, we retained the data at a broader taxonomic level (i.e., order, family or genus). In addition, we also included data related to human activities captured by camera traps, such as people and domestic animals. These have been used solely for assessing the habitat factors driving species assemblages, as the influence of anthropic disturbances on habitat selection by terrestrial species is clearly relevant, but were not incorporated in models of biodiversity distribution. For each species, we used the number of detections per camera trap station, applying a filtering method to ensure the independence of data and to reduce overestimation bias: we discounted records of the same species at the same camera trap station within 1 hour, except when animals were individually recognizable and when genders and/or age classes were unambiguous.

### *2.3.*Landscape covariates

We selected a preliminary set of 28 covariates covering a broad range of habitat gradients to investigate habitat requirements for the sampled species (Hughes 2017a). We included twelve landscape, four anthropic, three topographic, one climatic and eight spatial covariates. Species-habitat relationships are spatially structured at a number of scales and patterns, which interact with the perception and the behaviour of organisms, driving higher levels processes of population dynamics and community structure (Johnson 1980). Therefore, since environmental spatial patterns influence ecological processes and species-habitat relationships, to investigate more biologically meaningful derivatives of the original covariates, they were transformed into 46 covariates by applying composition (i.e., class proportion on the landscape) and configuration (i.e., landscape continuity) metrics using FRAGSTATS (McGarigal et al. 2012) (Table S1).

We obtained the original raster layers of the preliminary covariates from different sources and therefore they had different spatial resolutions and projections. To harmonise the raster layers, we followed the framework recently used in our work (Macdonald et al. 2019), where we used a similar set of covariates. We first re-projected all the layers to Asia South Albers Equal Area Conic projection in ArcMap v10.6.1, by applying a nearest neighbour re-sampling technique for discrete layers and a bilinear interpolation re-sampling technique for continuous layers. Then, by applying the same re-sampling techniques, we re-sampled all raster layers to 250m resolution, a resolution commonly used to model habitat suitability and biodiversity hotspots at continental and global extents (Rondinini et al. 2011).

Species select their environmental resources and conditions at different spatial scales (McGarigal et al. 2016, Macdonald et al. 2018, Macdonald et al. 2019). To investigate scalar relationships between sampled species and covariates, we calculated each metric at eight

different scales, by using circular buffers of 250m, 500m, 1,000m, 2,000m, 4,000m, 8,000m, 16,000m and 32,000m radii, centred on each camera trap location.

### *2.4.*Covariate selection and variance partitioning

Since zero-inflation of explanatory variables are likely to cause inaccurate parameter estimations and unreliable inferences (Martin et al. 2005), we removed poorly sampled covariates occurring at <10% of camera stations to avoid assessing unrepresentative habitat features. To investigate the most representative scales for sampled species, we performed Canonical Correspondence Analysis (CCA) (ter Braak 1986, McGarigal et al. 2000) independently at each scale for each covariate, using the vegan package (Oksanen et al. 2018) in R v3.5.1 (R Core Team 2018). For each covariate, we retained the scale whose univariate CCA showed the highest canonical eigenvalue (Borcard et al. 1992).

We then assessed multicollinearity by calculating Pearson's correlation coefficient between all covariate pairs. When two covariates were highly correlated ($|r| \geq 0.7$), we dropped the covariate whose univariate CCA showed the lowest adjusted-$R^2$ (Guisan and Zimmermann 2000). We selected the final set of covariates by performing forward selection for each group of covariates, retaining only the significant ones ($p < 0.001$) (Cushman and McGarigal 2004).

It should be noted that the preliminary covariates were selected to represent as wide habitat gradients as necessary to evaluate biodiversity distribution. The additional steps to assess composition and configuration metrics of the original covariates, as well as to evaluate their representative spatial scales, were performed to analyse how, and at what scale, habitat factors that we had already identified as fundamental for biodiversity, affected its geographic distribution.

To investigate the relative contribution of each group of covariates, we performed a variance partitioning analysis (Borcard et al. 1992) using the vegan package. Variance partitioning quantifies the independent contribution of each group of covariates to the global variance explained, as well as the shared variance explained by interacting combinations of covariates (Borcard et al. 1992, Cushman and McGarigal 2004).

### *2.5.*Modelling species richness

Since our data were counts of species detected at each camera station, to model species richness we performed Poisson generalized linear model (GLM) for each species using covariates at their representative spatial scales, in R v3.5.1. GLMs are commonly used regression models that allow the response variable to have different distributions than the normal one (Guisan and Zimmermann 2000), and Poisson distribution is used when the response variable is composed of abundance data (Vincent and Haworth 1983), as in our case. Projected models were reclassified to binary form, with zero and negative value pixels treated as absences and pixels with positive values as presences. Finally, single-species presence-absence maps were summed to predict species richness (Grand et al. 2004).

To evaluate the performance of the multi-species model, we trained the models with 80% of camera trap stations, and used the remaining 20% for validation. The multi-species model was validated by performing GLM between the modelled and the empirical number of species sampled at camera stations, and calculating the Nagelkerke-pseudo-$R^2$. The Nagelkerke-pseudo-$R^2$ (Nagelkerke 1991) is an index, ranging from 0 to 1, that provides a measure of the goodness-of-fit of logistic regressions. It is important to specify that, differently from linear regressions for which $R^2$ is a real measure of the goodness-of-fit that calculates models' explained variance, for logistic regressions a similar measure does not

exist. However, pseudo-$R^2$ is a relative measure of how well a model explains the data, and can be used to compare different models. Additionally, we compared the performance of the multi-species model with a "null" model obtained by summing the IUCN geographic range layers of the sampled species (IUCN 2019). The range layers considered were solely the polygons of the extent of occurrence (EOO) in which the species were considered extant and resident. We performed GLM between the number of species predicted by the IUCN model and the empirical number of species at camera stations, and calculated the Nagelkerke-pseudo-$R^2$.

### *2.6.* Gap analysis and species importance

We quantified the amount of protected biodiversity by calculating the ratio between the cumulative number of species predicted within protected areas (i.e., the number of species obtained by summing the number of species predicted within each pixel encompassed within protected areas) and the cumulative number of species predicted in the study area (i.e., the number of species obtained by summing the number of species predicted within each pixel in the study area) (Grand et al. 2004). Hence, we evaluated the effectiveness of protected areas and highlighted where additional ones should be implemented to fill the gaps in habitat protection.

We assessed the importance of sampled species as indicators of biodiversity by performing GLMs independently for each species. Predictor and response variables were random samples of points, in number proportional to the pixels representing the study area (1: 1,000), selected from the species richness model and from each single-species presence-absence model, respectively. We investigated deviance explained by each model and ranked each species according to its ability to predict overall biodiversity.

### *2.7.*Drivers of biodiversity patterns

To evaluate how well our model predicted empirical species richness for each country, we performed GLM between the modelled and empirical number of species at the camera traps used for validation. Then, using the residual values of each location, we performed ANOVA to test for significant differences between countries, followed by Tukey's test to assess which countries were significantly different from the others.

We assessed the drivers of predicted biodiversity patterns by sampling 10,000 random points from the species richness model and from covariate layers to evaluate our four hypotheses (i.e., climatic, low human pressure, protected status and interactions between them). The tested covariates were mean annual temperature and mean annual precipitation (Fick and Hijmans 2017) for the climatic hypothesis, human footprint (WCS and CIESIN 2005) and roughness for the anthropic hypothesis, and protected areas (UNEP-WCMC and IUCN 2022) for the management hypothesis. Roughness layer was derived from a digital elevation model (Jarvis et al. 2008) and obtained by applying geomorphometric transformations (Evans et al. 2014), and was considered as a proxy of landscape inaccessibility to human activities (Cushman et al. 2017). We followed the same procedure we employed to model species richness, to re-project and re-sample the layers. Next, we performed linear model (LM) to assess the ecological relationships between modelled species richness and the aforementioned covariates. Last, we assessed changes in predicted species richness as functions of increasing each explanatory variable from the $10^{th}$ to the $100^{th}$ percentile, while holding all other covariates at their median value, illustrating which covariate had the strongest effect in driving biodiversity richness (Wasserman et al. 2012).

## 3. Results

### *3.1.*Sampling effort

Camera traps were deployed at 1,384 camera trap stations in 15 landscapes across 7 countries in Southeast Asia, yielding a combined sampling effort of 115,389 trap nights (average trap nights per camera trap station= 83.4±2.6 SE). We sampled 90 species, of which 74 were mammals, 15 terrestrial birds and 1 reptile. In addition, we also sampled 4 anthropic elements: forest guards, including all rangers and patrollers in protected areas, humans (all other people excluding park staff), domestic dogs and domestic cattle. We assessed the effect of these anthropic elements on the habitat selection of sampled species, but we discounted them when we modelled species richness. The maximum numbers of species sampled per station were 15 in Cambodia, 15 in India, 17 in Laos, 14 in Myanmar, 13 in Nepal, 21 in Peninsular Malaysia and 33 in Thailand (Table 1). The list of sampled species is provided in Table S2, and the number of species sampled per country is reported in Table S3.

**Table 1.** Sampling effort in Southeast Asia. Shown are sampling locations per country, number of camera trap stations, total number of trap nights and average number of trap nights per camera trap station and maximum number of species detected per sampling location.

| Countries | Sampling locations | Camera trap stations | Total # trap nights (mean) | Max species detected |
|---|---|---|---|---|
| Cambodia | Cardamom | 76 | 7,740 (101.84) | 15 |
| | PPWS | 77 | 5,283 (68.61) | 12 |
| India | Dampa | 74 | 4,962 (67.05) | 15 |
| Laos | NPNV | 78 | 3,729 (47.81) | 17 |

| | | | | |
|---|---|---|---|---|
| | PD | 78 | 3,666 (47.00) | 14 |
| | PS | 72 | 3,446 (47.86) | 11 |
| Myanmar | Htamanthi | 162 | 14,546 (89.79) | 14 |
| Nepal | Langtang National Park | 83 | 7,105 (85.60) | 13 |
| Peninsular Malaysia | Ulu Muda Grid 1 | 76 | 9,251 (121.72) | 18 |
| | Ulu Muda Grid 2 | 54 | 5,995 (111.02) | 21 |
| Thailand | Dong Yai Wildlife Sanctuary | 58 | 2,709 (46.71) | 14 |
| | Khao Yai National Park | 67 | 6,266 (93.52) | 23 |
| | Pang Sida National Park | 163 | 17,750 (108.90) | 30 |
| | Ta Phraya National Park | 56 | 4,636 (82.79) | 14 |
| | Thap Lan National Park | 210 | 18,305 (87.17) | 33 |
| Total | | 1,384 | 115,389 (83.37) | |

### *3.2.* Univariate scaling and forward selection analysis

Scale optimization revealed relatively high homogeneity in the scales selected across predictor variables. The broadest scale (32,000m) had the strongest relationship with species occurrence patterns for 26 covariates (68.4%). Overall, broad scales (16,000m and 32,000m) were selected for 32 covariates (84.2%), while medium scales (2,000m, 4,000m and 8,000m) and fine scales (250m, 500m and 1,000m) were both selected for two covariates (5.3%) (Table 2).

Two covariates were poorly represented and were excluded due to lack of predictive power. We also dropped 33 highly correlated covariates. We then applied forward selection to identify the significant covariates from each group. The stringent threshold did not lead to the removal of any covariates. The final set was composed of 11 covariates (Table 2).

**Table 2.** Preliminary set of covariates and corresponding representative spatial scales.

| Class | Covariate | Metric | Best scale (m) |
|---|---|---|---|
| Landscape | Landcover | CWED | 16,000* |
| | Bare ground | PLAND | 32,000* |
| | | GYRATE_AM | 32,000* |
| | Croplands | PLAND | 32,000* |
| | | GYRATE_AM | 32,000* |
| | Mosaic areas | PLAND | 32,000* |
| | | GYRATE_AM | 32,000 |
| | Forested areas | PLAND | 32,000 |

| | | | |
|---|---|---|---|
| | | GYRATE_AM | 4,000 |
| | Shrublands and grasslands | PLAND | 32,000 |
| | | GYRATE_AM | 16,000* |
| | Permanent snow and ice | PLAND | NA† |
| | | GYRATE_AM | NA† |
| | Urban areas | PLAND | 32,000* |
| | | GYRATE_AM | 32,000* |
| | Water | PLAND | 32,000 |
| | | GYRATE_AM | 32,000 |
| | Non-forested areas | PLAND | 32,000* |
| | | GYRATE_AM | 32,000 |
| | Open forests | PLAND | 8,000* |
| | | GYRATE_AM | 16,000* |
| | Closed forests | PLAND | 32,000* |
| | | GYRATE_AM | 32,000* |
| Anthropic | Forest loss | FM | 250* |
| | | SD | 250* |
| | Human footprint | FM | 16,000* |
| | | SD | 32,000* |
| | Human population density | FM | 32,000* |
| | | SD | 16,000* |
| | Protected areas | PLAND | 32,000* |

| | | | |
|---|---|---|---|
| | | GYRATE_AM | 32,000 |
| Topographic | Elevation | FM | 32,000 |
| | | SD | 32,000* |
| | Compound topographic index | FM | 32,000* |
| | | SD | 32,000* |
| | Slope position | FM | 32,000 |
| | | SD | 32,000* |
| Climatic | Mean annual precipitation | FM | 16,000* |
| Spatial | X | NA | NA* |
| | Y | NA | NA |
| | X*Y | NA | NA* |
| | $X^2$ | NA | NA* |
| | $Y^2$ | NA | NA* |
| | $X^2$*Y | NA | NA* |
| | X*$Y^2$ | NA | NA* |
| | $X^2$*$Y^2$ | NA | NA* |
| Camera effort | Camera trap nights | NA | NA |

† Covariates excluded from the model because occurring in <10% of the camera trap stations.

* Covariates excluded from the model after the multicollinearity analysis.

### *3.3.*Variance partitioning and species-habitat relationships

The covariates selected for the CCA explained 32.5% of the variance in species detections. Camera effort had the strongest relationship (78% of the explained variance individually and 80% conditionally), followed by landscape covariates (11% individually and 18% conditionally). Topographic and spatial covariates each explained individually only 1% of the variance (Figure 2).

The CCA highlighted that species assemblages were primarily driven by environmental components (protected areas, mosaic areas, non-forested areas, shrublands and grasslands, water and forest cover). However, topographic (elevation) and spatial (Y) gradients also had substantial relationships with assemblage patterns (Figure 3, Table S4).

Analysis of the biplot revealed assemblages of species associated, at varying degrees, with specific habitat features (Figure 3). A conspicuous group was clearly driven by elevational gradient, differentiating low-elevation and high-altitude species. Another group of species was mainly associated with forests, mosaic areas and water, while a more heterogeneous group of species was influenced by slope position and non-forested areas.

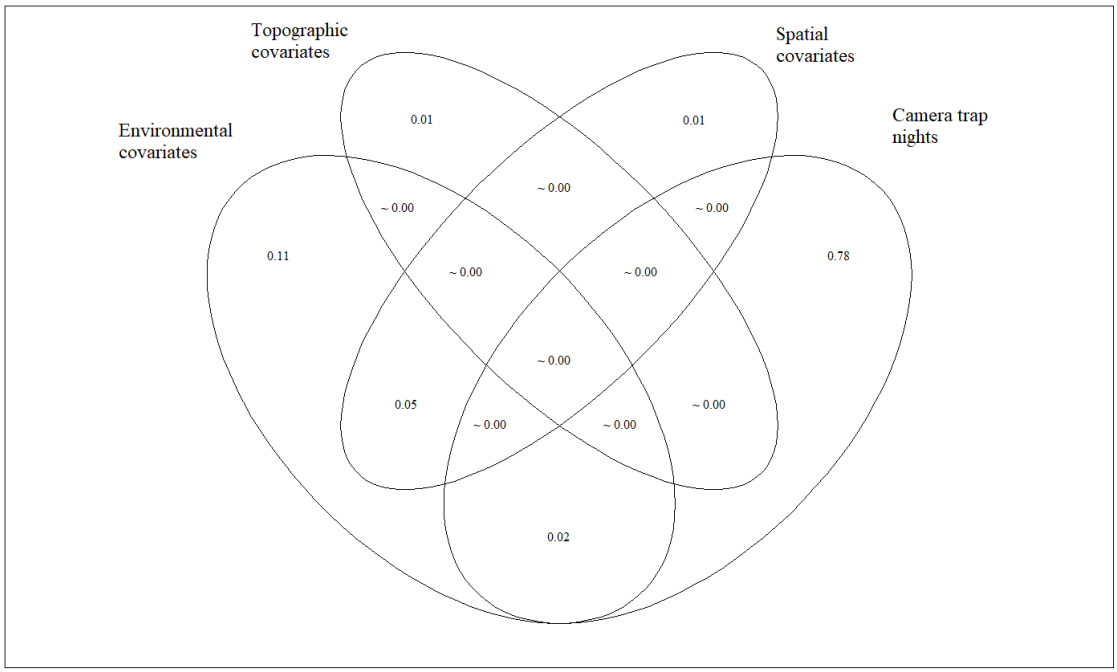



**Figure 2.** Variance partitioning analysis, showing the variance explained by different subsets of covariates. Shown are the relative values. Absolute variance explained= 0.325. Absolute residual variance= 0.675.

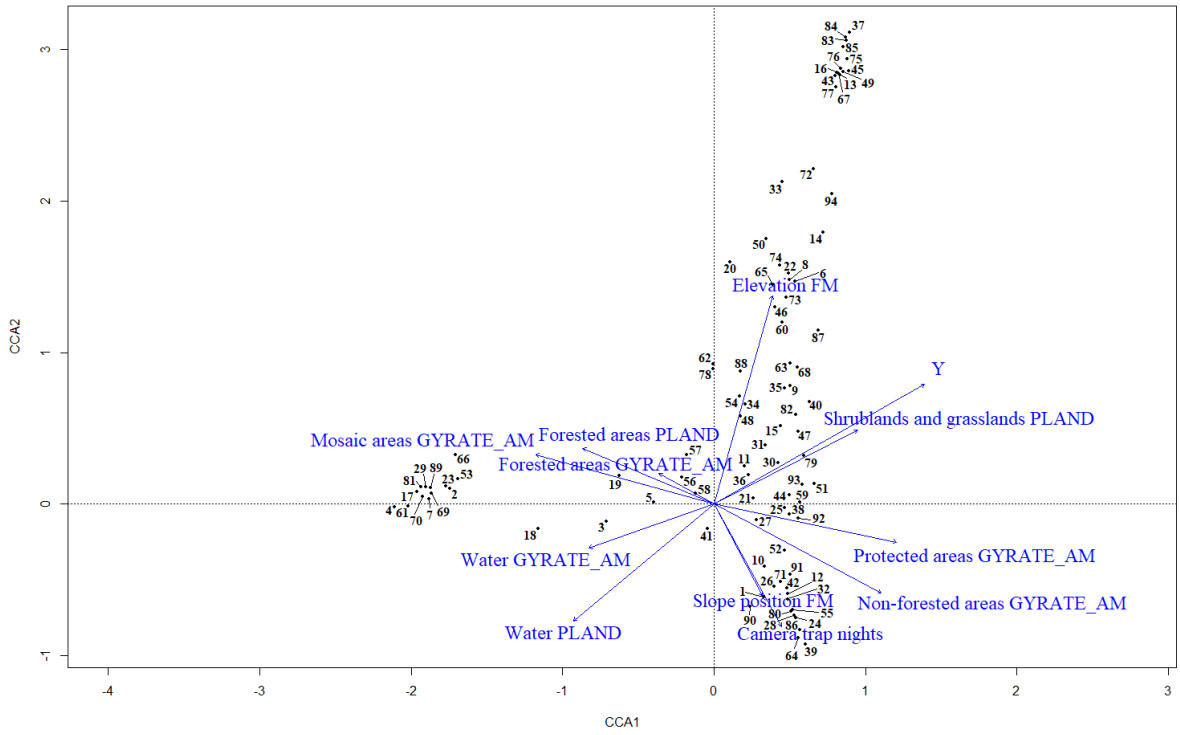



**Figure 3.** Biplot showing the relationships between species and habitat covariates. See Table S2 for species' codes.

### *3.4.*Single-species models

Single-species models showed that most species were positively associated with forests (GYRATE_AM and PLAND), protected areas, shrublands and grasslands and non-forested areas, with 72.1%, 65.1%, 64.0%, 64.0% and 57.0% of species showing positive coefficients for these variables, respectively. No pattern was shown by other covariates, with almost half of the species showing negative coefficients (Table S5).

### *3.5.*Model comparison, gap analysis and species importance

We identified the factors associated with species richness, and used them to predict biodiversity across the study area (Figure 4). We validated the multi-species model using independent test data, showing a positive model fit (Nagelkerke-pseudo-$R^2$= 0.14, p< 0.0001) (Figure S1). The IUCN model showed a much lower model fit (Nagelkerke-pseudo-$R^2$= 0.0017, p= 0.50) (Figure S2) and failed to highlight known biodiversity hotspots (Figures S3-S4).

The gap analysis revealed that 18.6% of the predicted biodiversity occurred within protected areas. The countries showing the lowest proportions of protected biodiversity were Singapore, Bangladesh and India, preserving 6.51%, 12.23% and 12.93% of their biodiversity, respectively (Table S6). Surprisingly, even the most effective protected areas preserved only a small percentage of biodiversity, with the top three protecting, respectively, 1.37%, 1.35% and 0.84% of the total cumulative number of species, and 7.38%, 7.27% and 4.52% of the cumulative number of species occurring within protected areas (Table S7). In addition, we identified biodiverse areas that are not protected, and whose preservation would dramatically increase the amount of protected biodiversity (Figure 4).

The biodiversity indicator analysis revealed that species richness was mainly predicted by carnivores, with ten carnivores occurring among the first twenty indicator species. These included members of the Viverridae, Herpestidae, Canidae, Ursidae, Mustelidae and Felidae families, with Viverridae as the most predictive indicator of overall biodiversity. The role of artiodactyls was also noteworthy, with three Bovidae, two Cervidae and mouse-deer species occurring in the first twenty positions. Among the most informative indicators were also rat species, two gallinaceous birds and one reptile (Table S8).

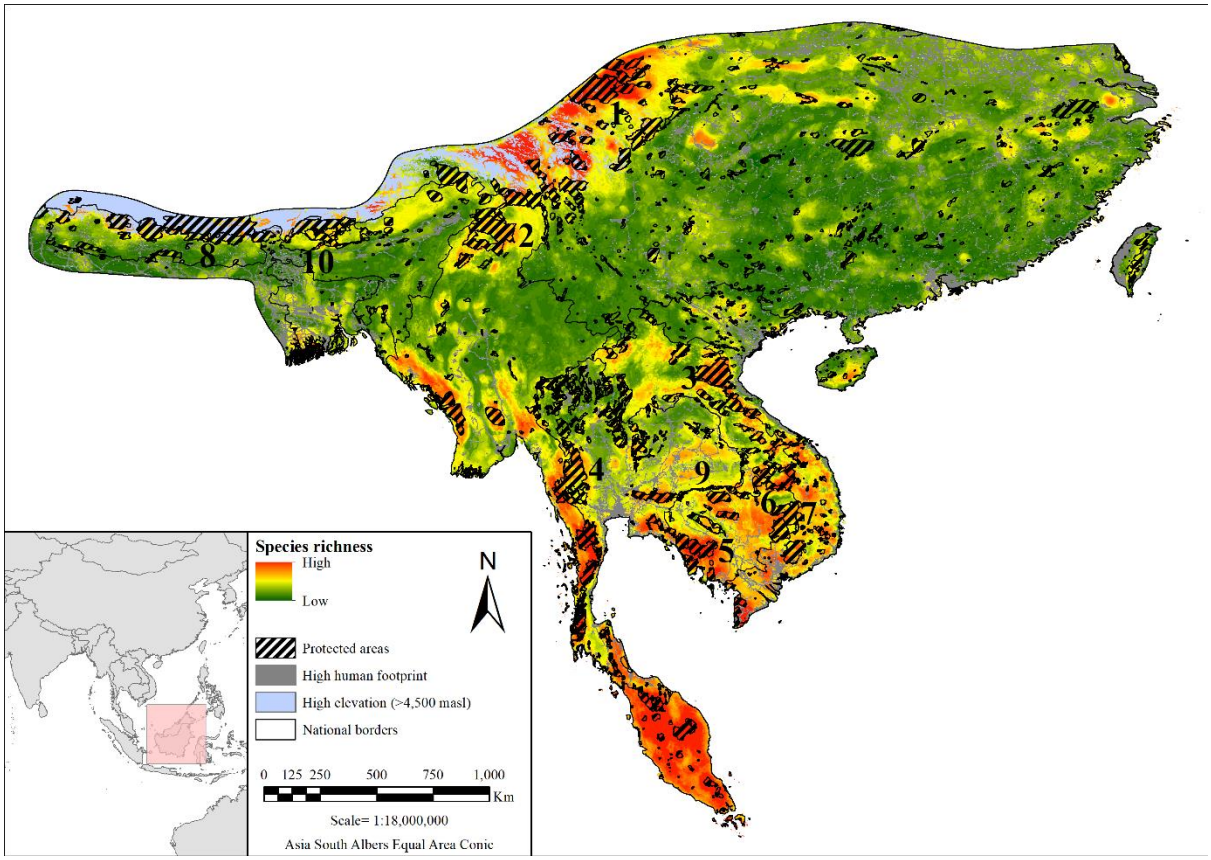



**Figure 4.** Predictive map of species richness in Southeast Asia, showing protected areas overlaid in black hash marks. Numbers represent codes of the main systems of protected areas reported in Table S7.

### *3.6.*Drivers of biodiversity patterns

ANOVA on the residuals of the GLM between the modelled and the empirical number of species showed a significant difference among countries ($F_{(6, 270)}$= 72.5, p< 0.0001). Specifically, Peninsular Malaysia showed positive distribution of the residuals, meaning that the country has more species richness than predicted, given its environmental conditions (Figure S5). Contrarily, Nepal, India and, to a lesser degree, Laos and Myanmar, showed negative residuals, meaning that our model overestimated the number of species in these

countries. Additionally, Tukey's test revealed significant differences between almost all the country dyads, with the exception of Thailand-Cambodia and Myanmar-Laos (Table S9).

We found that climatic covariates contributed the most to the predicted patterns of species richness. Mean annual temperature was negatively associated with species richness, although it showed nonlinear relationship (Figure 5), while precipitation showed a positive relationship (Figure 6). Protected areas also showed positive relationship with biodiversity. Anthropic factors showed negative but weaker effect on patterns of biodiversity than both climatic factors and protected status (Table 3).

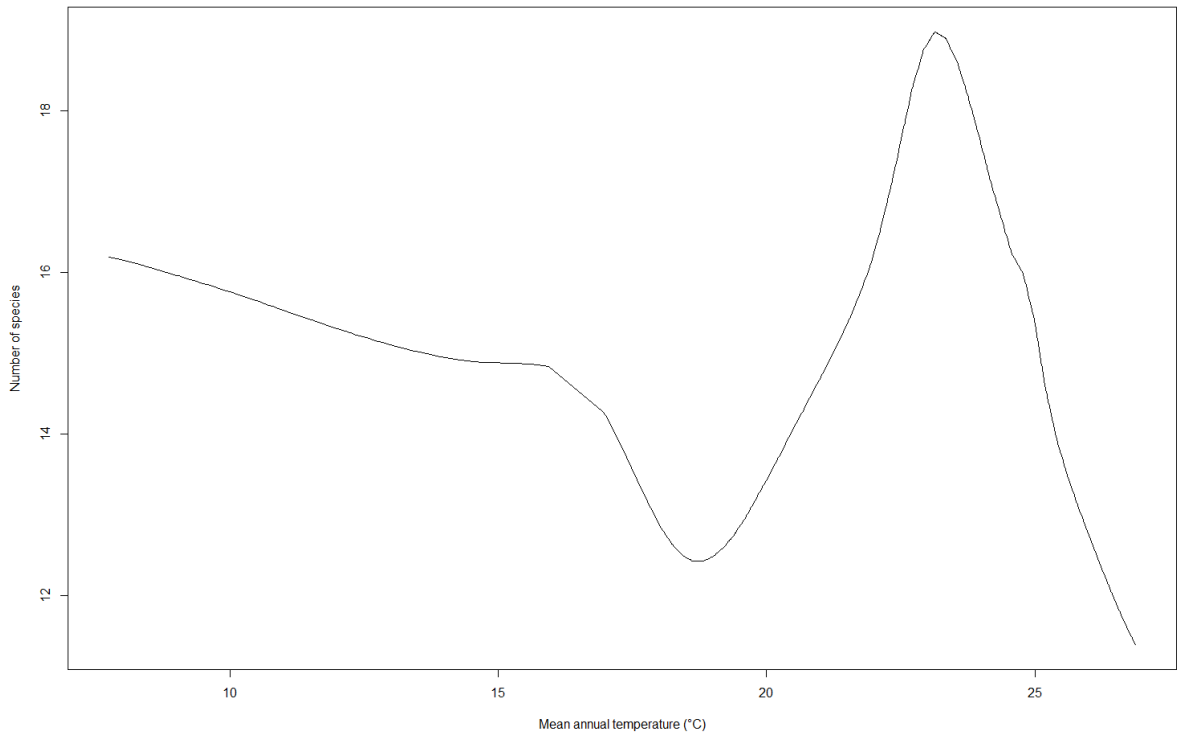



**Figure 5.** Relationship between the sampled number of species and mean annual temperature of the sampling locations.

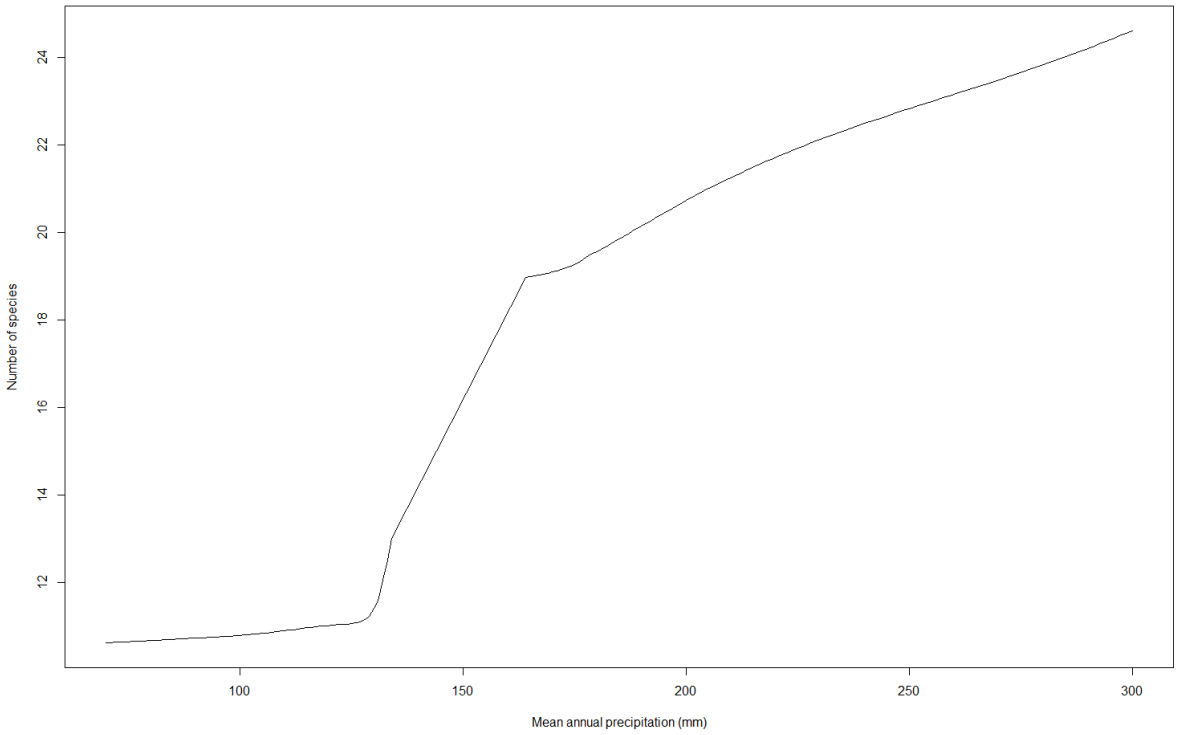



**Figure 6.** Relationship between the sampled number of species and mean annual precipitation of the sampling locations.

Plotting changes in species richness as function of increasing values of explanatory covariates from the 10th to the 100th percentile corroborated the importance of protected status, and the negative impact of accessibility for anthropic activities on biodiversity (Figure 7).

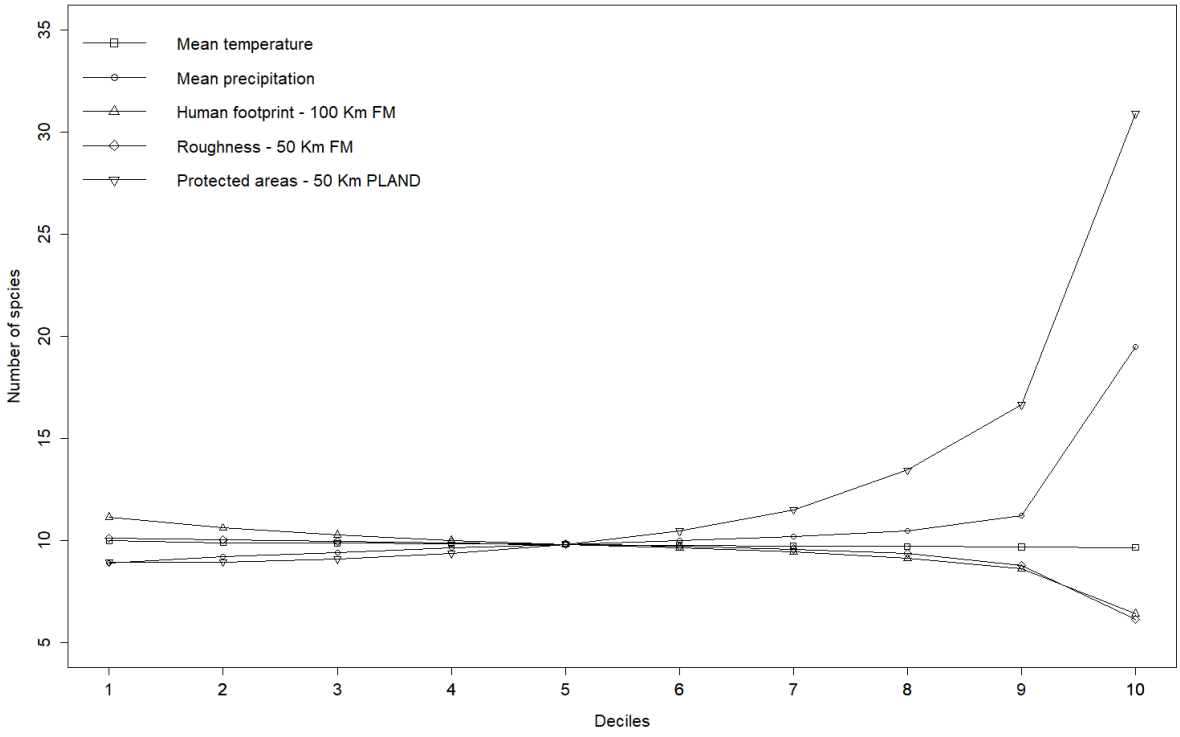



**Figure 7.** Changes in species richness as the explanatory covariates increase from the 10th to the 100th percentile, holding all other covariates constant at their medians.

**Table 3.** Sources, native resolutions, coefficients and p-values of the explanatory covariates analysed to investigate the drivers of the predicted biodiversity patterns.

| Covariate | Source | Native resolution | Coefficient | P-value |
|---|---|---|---|---|
| Intercept | NA | NA | 11.46 | <0.0001 |
| Temperature | Fick and Hijmans (2017) | 1,000m | -5.18 | <0.0001 |
| Precipitation | Fick and Hijmans (2017) | 1,000m | 5.18 | <0.0001 |
| Human footprint | WCS and CIESIN (2005) | 1,000m | -1.02 | <0.0001 |
| Roughness | Jarvis et al. (2008); Evans et al. (2014) | 90m | -0.62 | <0.0001 |
| Protected areas | UNEP-WCMC and IUCN (2022) | NA (polygons) | 3.68 | <0.0001 |

## 4. Discussion

We provide new insights into spatial patterns and drivers of species richness in Southeast Asia, highlighting that biodiversity is not evenly distributed across the region, but exhibits a complex pattern with two large primary biodiverse areas: the Thai-Malay Peninsula (TMP) and the mountainous region of Southwest China (MSC), with a scattering of smaller biodiverse areas between them. Hughes (2017a) described similar pattern, producing distinct models for different taxa in Southeast Asia, although the MSC was highlighted as hotspot only in the birds' model. However, it is likely that a more taxonomically inclusive model would have corresponded more closely with ours.

The habitat features of the two main hotspots are distinct, leading to unique communities of species. Yet both have in common landscapes characterized by high environmental and climatic heterogeneity. In the TMP, the highly productive, wet and stable tropical rainforest climate, coupled with strong topographic heterogeneity along altitudinal gradients, likely facilitate high levels of biodiversity (Wikramanayake 2002). Similarly, the MSC environment is characterized by ridge-valley complexes shaped by large rivers, and divided by tall mountain ranges. This landscape is characterized by high environmental and climatic heterogeneity, promoting ecological niche diversity, and supporting a large variety of species in a relatively small area (Wikramanayake 2002).

### *4.1.*Species-habitat relationships

The multi-scale optimization approach revealed broad-scale selection for almost all covariates. Given the vastness of our study area, this result likely reflects the fact that habitat loss, poaching pressure and other human perturbations are broad-scale and that remaining areas of high biodiversity are in large remnant ecosystems. However, this result might also be due to an insufficiency of the spatial scales assessed in the multi-scale optimization. As pointed out by Jackson and Fahrig (2015) if, in a multi-scale assessment, a species selects most of the habitat factors at the largest or at the smallest scale, this might be symptomatic of a limitation in the number, and in the magnitude, of the scales assessed. In other words, the true representative scale could be outside the range of evaluated ones. While acknowledging this possibility, insofar as it might be a potential limitation of our study, it can arise only if the true, but undetected, representative scales are broader than those we report, which extended up to 32km. It is very unlikely that finer un-sampled scales were important, given

the scale range we analysed. The main implication of our results is that species richness is mostly related to environmental and anthropic features at very broad scales.

Southeast Asian environmental features have a predominant influence on species richness. Presence of forests and water are particularly important in explaining species presence, with both composition (habitat percentage) and configuration (aggregation index) metrics retained in the final model. Variance partitioning supported these results, showing that model's variance was mainly explained by environmental covariates, while topographic and spatial factors only marginally contributed to explain species assemblages.

The CCA biplot elucidates the relationships of individual species with habitat features, and their contribution to shaping communities. Forest cover, non-forested areas, shrublands and grasslands, mosaic areas and water showed the highest eigenvalues. However, Y (latitude) and elevation also showed high values revealing their importance in shaping the distribution of some species, even with limited weights in the variance partitioning. This was not surprising, as it is widely acknowledged that species richness increases along altitudinal and latitudinal gradients, driven by combinations of different climatic and ecological conditions that, together, promote ecological niche diversity, allowing a variety of different species to occur in the same area (Willig and Presley 2016, Cuesta et al. 2017). Indeed, the biplot (Figure 3) clearly highlights that a number of species follow an elevational and latitudinal gradient. Additionally, two other distinct groups emerged: the first, on the left side of the biplot, driven mostly by forest cover, mosaic areas and water, and including species associated with forests such as banded palm civet (*Hemigalus derbyanus*), banded linsang (*Prionodon linsang*), Malay tapir (*Tapirus indicus*) and Southern pig-tailed macaque (*Macaca nemestrina*), and the second, on the bottom right corner of the biplot, driven by presence of protected areas, slope position and non-forested areas, and composed of a more

heterogeneous group of species including golden jackal (*Canis aureus*), tiger (*Panthera tigris*), Asian black bear (*Ursus thibetanus*) and Asian elephant (*Elephas maximus*).

Importantly, most of the species in this latter group are known to be highly sensitive to forest loss and anthropogenic killing. Their association with low elevation suggests their preference for warm tropical habitats. However, the joint association with non-forested areas indicates that very little extensive native forests persists at low elevations, and that it is concentrated in protected areas.

### *4.2.*Biodiversity indicators

The use of single indicator taxa to prioritise management actions has long been employed in conservation (Noss 1990, Rodrigues and Brooks 2007), although the assumption that preserving the habitat for a taxon would automatically guarantee protection for other sympatric species has been contested, and its effectiveness is debated (Cushman et al. 2010). We found that carnivores are a good surrogate of sampled terrestrial biodiversity, with several species from a variety of families being potential indicators. The role of carnivores as indicators of the sampled biodiversity likely reflects their role as umbrella species. Specifically, these species are among the most sensitive to habitat loss, mortality risk and other perturbations due to their large size, high vulnerability and large home ranges. Therefore, their occurrence in specific areas suggests that site has high ecological integrity and low levels of perturbation, associated with high richness in other species. Our results are in line with previous findings on the importance of carnivores as indicator and umbrella species (Carroll et al. 2001, Dalerum et al. 2008), and on the potential role of large ungulates and birds as indicators (Larsen et al. 2012).

However, we agree that prioritisation strategies based on single taxa might fail to highlight important conservation areas for rare or cryptic species. Hence, we advocate the use of a combination of indicators to cover broader ranges of habitat gradients, ensuring the protection of a larger number of species, as recommended for conservation planning (Carroll et al. 2001, Macdonald et al. 2012) and for landscape connectivity (Cushman and Landguth 2012).

### *4.3.*Model performance

The model validation revealed our species richness model to be a relatively good predictor of biodiversity, capturing the high species richness of the MSC, a known biodiversity hotspot (Myers et al. 2000). Considering that the camera survey was largely carried out in the TMP, this congruence adds confidence that our model successfully predicted biodiversity patterns across the region. We recommend future studies to undertake data collection in the MSC to test our predictions for that region.

Comparing our model with current knowledge of species distribution based on IUCN geographic range layers, revealed that our model performed much better based on the Nagelkerke-pseudo-$R^2$. The IUCN model, in addition to misclassifying the empirical species richness showing lower Nagelkerke-pseudo-$R^2$, failed to highlight some of the important biodiversity hotspots, such as the MSC, and classified as hotspot less biodiverse areas. Similar discrepancies were also highlighted by Hughes (2017a). This lack of congruence raises concerns, as the IUCN layers are often the main source to produce multi-species assessments (Rondinini et al. 2011). We argue that models based on extensive empirical data and robust analytical approaches are more accurate (Di Marco et al. 2017), and that continued

investment in direct biodiversity monitoring is essential to collect reliable empirical data (Cushman and McKelvey 2010).

Comparing our model with the IUCN model, we found that the region between the Chinese province of Yunnan, Central Myanmar and East India, differed the most (Figure S4). Here the IUCN model predicted more species than ours. The MSC also showed strong discrepancy, with our model predicting more species than the IUCN one. This suggests that other factors, likely very high past deforestation, are responsible for the lower species richness in these regions.

### *4.4.*Drivers of biodiversity pattern

We found significant differences between countries in the residuals of predicted biodiversity, suggesting potential geopolitical causes behind these discrepancies. Peninsular Malaysia show higher observed biodiversity than predicted by our model, while Nepal and India have lower observed biodiversity. This suggests that differences in factors distinct than those we evaluated here are also critical in driving regional biodiversity. Specifically, it is likely that Peninsular Malaysia has higher biodiversity partially due to better governance, rule of law or enforcement, while the overestimation of biodiversity in Nepal and India might reflect greater ineffectiveness in management and enforcement. We recommend future work to highlight the geopolitical factors affecting species richness in these countries.

Climatic factors were revealed as the main drivers of the overall spatial pattern. Temperature and precipitation were strongly associated with predicted biodiversity, the former showing a negative, nonlinear, relationship (Figure 5), and the latter with a positive effect (Figure 6). We suggest these results to be due to the importance of the MSC in our model, where the lowest temperatures across the study area are recorded. Protected status

also have strong positive effect, as demonstrated by the quantile plot showing that changes in protection have the largest impact on increasing species richness (Figure 7).

Hence, the current pattern of biodiversity is the result of the combination of climatic and, consequently, environmental conditions, with protection status. Human footprint and topographic roughness were associated with the largest decreases in modelled species richness, indicating strong reduction of biodiversity in areas highly accessible to human activities. Overall, these results show a complex pattern of environmental and anthropic factors driving local patterns of species richness, within broader gradients defined by climatic conditions.

### *4.5.* Gap analysis and conservation recommendations

We revealed poor match between existing protected areas and regions of predicted high biodiversity, suggests two conservation imperatives: (1) to extend the protected areas already existing, as well as the corridors linking them (Cushman et al. 2018, Kaszta et al. 2020b), and (2) to encourage cultural tolerance of coexistence and an ethic of wildlife protection (Western et al. 2019).

In line with classic biogeographic thinking, it is desirable to expand the protected areas network into a vaster, more connected system. One of the main hotspots, the MSC, has very limited coverage by protected areas. These should be extended, particularly in Southern Yunnan, including the Indian state of Assam and northern Myanmar, and designed to provide stepping-stones along the major connectivity corridors of the region (Kaszta et al. 2020b).

Biodiversity in the southern extent of the study area is also most lacking in protection, despite its importance. Almost none of the predicted highly biodiverse areas in Cambodia fell within protected areas. We also identified several critical areas occurring across countries,

underlining the importance of transboundary conservation programmes (Kaszta et al. 2020b). The biodiverse region spanning the border between Vietnam and Laos is only partially protected. Here, the Vietnamese system should be expanded to cover broader portions, especially in the south where we predicted large biodiversity areas but protection is inadequate and human impact is high. Myanmar shares a hotspot with Thailand, but only the Thai side is substantially protected. Similarly, the western coast of Myanmar adjoining Bangladesh is mostly unprotected despite its very high biodiversity. These areas are also identified as the most important nodes in a regional connectivity network for clouded leopard (Kaszta et al. 2020b). Finally, Peninsular Malaysia has the highest levels of species richness, but only a small portion is protected, despite its vulnerability to isolation from other areas of species richness due to its long, thin and vulnerable corridors (Kaszta et al. 2020b).

### *4.6.*Scope and limitations

We selected the final set of covariates by implementing forward selection analysis, a method that adds variables to the model one by one, until any variable not included in the model can add any significant contribution to the model itself (Chowdhury and Turin 2020). However, this method has often been criticised for choosing nuisance variables rather than true variables (Smith 2018), and other metrics (e.g., AUC and TSS) are often preferred for variable selection. In this work, a forward selection approach has been preferred over other methods because it can accommodate multiple response variables, such as the different sampled species that we analysed in a multivariate modelling framework. However, we recognise the limitations of this approach, and acknowledge that a different variable selection technique might have led to a different set of covariates.

Additionally, we also acknowledge the limitations of having used ANOVA test on the residuals of the GLM performed between predicted and empirical number of species to evaluate differences between countries, since residuals might not have been normally distributed, a stringent criteria for ANOVA test.

A possible source of bias in our analyses was also given by the clustered nature of the camera trap surveys, deployed mostly within protected areas and failing to describe the full gradients of the habitat factors used to model species richness. Although in the data pre-processing we removed all the camera traps closer than 250m from each other (i.e., falling within the same pixel) by retaining only one of them, it is still likely that the clustered nature of the camera traps was spatially autocorrelated. Spatial autocorrelation might have affected our model by leading to overestimation and underestimation of the suitability of certain regions, leading to inaccurate model predictions. Additionally, the analysis of the residuals of the GLM between the modelled and the empirical number of species at the camera trap stations revealed clear groups between countries (Figure S5), which might have suggested the use of models allowing the incorporation of random effects, such as Generalised Linear Mixed Models (GLMMs). While acknowledging that we could have fitted different GLMMs independently for each species, we decided to apply a multivariate approach where we first highlighted the most important covariates to define the simultaneous abundance of the sampled species, and then fitted GLMs for each species, but using the same set of covariates. We acknowledge the possible limitations of this approach, and suggest future research to investigate the differential outcomes of other modelling frameworks.

## 5. Conclusions

Implementing and managing protected areas is not easy, especially over an area as vast as the one we assessed. In addition, the preservation of a high number of species as a consequence of their occurrence within protected areas is not a foregone conclusion (Hallmann et al. 2017). In recommending an expansion of the protected areas network we are mindful of the hazard of creating paper parks if protection does not advance hand in hand with enforcement, enhanced willingness to embrace co-existence, adequacy of finances, all set in the framework of regional, transboundary, conservation strategies that are often either unthought-out or unfulfilled (Di Minin and Toivonen 2015, Western et al. 2019). Our finding that only a small portion of biodiversity is protected, prompts us to prioritise thinking, and action, about both expanding that network and identifying a wider portfolio of complementary conservation strategies. In a region like Southeast Asia, where over-exploitation of natural resources and deforestation are the main drivers of biodiversity loss (Sodhi et al. 2004, Sodhi and Brook 2006), we argue for a holistic, integrated approach between conservationists, local stakeholders and governmental institutions, supported by trans-disciplinary evidence and insight (Macdonald 2019). Only through consideration of the well-being of both wildlife and people, at the intersection of conservation and social justice, can we be hopeful for co-existence (Vucetich et al. 2018).

As argued by Sodhi and Brook (2006), a route to tangible conservation outcomes in Southeast Asia is via the assessment and integration of social issues with conservation actions. In addition, among the main components likely to encourage coexistence are perceptions of benefits and costs, and rectifying both real and imagined imbalances between them (Dickman et al. 2011, Western et al. 2019). Hence, although providing local, fine-scale, socio-economic recommendations to foster conservation practices in Southeast Asia goes beyond the scope of this paper, overall our results should be framed by the wider social and

geopolitical circumstances of the region, and we argue for conservation interventions that fully embrace the holism of human-wildlife coexistence: striving, not only for the conservation of species and habitats, but also for the economic benefits, equity, rights and active participation for people depending on these areas (Chazdon et al. 2009).

Here we implemented a multi-scale, multi-species modelling framework to reveal the areas richest in biodiversity in Southeast Asia. We demonstrated the importance of undergoing a paradigmatic switch from single-species to multi-species models to better understand conservation priorities. We hope that our results will contribute to conservation of terrestrial vertebrates in the region, highlighting where conservation actions should be prioritised. Further, the modelling framework that we illustrated here could usefully be implemented in other areas of the world, to highlight biodiversity hotspots and to identify conservation gaps with greater precision that has been possible hitherto. Finally, based on the evidences reported here, we make a strident call for transboundary conservation planning to protect natural habitats from the imminent risks of degradation, and for strategic efforts to identify and prioritise core habitats and connectivity linkages among them at broad, regional scales.

## 6. References

Azhar, B., N. Saadun, M. Prideaux, and D. B. Lindenmayer. 2017. The global palm oil sector must change to save biodiversity and improve food security in the tropics. Journal of Environmental Management **203**:457-466.

Beck, H. E., N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, and E. F. Wood. 2018. Present and future Koppen-Geiger climate classification maps at 1-km resolution. Scientific Data **5**.

Borcard, D., P. Legendre, and P. Drapeau. 1992. Partialling out the Spatial Component of Ecological Variation. Ecology **73**:1045-1055.

Bruner, A. G., R. E. Gullison, R. E. Rice, and G. A. B. da Fonseca. 2001. Effectiveness of parks in protecting tropical biodiversity. Science **291**:125-128.

Carroll, C., R. F. Noss, and P. C. Paquet. 2001. Carnivores as focal species for conservation planning in the Rocky Mountain region. Ecological Applications **11**:961-980.

Chazdon, R. L., C. A. Harvey, O. Komar, D. M. Griffith, B. G. Ferguson, M. Martinez-Ramos, H. Morales, R. Nigh, L. Soto-Pinto, M. van Breugel, and S. M. Philpott. 2009. Beyond Reserves: A Research Agenda for Conserving Biodiversity in Human-modified Tropical Landscapes. Biotropica **41**:142-153.

Chowdhury, M. Z. I., and T. C. Turin. 2020. Variable selection strategies and its importance in clinical prediction modelling. Family Medicine and Community Health **8**.

Cuesta, F., P. Muriel, L. D. Llambí, S. Halloy, N. Aguirre, S. Beck, J. Carilla, R. I. Meneses, S. Cuello, A. Grau, L. E. Gámez, J. Irazábal, J. Jácome, R. Jaramillo, L. Ramírez, N. Samaniego, D. Suárez-Duque, N. Thompson, A. Tupayachi, P. Viñas, K. Yager, M. T. Becerra, H. Pauli, and W. D. Gosling. 2017. Latitudinal and altitudinal patterns of plant community diversity on mountain summits across the tropical Andes. Ecography **40**:1381-1394.

Cushman, S. A., N. B. Elliott, D. Bauer, K. Kesch, L. Bahaa-el-din, H. Bothwell, M. Flyman, G. Mtare, D. W. Macdonald, and A. J. Loveridge. 2018. Prioritizing core areas, corridors and conflict hotspots for lion conservation in southern Africa. Plos One **13**.

Cushman, S. A., and E. L. Landguth. 2012. Multi-taxa population connectivity in the Northern Rocky Mountains. Ecological Modelling **231**:101-112.

Cushman, S. A., E. Macdonald, E. Landguth, Y. Malhi, and D. Macdonald. 2017. Multiple-scale prediction of forest loss risk across Borneo. Landscape Ecology **32**:1581-1598.

Cushman, S. A., and K. McGarigal. 2004. Hierarchical analysis of forest bird species-environment relationships in the Oregon Coast Range. Ecological Applications **14**:1090-1105.

Cushman, S. A., and K. S. McKelvey. 2010. Data on Distribution and Abundance: Monitoring for Research and Management. Spatial Complexity, Informatics, and Wildlife Conservation:111-129.

Cushman, S. A., K. S. McKelvey, B. R. Noon, and K. McGarigal. 2010. Use of Abundance of One Species as a Surrogate for Abundance of Others. Conservation Biology **24**:830-840.

Dalerum, F., M. J. Somers, K. E. Kunkel, and E. Z. Cameron. 2008. The potential for large carnivores to act as biodiversity surrogates in southern Africa. Biodiversity and Conservation **17**:2939-2949.

Di Marco, M., C. Rondinini, L. Boitani, and K. A. Murray. 2013. Comparing multiple species distribution proxies and different quantifications of the human footprint map, implications for conservation. Biological Conservation **165**:203-211.

Di Marco, M., J. E. M. Watson, H. P. Possingham, and O. Venter. 2017. Limitations and trade-offs in the use of species distribution maps for protected area planning. Journal of Applied Ecology **54**:402-411.

Di Minin, E., and T. Toivonen. 2015. Global Protected Area Expansion: Creating More than Paper Parks. Bioscience **65**:637-638.

Dickman, A. J., E. A. Macdonald, and D. W. Macdonald. 2011. A review of financial instruments to pay for predator conservation and encourage human-carnivore coexistence. Proceedings of the National Academy of Sciences of the United States of America **108**:13937-13944.

Evans, J. S., J. Oakleaf, S. A. Cushman, and D. Theobald. 2014. An ArcGIS toolbox for surface gradient and geomorphometric modeling, version 2.0-0. Available: http://evansmurphy.wix.com/evansspatial.

Fahrig, L. 1997. Relative effects of habitat loss and fragmentation on population extinction. Journal of Wildlife Management **61**:603-610.

Fick, S. E., and R. J. Hijmans. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology **37**:4302-4315.

Gaveau, D. L. A., D. Sheil, Husnayaen, M. A. Salim, S. Arjasakusuma, M. Ancrenaz, P. Pacheco, and E. Meijaard. 2016. Rapid conversions and avoided deforestation: examining four decades of industrial plantation expansion in Borneo. Scientific Reports **6**.

Global Forest Watch. 2020. Tree cover loss. Accessed on 17/07/2020 from www.globalforestwatch.org.

Grand, J., J. Buonaccorsi, S. A. Cushman, C. R. Griffin, and M. C. Neel. 2004. A multiscale landscape approach to predicting bird and moth rarity hotspots, in a threatened pitch pine-scrub oak community. Conservation Biology **18**:1063-1077.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling **135**:147-186.

Hallmann, C. A., M. Sorg, E. Jongejans, H. Siepel, N. Hofland, H. Schwan, W. Stenmans, A. Muller, H. Sumser, T. Horren, D. Goulson, and H. de Kroon. 2017. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. Plos One **12**.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L.

Chini, C. O. Justice, and J. R. G. Townshend. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science **342**:850-853.

Hearn, A. J., S. A. Cushman, B. Goossens, E. Macdonald, J. Ross, L. T. B. Hunter, N. K. Abram, and D. W. Macdonald. 2018. Evaluating scenarios of landscape change for Sunda clouded leopard connectivity in a human dominated landscape. Biological Conservation **222**:232-240.

Hoffman, M., K. Koenig, G. Bunting, J. Costanza, and K. J. Williams. 2016. Biodiversity Hotspots (Version 2016.1).*in* Zenodo, editor.

Hughes, A. C. 2017a. Mapping priorities for conservation in Southeast Asia. Biological Conservation **209**:395-405.

Hughes, A. C. 2017b. Understanding the drivers of Southeast Asian biodiversity loss. Ecosphere **8**.

IUCN. 2019. The IUCN Red List of Threatened Species. Version 2019-5. https://www.iucnredlist.org. Downloaded on 17 May 2019.

Jackson, H. B., and L. Fahrig. 2015. Are ecologists conducting research at the optimal scale? Global Ecology and Biogeography **24**:52-63.

Jarvis, A., H. I. Reuter, A. Nelson, and E. Guevara. 2008. Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT). http://srtm.csi.cgiar.org.

Johnson, D. H. 1980. The Comparison of Usage and Availability Measurements for Evaluating Resource Preference. Ecology **61**:65-71.

Kaszta, Z., S. A. Cushman, S. Htun, H. Naing, D. Burnham, and D. W. Macdonald. 2020a. Simulating the impact of Belt and Road initiative and other major developments in Myanmar on an ambassador felid, the clouded leopard, Neofelis nebulosa. Landscape Ecology **35**:727-746.

Kaszta, Z., S. A. Cushman, and D. W. Macdonald. 2020b. Prioritizing habitat core areas and corridors for a large carnivore across its range. Animal Conservation.

Larsen, F. W., J. Bladt, A. Balmford, and C. Rahbek. 2012. Birds as biodiversity surrogates: will supplementing birds with other taxa improve effectiveness? Journal of Applied Ecology **49**:349-356.

Lim, M. 2016. Governance criteria for effective transboundary biodiversity conservation. International Environmental Agreements-Politics Law and Economics **16**:797-813.

Macdonald, D. W. 2019. Mammal Conservation: Old Problems, New Perspectives, Transdisciplinarity, and the Coming of Age of Conservation Geopolitics. Annual Review of Environment and Resources, Vol 44 **44**:61-88.

Macdonald, D. W., H. M. Bothwell, A. J. Hearn, S. M. Cheyne, I. Haidir, L. T. B. Hunter, A. Kaszta, M. Linkie, E. A. Macdonald, J. Ross, and S. A. Cushman. 2018. Multi-scale habitat selection modeling identifies threats and conservation opportunities for the Sunda clouded leopard (Neofelis diardi). Biological Conservation **227**:92-103.

Macdonald, D. W., H. M. Bothwell, Ż. Kaszta, E. Ash, G. Bolongon, D. Burnham, Ö. E. Can, A. Campos-Arceiz, P. Channa, G. R. Clements, A. J. Hearn, L. Hedges, S. Htun, J. F. Kamler, K. Kawanishi, E. A. Macdonald, S. W. Mohamad, J. Moore, H. Naing, M. Onuma, U. Penjor, A. Rasphone, D. Mark Rayan, J. Ross, P. Singh, C. K. W. Tan, J. Wadey, B. P. Yadav, and S. A. Cushman. 2019. Multi-scale habitat modelling identifies spatial conservation priorities for mainland clouded leopards (Neofelis nebulosa). Diversity and Distributions **0**:1-16.

Macdonald, D. W., D. Burnham, A. E. Hinks, and R. Wrangham. 2012. A problem shared is a problem reduced: seeking efficiency in the conservation of felids and primates. Folia Primatol (Basel) **83**:171-215.

Martin, T. G., B. A. Wintle, J. R. Rhodes, P. M. Kuhnert, S. A. Field, S. J. Low-Choy, A. J. Tyre, and H. P. Possingham. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecology Letters **8**:1235-1246.

McGarigal, K., S. Cushman, and S. G. Stafford. 2000. Multivariate statistics for wildlife and ecology research. Pages 1 online resource (xiii, 283 pages). Springer,, New York.

McGarigal, K., S. A. Cushman, and E. Ene. 2012. FRAGSTATS v4: spatial pattern analysis program for categorical and continuous maps. Computer software program produced by the authors at the University of Massachusetts, Amherst, MA. Available: http://www.umass.edu/landeco/research/fragstats/fragstats.html.

McGarigal, K., H. Y. Wan, K. A. Zeller, B. C. Timm, and S. A. Cushman. 2016. Multi-scale habitat selection modeling: a review and outlook. Landscape Ecology **31**:1161-1175.

Miettinen, J., C. H. Shi, and S. C. Liew. 2011. Deforestation rates in insular Southeast Asia between 2000 and 2010. Global Change Biology **17**:2261-2270.

Mittermeier, R. A., P. R. Gil, M. Hoffman, J. Pilgrim, T. Brooks, C. G. Mittermeier, J. Lamoreux, and G. A. B. da Fonseconda. 2004. Hotspots revisited: Earth's biologically richest and most endangered terrestrial ecoregions. Cemex, Mexico City.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. Nature **403**:853-858.

Nagelkerke, N. J. D. 1991. A Note on a General Definition of the Coefficient of Determination. Biometrika **78**:691-692.

Noss, R. F. 1990. Indicators for Monitoring Biodiversity - a Hierarchical Approach. Conservation Biology **4**:355-364.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. Henry, H. Stevens, E. Szoecs, and H.

Wagner. 2018. Vegan: community ecology package. R package version 2.4-6. http://CRAN.R-project.org/package=vegan.

Olson, D. M., E. Dinerstein, E. D. Wikramanayake, N. D. Burgess, G. V. N. Powell, E. C. Underwood, J. A. D'Amico, I. Itoua, H. E. Strand, J. C. Morrison, C. J. Loucks, T. F. Allnutt, T. H. Ricketts, Y. Kura, J. F. Lamoreux, W. W. Wettengel, P. Hedao, and K. R. Kassem. 2001. Terrestrial ecoregions of the worlds: A new map of life on Earth. Bioscience **51**:933-938.

Pearson, R. G., and T. P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography **12**:361-371.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rodrigues, A. S. L., and T. M. Brooks. 2007. Shortcuts for biodiversity conservation planning: The effectiveness of surrogates. Annual Review of Ecology Evolution and Systematics **38**:713-737.

Rondinini, C., M. Di Marco, F. Chiozza, G. Santulli, D. Baisero, P. Visconti, M. Hoffmann, J. Schipper, S. N. Stuart, M. F. Tognelli, G. Amori, A. Falcucci, L. Maiorano, and L. Boitani. 2011. Global habitat suitability models of terrestrial mammals. Philosophical Transactions of the Royal Society B-Biological Sciences **366**:2633-2641.

Ruter, S., C. C. Vos, M. van Eupen, and H. Ruhmkorf. 2014. Transboundary ecological networks as an adaptation strategy to climate change: The example of the Dutch - German border. Basic and Applied Ecology **15**:639-650.

Smith, G. 2018. Step away from stepwise. Journal of Big Data **5**.

Sodhi, N. S., and B. W. Brook. 2006. Southeast Asian biodiversity in crisis. Cambridge University Press, Cambridge.

Sodhi, N. S., L. P. Koh, B. W. Brook, and P. K. L. Ng. 2004. Southeast Asian biodiversity: an impending disaster. Trends in Ecology & Evolution **19**:654-660.

ter Braak, C. J. F. 1986. Canonical Correspondence-Analysis - a New Eigenvector Technique for Multivariate Direct Gradient Analysis. Ecology **67**:1167-1179.

UNEP-WCMC, and IUCN. 2022. Protected Planet: The World Database on Protected Areas (WDPA), Accessed April 2022, Cambridge, UK: UNEP-WCMC and IUCN. Available at: www.protectedplanet.net.

United Nations, Department of Economic and Social Affairs, Population Division. 2019. World Population Prospects 2019, custom data acquired via website.

Vincent, P. J., and J. M. Haworth. 1983. Poisson Regression-Models of Species Abundance. Journal of Biogeography **10**:153-160.

Vucetich, J. A., D. Burnham, E. A. Macdonald, J. T. Bruskotter, S. Marchini, A. Zimmermann, and D. W. Macdonald. 2018. Just conservation: What is it and should we pursue it? Biological Conservation **221**:23-33.

Wasserman, T. N., S. A. Cushman, D. O. Wallin, and J. Hayden. 2012. Multi Scale Habitat Relationships of Martes americana in Northern Idaho, USA. Environmental Sciences Faculty and Staff Publications. 20. https://cedar.wwu.edu/esci_facpubs/20.

WCS, and CIESIN. 2005. Last of The Wild data version 2, (LTW-2): global human footprint dataset (geographic). http://sedac.ciesin.columbia.edu/wildareas/.

Western, G., D. W. Macdonald, A. J. Loveridge, and A. J. Dickman. 2019. Creating Landscapes of Coexistence: Do Conservation Interventions Promote Tolerance of Lions in Human-dominated Landscapes? Conservation & Society **17**:204-217.

Wikramanayake, E. D. 2002. Terrestrial ecoregions of the Indo-Pacific : a conservation assessment. Island Press, Washington D.C.; London.

Willig, M. R., and S. J. Presley. 2016. Biodiversity and metacommunity structure of animals along altitudinal gradients in tropical montane forests. Journal of Tropical Ecology **32**:421-436.

Zhao, M., W. M. Cheng, C. H. Zhou, M. C. Li, K. Huang, and N. Wang. 2018. Assessing Spatiotemporal Characteristics of Urbanization Dynamics in Southeast Asia Using Time Series of DMSP/OLS Nighttime Light Data. Remote Sensing **10**.

# Supplementary material for Chapter 2

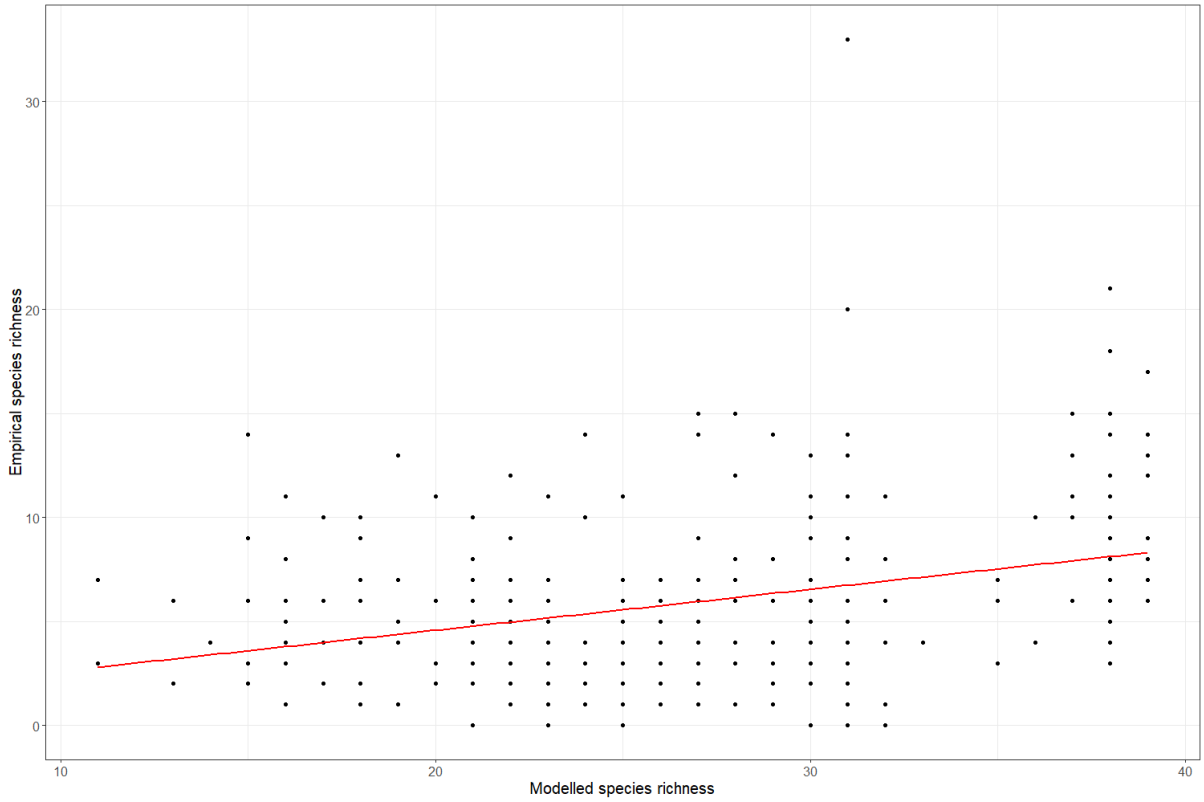



**Figure S1.** Scatter plot between the modelled and the empirical number of species at the camera trap stations used for validation. *Nagelkerke-pseudo-*$R^2$= 0.14, $p$< 0.0001.

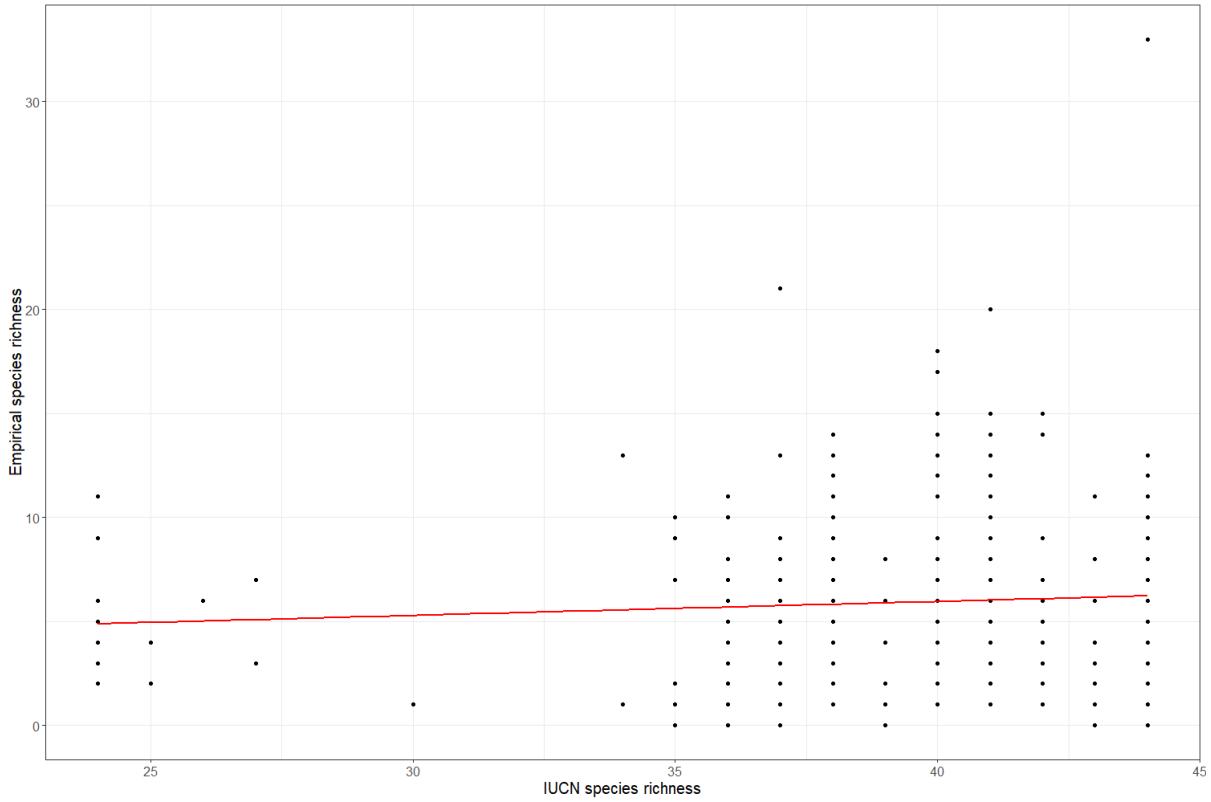



**Figure S2.** Scatter plot between the number of species predicted by the IUCN model and the empirical number of species at the camera trap stations used for validation. *Nagelkerke-pseudo-*$R^2$= 0.0017, *p*= 0.50.

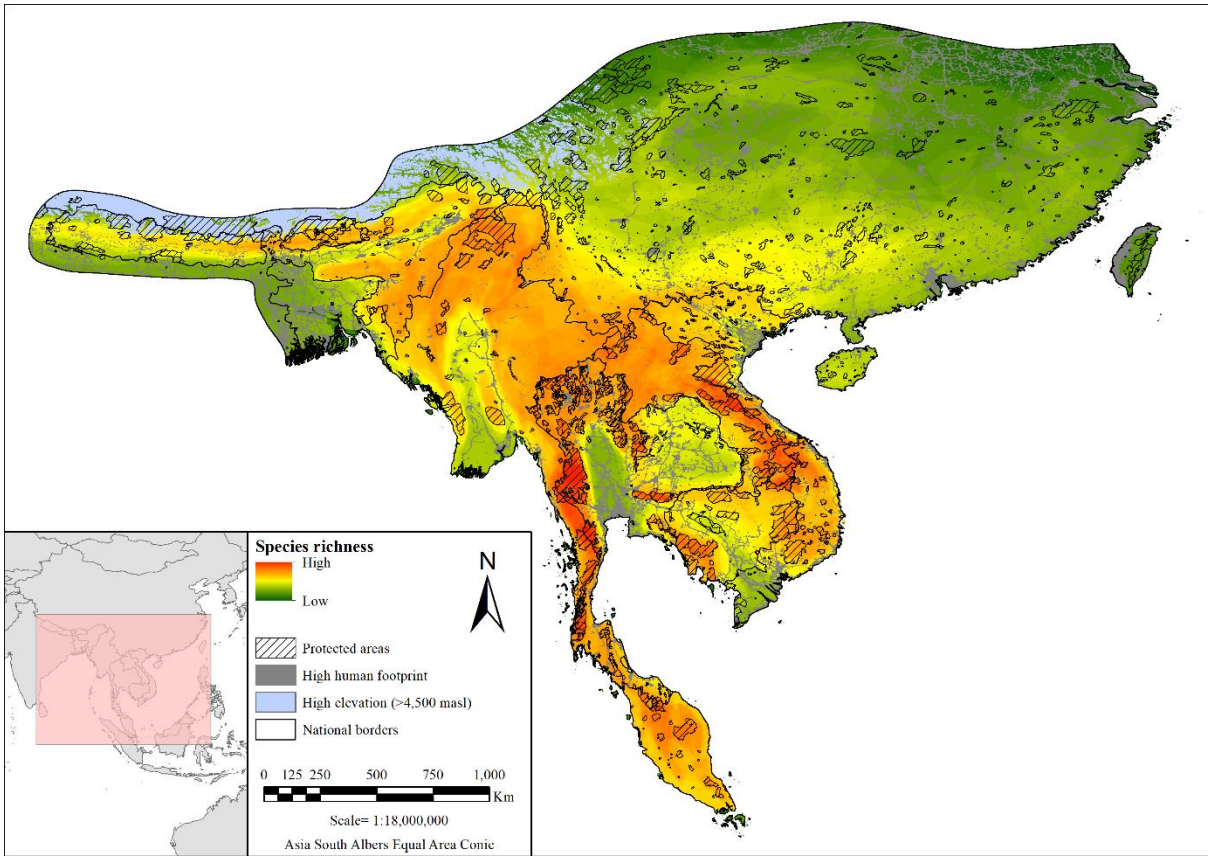



**Figure S3.** Predictive map of species richness in Southeast Asia as obtained by adding up the IUCN geographic range layers of the sampled species. Protected areas are overlaid in black hash marks.

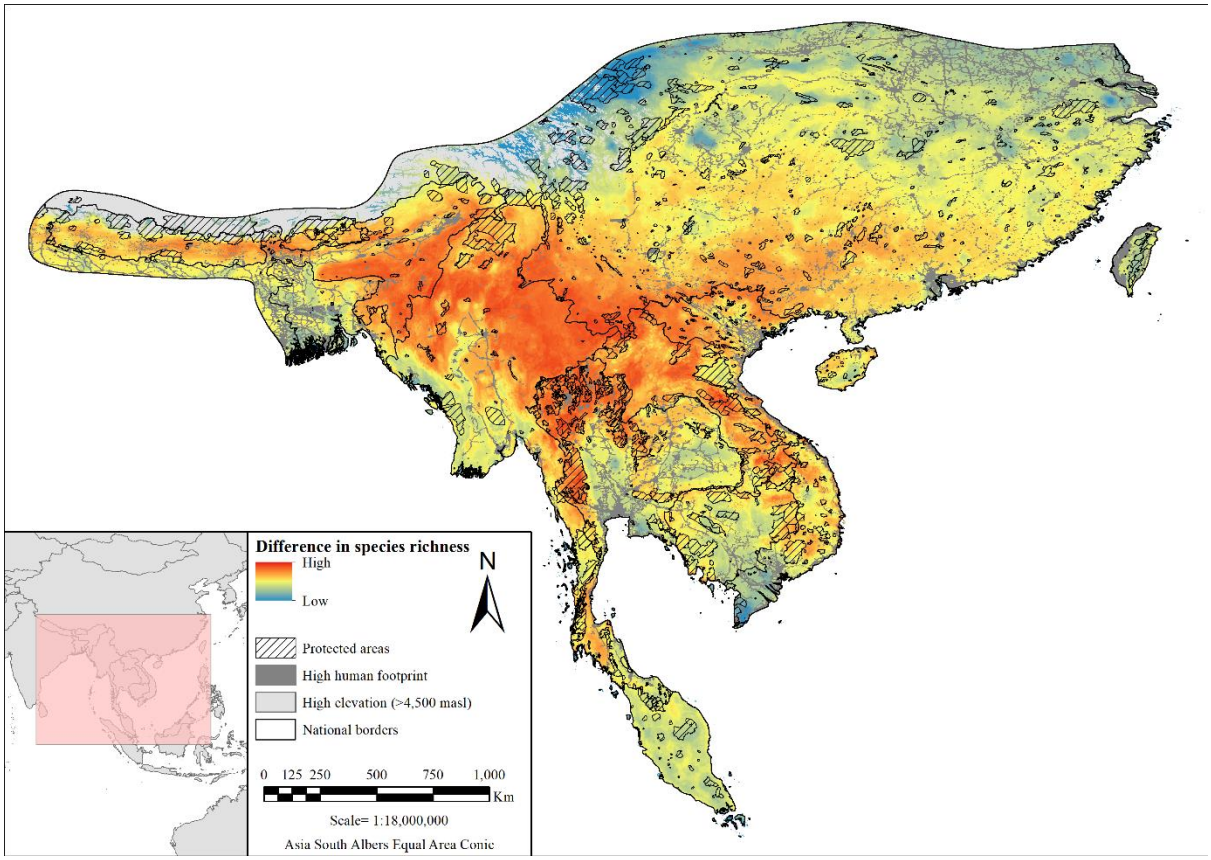



**Figure S4.** Difference between the predictive maps of the IUCN and of the modelled multi-species models. Warm colours represent areas where the IUCN model predicts more species than the modelled multi-species model, while cold colours represent areas where the modelled multi-species model predicts more species than the IUCN model. Protected areas are overlaid in black hash marks.

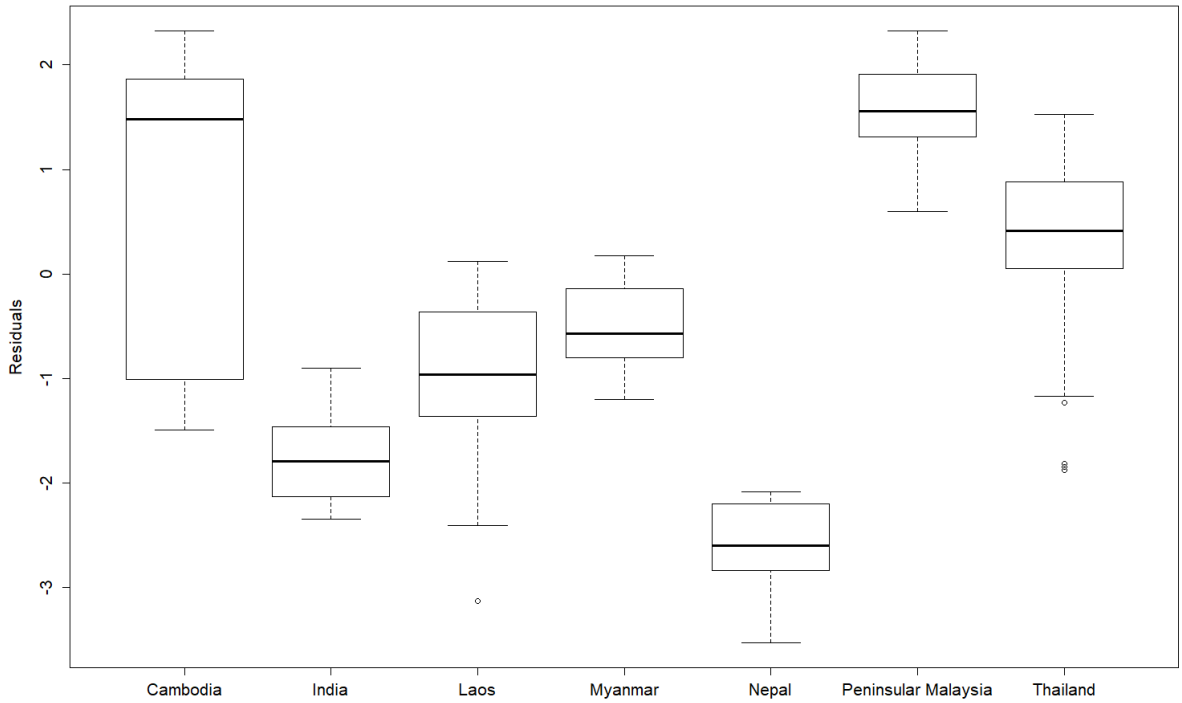



**Figure S5.** Boxplots of the residuals of the GLM between the modelled and empirical number of species at the camera trap stations used for validation, per country.

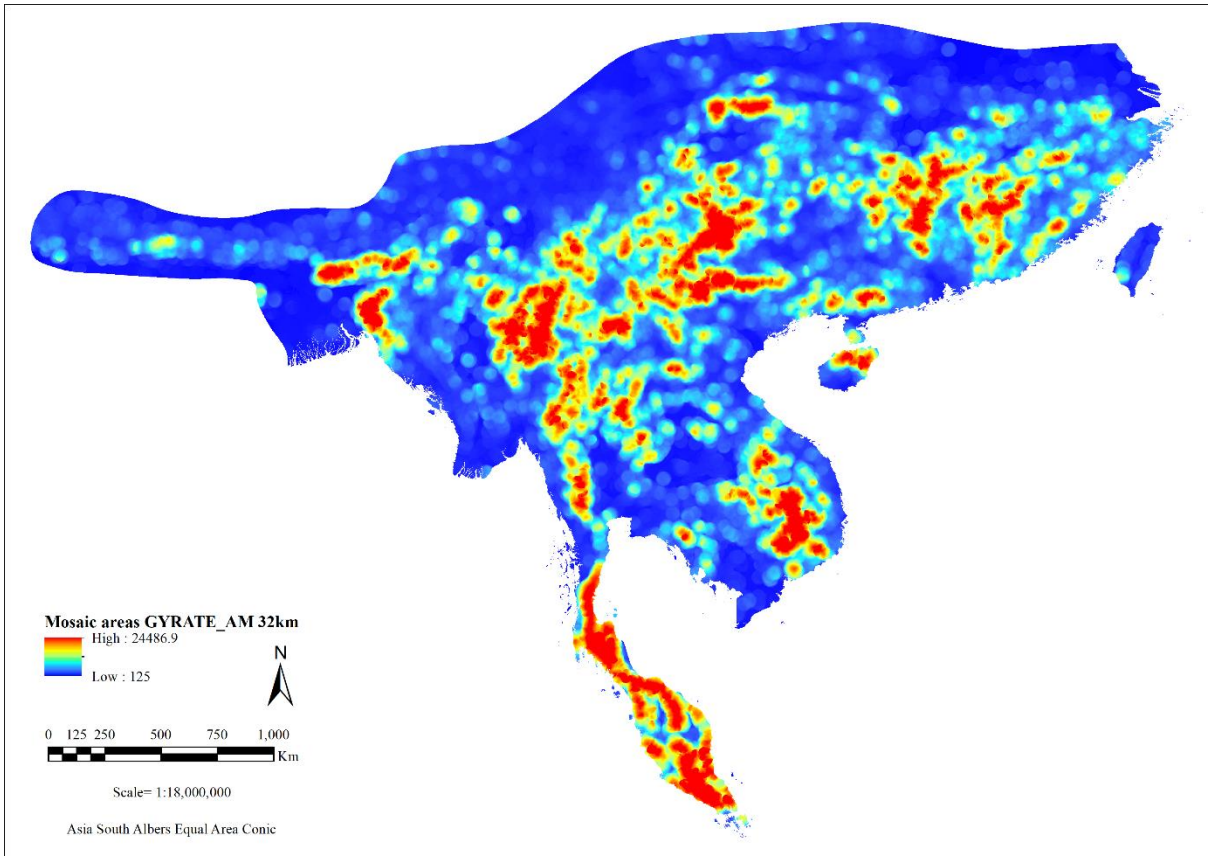



**Figure S6.** Southeast Asia mosaic areas GYRATE_AM 32,000m-radius layer.

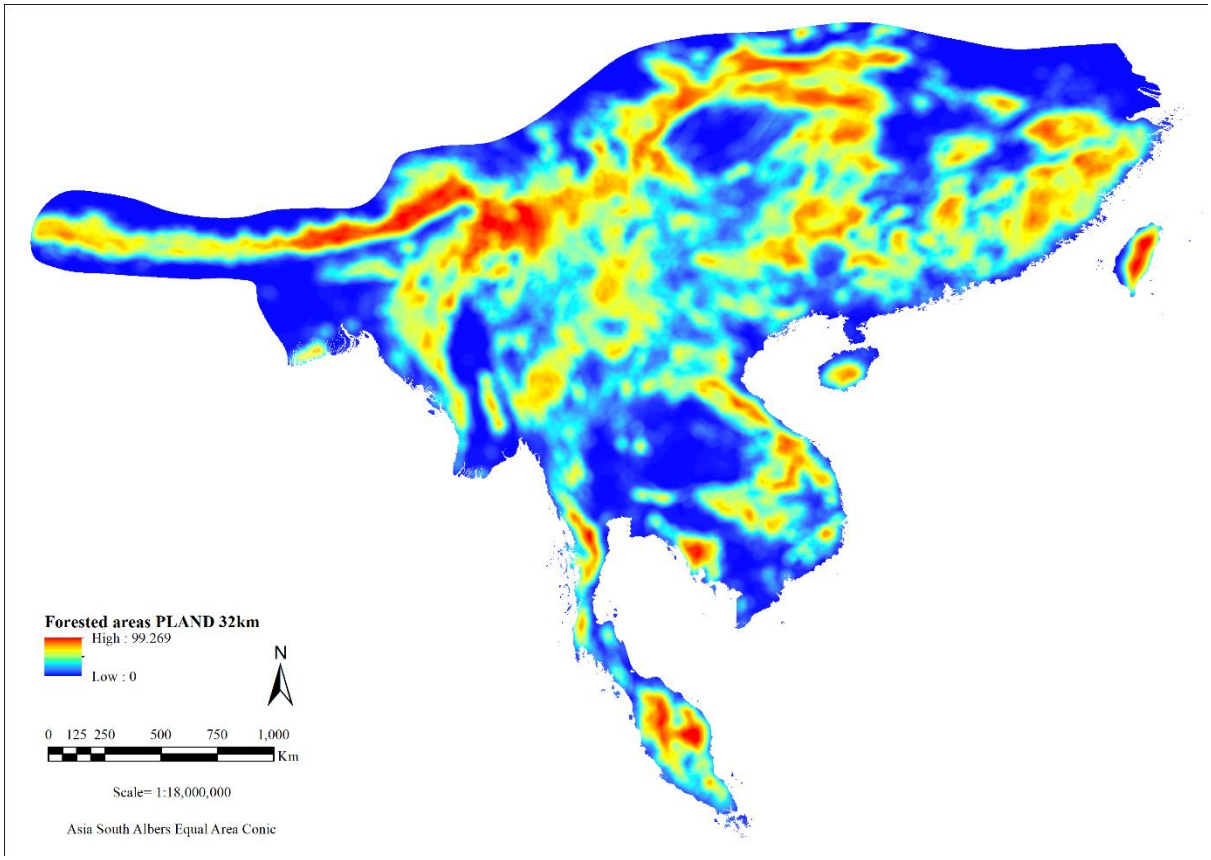



**Figure S7.** Southeast Asia forested areas PLAND 32,000m-radius layer.

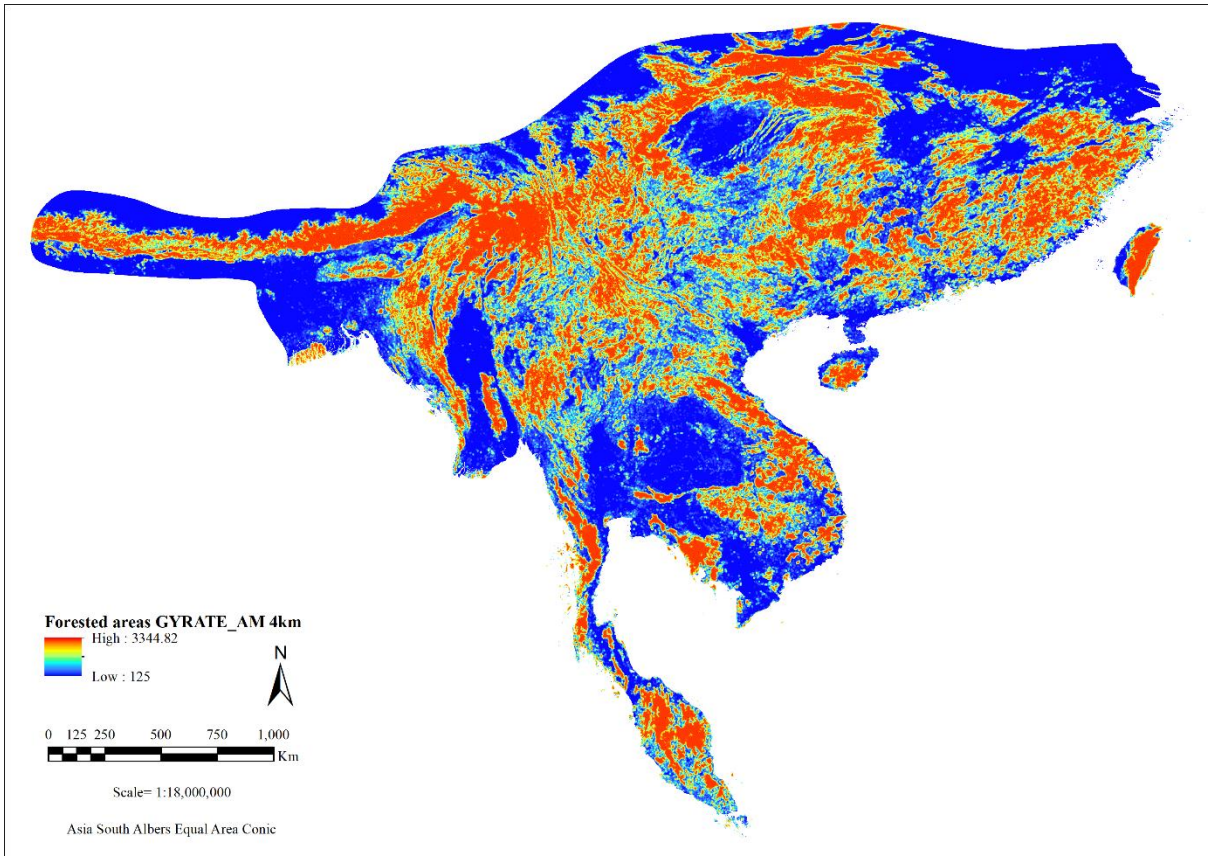



**Figure S8.** Southeast Asia forested areas GYRATE_AM 4,000m-radius layer.

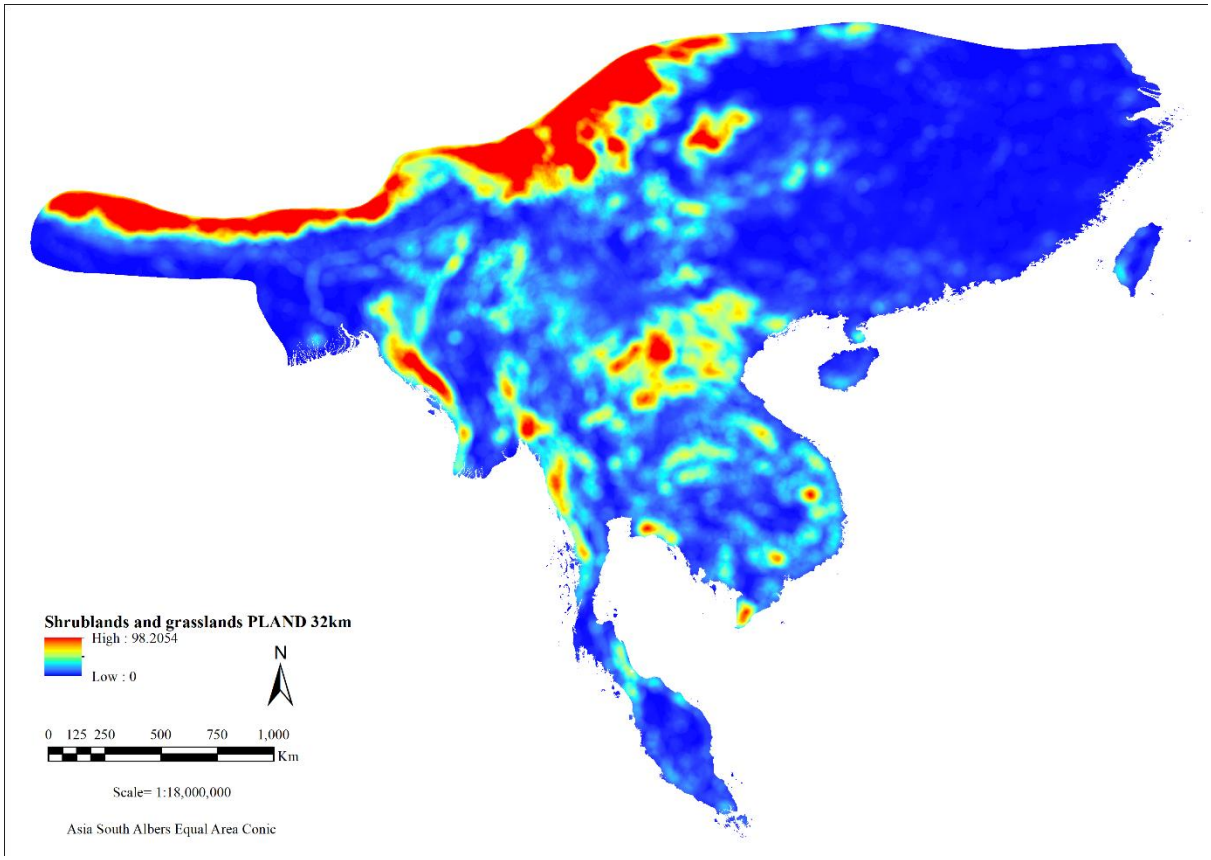



**Figure S9.** Southeast Asia shrublands and grasslands PLAND 32,000m-radius layer.

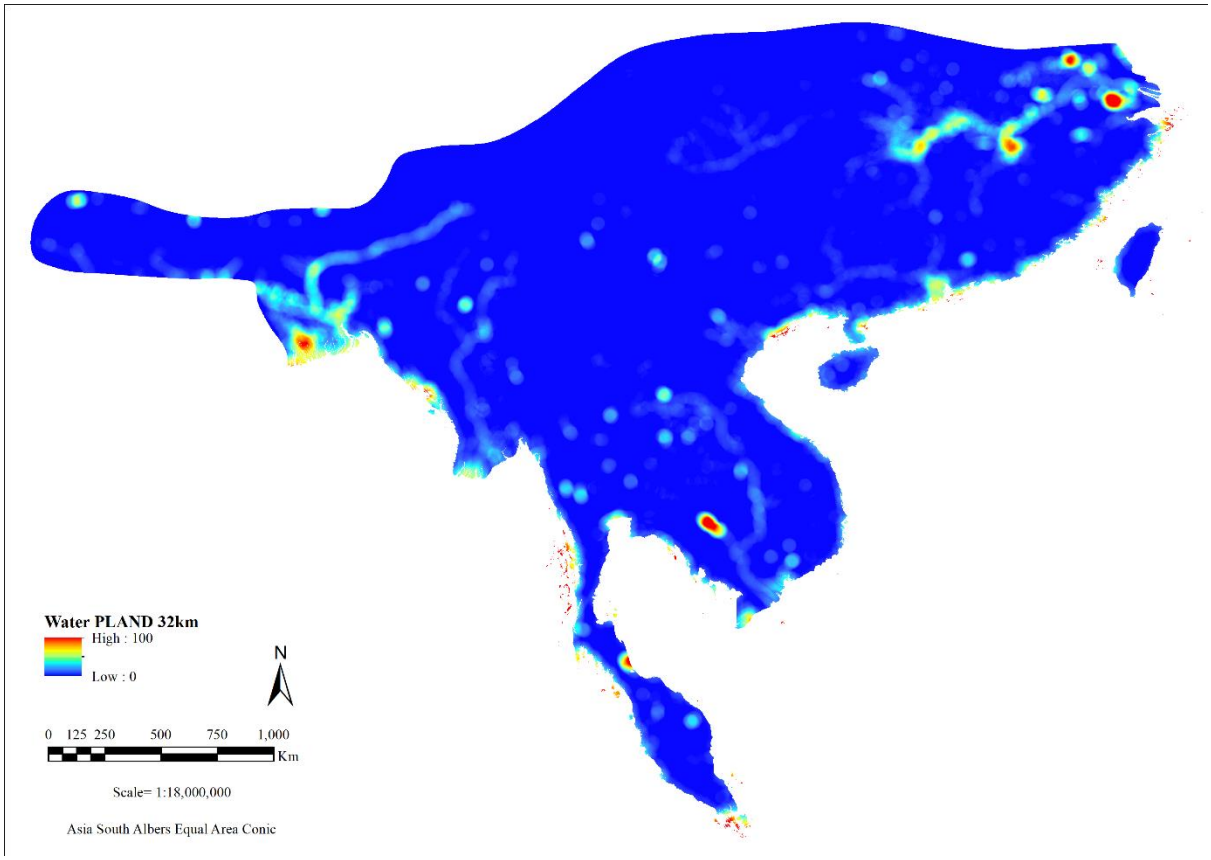



**Figure S10.** Southeast Asia water PLAND 32,000m-radius layer.

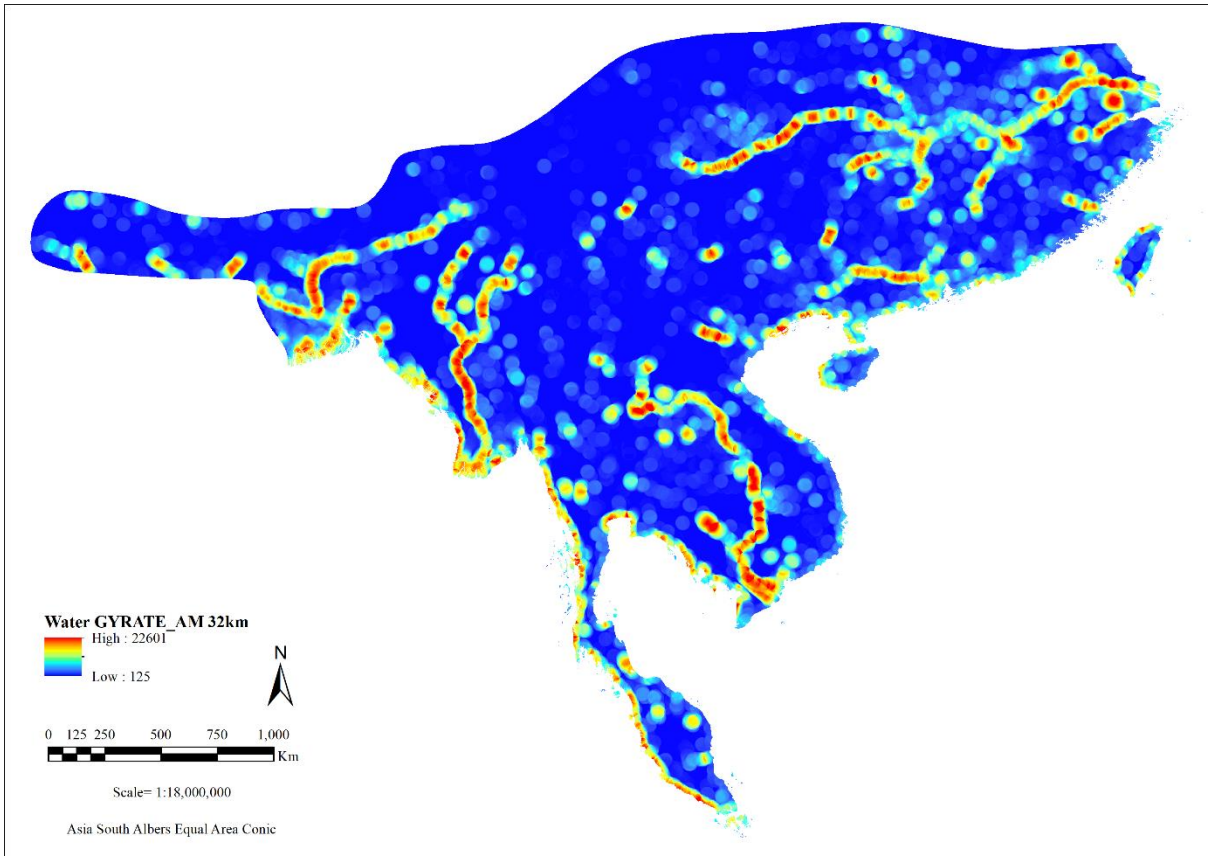



**Figure S11.** Southeast Asia water GYRATE_AM 32,000m-radius layer.

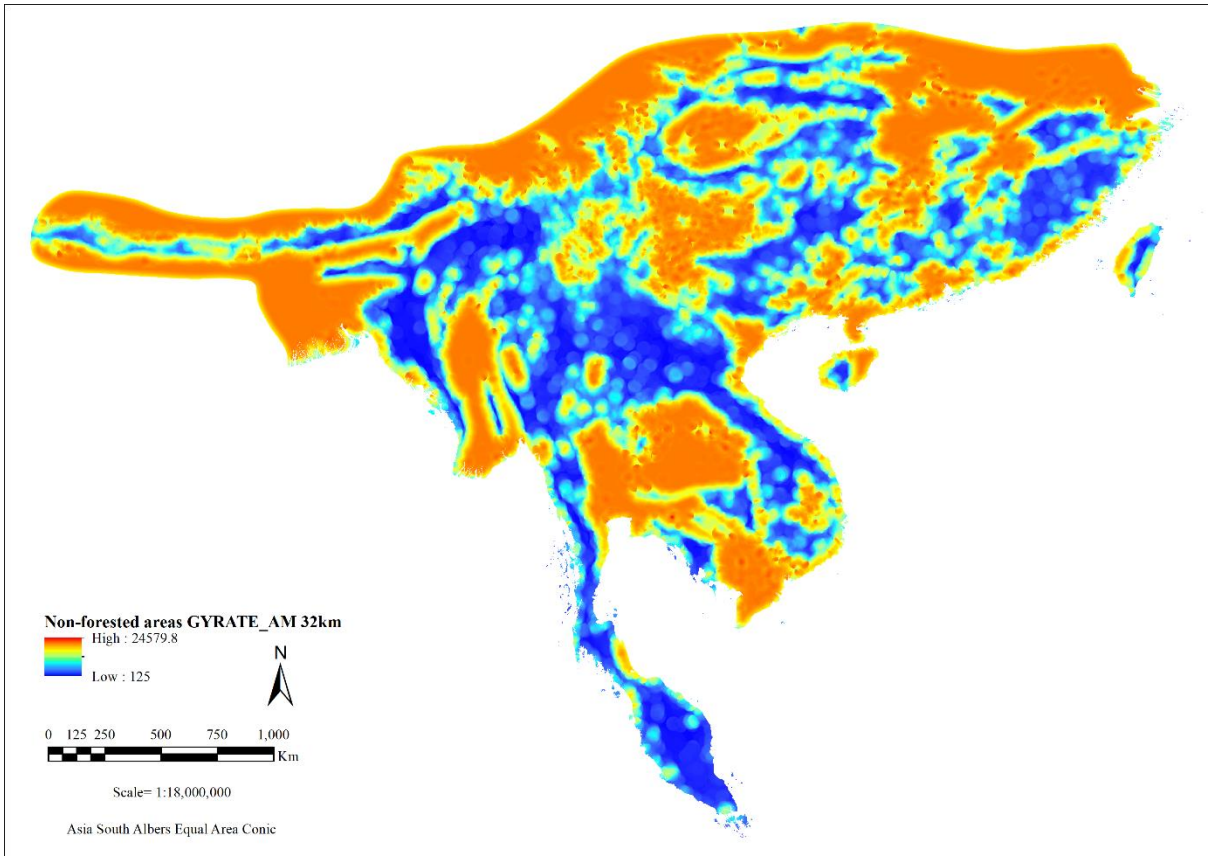



**Figure S12.** Southeast Asia non-forested areas GYRATE_AM 32,000m-radius layer.

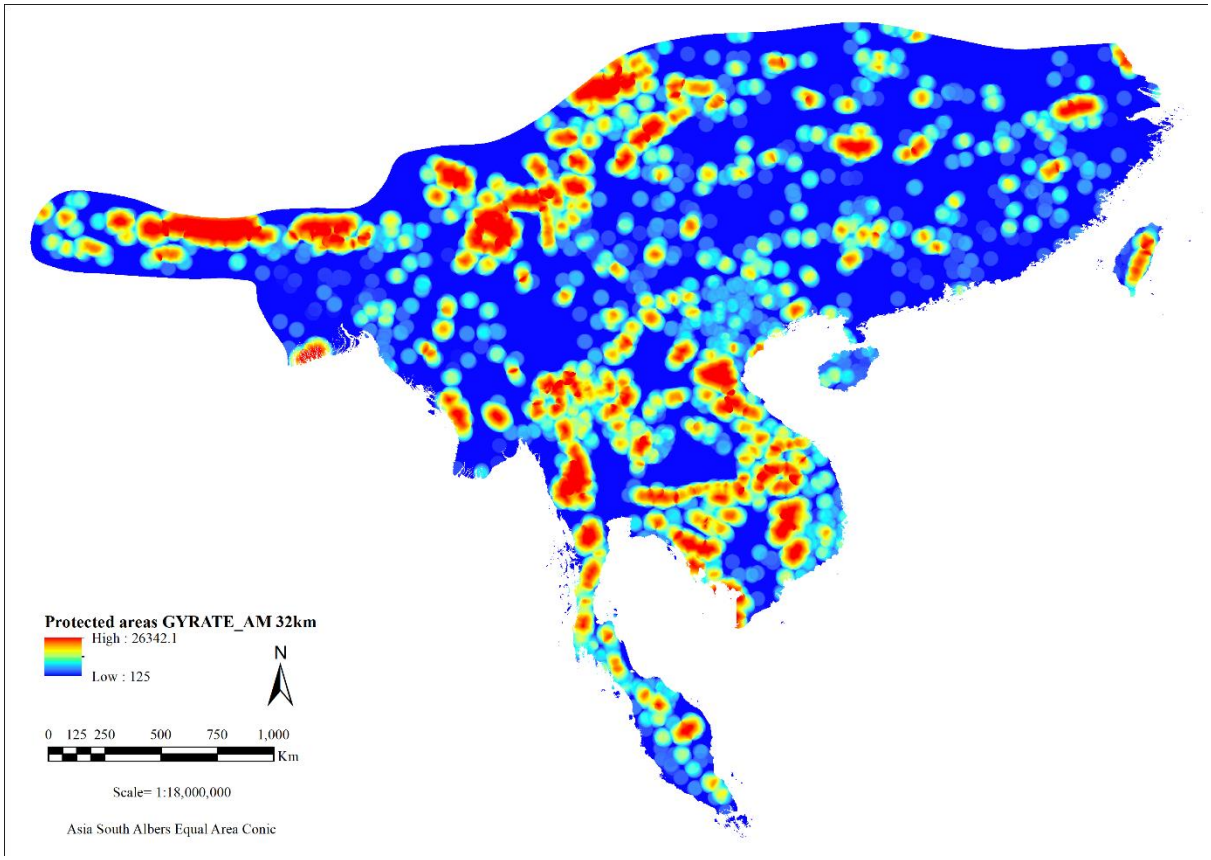



**Figure S13.** Southeast Asia protected areas GYRATE_AM 32,000m-radius layer.

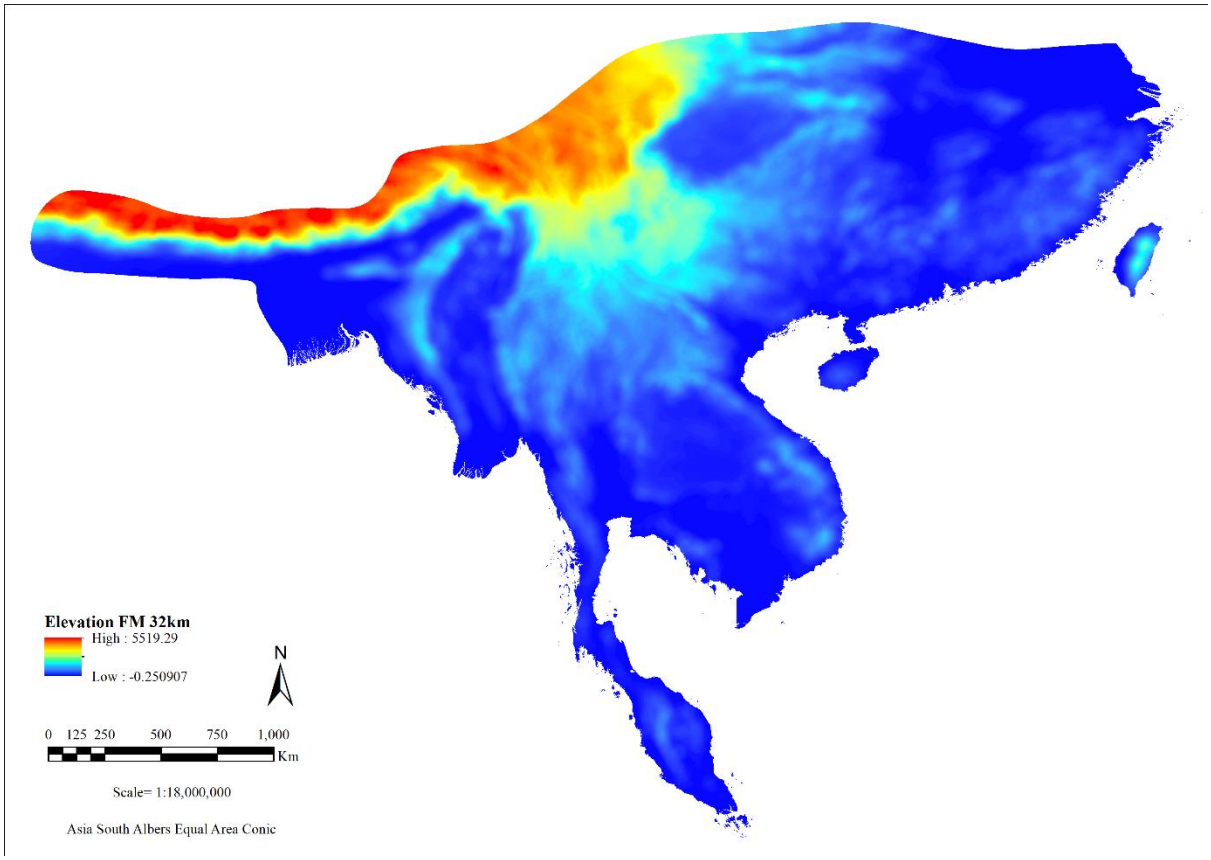



**Figure S14.** Southeast Asia elevation FM 32,000m-radius layer.

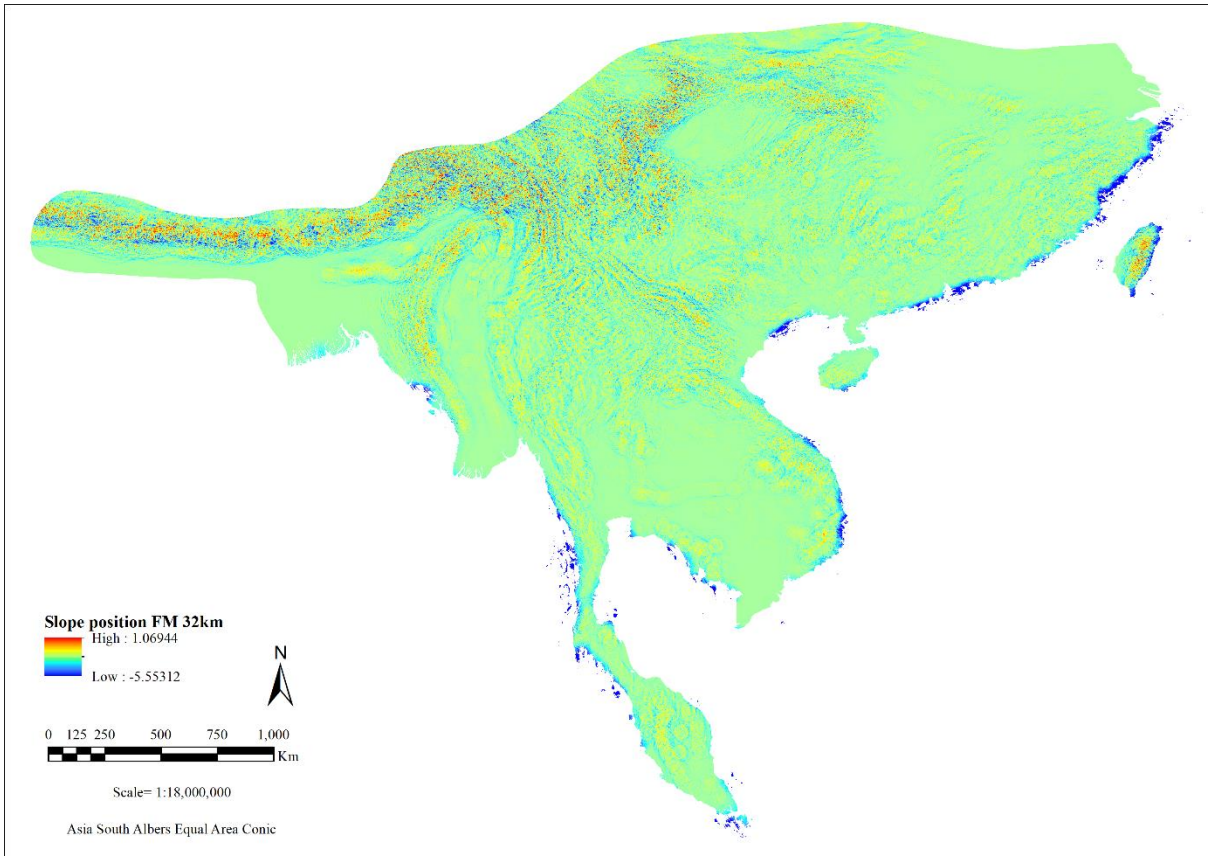



**Figure S15.** Southeast Asia slope position FM 32,000m-radius layer.

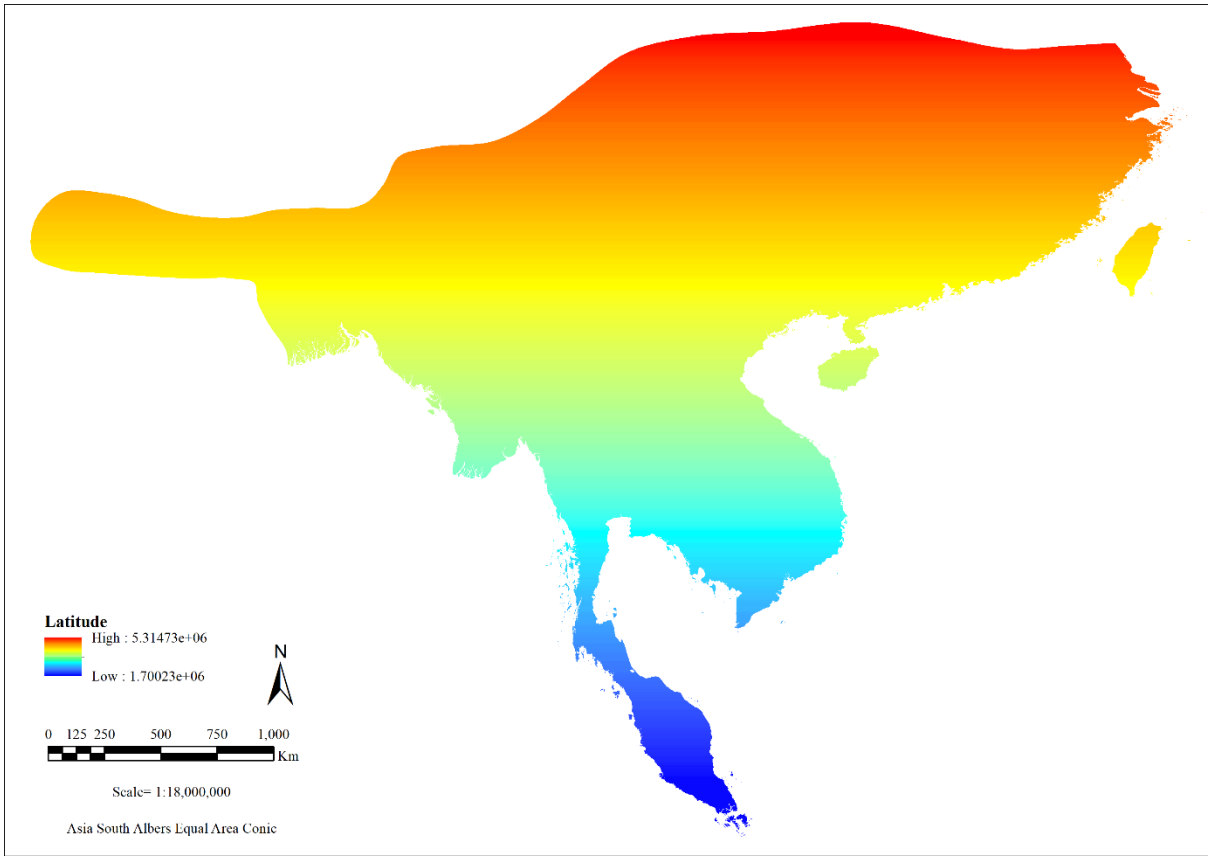



**Figure S16.** Southeast Asia Y (latitude) layer.

**Table S1.** Description of the primary set of covariates selected for the Southeast Asian biodiversity model. For the analyses, landscape and anthropic covariates were collapsed into the macro-group of environmental covariates. We used FRAGSTATS (McGarigal et al. 2012) to calculate additional metrics. For each landcover and forest cover binary layers, and for forest loss and protected areas layers, we calculated class level statistics: PLAND and GYRATE_AM. PLAND represents the percentage of the landscape occupied by a specific class. GYRATE_AM provides a measure of landscape continuity representing the average traversability from any random starting point within the class. For the overall landcover layer, we calculated landscape level statistics: contrast-weighted edge density (CWED), representing the habitat fragmentation weighted by the degree of contrast between different classes in the landscape. For the continuous layers, we used ArcMap v10.6.1 to calculate focal mean (FM) and standard deviation (SD). All raster layers were first re-projected to Asia South Albers Equal Area Conic projection in ArcMap v10.6.1, by applying a nearest neighbour re-sampling technique for discrete layers and a bilinear interpolation re-sampling technique for continuous layers. Then all raster layers were re-sampled to 250m resolution by applying the same re-sampling techniques.

| Class | Covariate | Source | Native resolution | Metrics |
|---|---|---|---|---|
| Landscape | Landcover | ESA Land Cover CCI (2017) | 300m | CWED |
| | Bare ground | ESA Land Cover CCI (2017) | 300m | PLAND, GYRATE_AM |

| | | | | |
|---|---|---|---|---|
| | Croplands | ESA Land Cover CCI (2017) | 300m | PLAND, GYRATE_AM |
| | Mosaic areas | ESA Land Cover CCI (2017) | 300m | PLAND, GYRATE_AM |
| | Forested areas | ESA Land Cover CCI (2017) | 300m | PLAND, GYRATE_AM |
| | Shrublands and grasslands | ESA Land Cover CCI (2017) | 300m | PLAND, GYRATE_AM |
| | Permanent snow and ice | ESA Land Cover CCI (2017) | 300m | PLAND, GYRATE_AM |
| | Urban areas | ESA Land Cover CCI (2017) | 300m | PLAND, GYRATE_AM |
| | Water | ESA Land Cover CCI (2017) | 300m | PLAND, GYRATE_AM |
| | Non-forested areas (0-20%) | Hansen et al. (2013) | 30m | PLAND, GYRATE_AM |
| | Open forests (20-40%) | Hansen et al. (2013) | 30m | PLAND, GYRATE_AM |
| | Closed forests (>40%) | Hansen et al. (2013) | 30m | PLAND, GYRATE_AM |
| Anthropic | Forest loss | Hansen et al. (2013) | 30m | FM, SD |
| | Human footprint | WCS and CIESIN (2005) | 1,000m | FM, SD |
| | Human population density | CIESIN and CIAT (2016) | 1,000m | FM, SD |

| | | | | |
|---|---|---|---|---|
| | Protected areas | UNEP-WCMC and IUCN (2022) | NA (polygons) | PLAND, GYRATE_AM |
| Topographic | Elevation | Jarvis et al. (2008) | 90m | FM, SD |
| | CTI | Jarvis et al. (2008); Evans et al. (2014) | 90m | FM, SD |
| | Slope position | Jarvis et al. (2008); Evans et al. (2014) | 90m | FM, SD |
| Climatic | Mean annual precipitation | Fick and Hijmans (2017) | 1,000m | FM |
| Spatial | X | NA | NA | NA |
| | Y | NA | NA | NA |
| | X*Y | NA | NA | NA |
| | $X^2$ | NA | NA | NA |
| | $Y^2$ | NA | NA | NA |
| | $X^2$*Y | NA | NA | NA |
| | X*$Y^2$ | NA | NA | NA |
| | $X^2$*$Y^2$ | NA | NA | NA |
| Camera effort | Camera trap nights | NA | NA | NA |

**Table S2.** Species sampled in Southeast Asia with relative identification codes.

| Species | Class | Order | Family | Genus | Code |
|---|---|---|---|---|---|
| *Elephas maximus* | Mammalia | Proboscidea | Elephantidae | *Elephas* | 1 |
| *Echinosorex gymnura* | Mammalia | Eulipotyphla | Erinaceidae | *Echinosorex* | 2 |
| *Sus scrofa* | Mammalia | Cetartiodactyla | Suidae | *Sus* | 3 |
| Mouse-deer species | Mammalia | Cetartiodactyla | Tragulidae | NA | 4 |
| *Tragulus kanchil* | Mammalia | Cetartiodactyla | Tragulidae | *Tragulus* | 5 |
| Muntjac species | Mammalia | Cetartiodactyla | Cervidae | *Muntiacus* | 6 |
| *Muntiacus muntjak* | Mammalia | Cetartiodactyla | Cervidae | *Muntiacus* | 7 |
| *Muntiacus rooseveltorum* | Mammalia | Cetartiodactyla | Cervidae | *Muntiacus* | 8 |
| *Muntiacus vaginalis* | Mammalia | Cetartiodactyla | Cervidae | *Muntiacus* | 9 |
| *Rusa unicolor* | Mammalia | Cetartiodactyla | Cervidae | *Rusa* | 10 |
| *Bos javanicus* | Mammalia | Cetartiodactyla | Bovidae | *Bos* | 11 |
| *Bos gaurus* | Mammalia | Cetartiodactyla | Bovidae | *Bos* | 12 |
| *Naemorhedus goral* | Mammalia | Cetartiodactyla | Bovidae | *Naemorhedus* | 13 |

| | | | | | |
|---|---|---|---|---|---|
| *Capricornis thar* | Mammalia | Cetartiodactyla | Bovidae | *Capricornis* | 14 |
| *Capricornis milneedwardsii* | Mammalia | Cetartiodactyla | Bovidae | *Capricornis* | 15 |
| *Moschus leucogaster* | Mammalia | Cetartiodactyla | Moschidae | *Moschus* | 16 |
| *Tapirus indicus* | Mammalia | Perissodactyla | Tapiridae | *Tapirus* | 17 |
| *Manis javanica* | Mammalia | Pholidota | Manidae | *Manis* | 18 |
| *Arctictis binturong* | Mammalia | Carnivora | Viverridae | *Arctictis* | 19 |
| *Paguma larvata* | Mammalia | Carnivora | Viverridae | *Paguma* | 20 |
| *Paradoxurus hermaphroditus* | Mammalia | Carnivora | Viverridae | *Paradoxurus* | 21 |
| *Chrotogale owstoni* | Mammalia | Carnivora | Viverridae | *Chrotogale* | 22 |
| *Hemigalus derbyanus* | Mammalia | Carnivora | Viverridae | *Hemigalus* | 23 |
| *Viverricula indica* | Mammalia | Carnivora | Viverridae | *Viverricula* | 24 |
| *Viverra zibetha* | Mammalia | Carnivora | Viverridae | *Viverra* | 25 |
| *Viverra megaspila* | Mammalia | Carnivora | Viverridae | *Viverra* | 26 |
| *Herpestes urva* | Mammalia | Carnivora | Herpestidae | *Herpestes* | 27 |
| *Herpestes javanicus* | Mammalia | Carnivora | Herpestidae | *Herpestes* | 28 |
| *Prionodon linsang* | Mammalia | Carnivora | Prionodontidae | *Prionodon* | 29 |

| | | | | | |
|---|---|---|---|---|---|
| *Prionodon pardicolor* | Mammalia | Carnivora | Prionodontidae | *Prionodon* | 30 |
| *Neofelis nebulosa* | Mammalia | Carnivora | Felidae | *Neofelis* | 31 |
| *Panthera tigris* | Mammalia | Carnivora | Felidae | *Panthera* | 32 |
| *Panthera pardus* | Mammalia | Carnivora | Felidae | *Panthera* | 33 |
| *Pardofelis marmorata* | Mammalia | Carnivora | Felidae | *Pardofelis* | 34 |
| *Catopuma temminckii* | Mammalia | Carnivora | Felidae | *Catopuma* | 35 |
| *Prionailurus bengalensis* | Mammalia | Carnivora | Felidae | *Prionailurus* | 36 |
| *Vulpes vulpes* | Mammalia | Carnivora | Canidae | *Vulpes* | 37 |
| *Cuon alpinus* | Mammalia | Carnivora | Canidae | *Cuon* | 38 |
| *Canis aureus* | Mammalia | Carnivora | Canidae | *Canis* | 39 |
| Bear species | Mammalia | Carnivora | Ursidae | NA | 40 |
| *Helarctos malayanus* | Mammalia | Carnivora | Ursidae | *Helarctos* | 41 |
| *Ursus thibetanus* | Mammalia | Carnivora | Ursidae | *Ursus* | 42 |
| *Ailurus fulgens* | Mammalia | Carnivora | Ailuridae | *Ailurus* | 43 |
| *Arctonyx collaris* | Mammalia | Carnivora | Mustelidae | *Arctonyx* | 44 |
| Marten species | Mammalia | Carnivora | Mustelidae | *Martes* | 45 |

| | | | | | |
|---|---|---|---|---|---|
| *Martes flavigula* | Mammalia | Carnivora | Mustelidae | *Martes* | 46 |
| Ferret-badger species | Mammalia | Carnivora | Mustelidae | *Melogale* | 47 |
| *Melogale personata* | Mammalia | Carnivora | Mustelidae | *Melogale* | 48 |
| Weasel species | Mammalia | Carnivora | Mustelidae | *Mustela* | 49 |
| *Mustela kathiah* | Mammalia | Carnivora | Mustelidae | *Mustela* | 50 |
| *Aonyx cinereus* | Mammalia | Carnivora | Mustelidae | *Aonyx* | 51 |
| *Lutrogale perspicillata* | Mammalia | Carnivora | Mustelidae | *Lutrogale* | 52 |
| *Tupaia glis* | Mammalia | Scandentia | Tupaiidae | *Tupaia* | 53 |
| *Tupaia belangeri* | Mammalia | Scandentia | Tupaiidae | *Tupaia* | 54 |
| *Lepus peguensis* | Mammalia | Lagomorpha | Leporidae | *Lepus* | 55 |
| Rat species | Mammalia | Rodentia | NA | NA | 56 |
| *Atherurus macrourus* | Mammalia | Rodentia | Hystricidae | *Atherurus* | 57 |
| *Hystrix brachyura* | Mammalia | Rodentia | Hystricidae | *Hystrix* | 58 |
| Squirrel species | Mammalia | Rodentia | Sciuridae | NA | 59 |
| *Callosciurus erythraeus* | Mammalia | Rodentia | Sciuridae | *Callosciurus* | 60 |
| *Rhinosciurus laticaudatus* | Mammalia | Rodentia | Sciuridae | *Rhinosciurus* | 61 |

| | | | | | |
|---|---|---|---|---|---|
| *Tamiops rodolphii* | Mammalia | Rodentia | Sciuridae | *Tamiops* | 62 |
| *Dremomys rufigenis* | Mammalia | Rodentia | Sciuridae | *Dremomys* | 63 |
| *Rhizomys pruinosus* | Mammalia | Rodentia | Spalacidae | *Rhizomys* | 64 |
| *Rattus exulans* | Mammalia | Rodentia | Muridae | *Rattus* | 65 |
| *Leopoldamys sabanus* | Mammalia | Rodentia | Muridae | *Leopoldamys* | 66 |
| *Semnopithecus schistaceus* | Mammalia | Primates | Cercopithecidae | *Semnopithecus* | 67 |
| *Trachypithecus shortridgei* | Mammalia | Primates | Cercopithecidae | *Trachypithecus* | 68 |
| *Trachypithecus obscurus* | Mammalia | Primates | Cercopithecidae | *Trachypithecus* | 69 |
| *Macaca nemestrina* | Mammalia | Primates | Cercopithecidae | *Macaca* | 70 |
| *Macaca leonina* | Mammalia | Primates | Cercopithecidae | *Macaca* | 71 |
| *Macaca assamensis* | Mammalia | Primates | Cercopithecidae | *Macaca* | 72 |
| *Macaca arctoides* | Mammalia | Primates | Cercopithecidae | *Macaca* | 73 |
| *Macaca mulatta* | Mammalia | Primates | Cercopithecidae | *Macaca* | 74 |
| Pheasant species | Aves | Galliformes | Phasianidae | NA | 75 |
| Partridge species | Aves | Galliformes | Phasianidae | NA | 76 |
| *Arborophila torqueola* | Aves | Galliformes | Phasianidae | *Arborophila* | 77 |

| | | | | | |
|---|---|---|---|---|---|
| *Arborophila cambodiana* | Aves | Galliformes | Phasianidae | *Arborophila* | 78 |
| *Arborophila atrogularis* | Aves | Galliformes | Phasianidae | *Arborophila* | 79 |
| *Gallus gallus* | Aves | Galliformes | Phasianidae | *Gallus* | 80 |
| *Argusianus argus* | Aves | Galliformes | Phasianidae | *Argusianus* | 81 |
| *Polyplectron bicalcaratum* | Aves | Galliformes | Phasianidae | *Polyplectron* | 82 |
| *Ithaginis cruentus* | Aves | Galliformes | Phasianidae | *Ithaginis* | 83 |
| *Tragopan satyra* | Aves | Galliformes | Phasianidae | *Tragopan* | 84 |
| *Lophophorus impejanus* | Aves | Galliformes | Phasianidae | *Lophophorus* | 85 |
| *Lophura diardi* | Aves | Galliformes | Phasianidae | *Lophura* | 86 |
| *Lophura leucomelanos* | Aves | Galliformes | Phasianidae | *Lophura* | 87 |
| *Lophura nycthemera* | Aves | Galliformes | Phasianidae | *Lophura* | 88 |
| *Lophura rufa* | Aves | Galliformes | Phasianidae | *Lophura* | 89 |
| *Varanus salvator* | Reptilia | Squamata | Varanidae | *Varanus* | 90 |
| Forest guard | NA | NA | NA | NA | 91 |
| Human | NA | NA | NA | NA | 92 |
| Domestic dog | NA | NA | NA | NA | 93 |

| | | | | | |
|---|---|---|---|---|---|
| Domestic cattle | NA | NA | NA | NA | 94 |

**Table S3.** Species sampled in Southeast Asia with number of detections per country. We ensured the independence of data by discounting records of the same species at the same camera trap station within 1 hour, except when animals were individually recognizable and when genders and/or age classes were unambiguous.

| Species | Cambodia | India | Laos | Myanmar | Nepal | Peninsular Malaysia | Thailand | Total |
|---|---|---|---|---|---|---|---|---|
| *Ailurus fulgens* | 0 | 0 | 0 | 0 | 14 | 0 | 0 | 14 |
| *Aonyx cinereus* | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 8 |
| *Arborophila atrogularis* | 0 | 14 | 0 | 0 | 0 | 0 | 0 | 14 |
| *Arborophila cambodiana* | 42 | 0 | 0 | 0 | 0 | 0 | 0 | 42 |
| *Arborophila torqueola* | 0 | 0 | 0 | 0 | 12 | 0 | 0 | 12 |
| *Arctictis binturong* | 4 | 0 | 6 | 7 | 0 | 31 | 15 | 63 |
| *Arctonyx collaris* | 3 | 9 | 118 | 0 | 0 | 0 | 228 | 358 |
| *Argusianus argus* | 0 | 0 | 0 | 0 | 0 | 697 | 0 | 697 |
| *Atherurus macrourus* | 82 | 173 | 75 | 207 | 0 | 238 | 144 | 919 |
| Bear species | 0 | 4 | 2 | 0 | 0 | 0 | 0 | 6 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Bos gaurus* | 27 | 3 | 1 | 79 | 0 | 0 | 1180 | 1290 |
| *Bos javanicus* | 100 | 0 | 0 | 0 | 0 | 0 | 11 | 111 |
| *Callosciurus erythraeus* | 0 | 5 | 15 | 0 | 0 | 0 | 0 | 20 |
| *Canis aureus* | 0 | 1 | 0 | 0 | 0 | 0 | 41 | 42 |
| *Capricornis milneedwardsii* | 10 | 0 | 103 | 2 | 0 | 0 | 90 | 205 |
| *Capricornis thar* | 0 | 35 | 0 | 0 | 47 | 0 | 0 | 82 |
| *Catopuma temminckii* | 18 | 30 | 71 | 166 | 0 | 2 | 19 | 306 |
| *Chrotogale owstoni* | 0 | 0 | 16 | 0 | 0 | 0 | 0 | 16 |
| *Cuon alpinus* | 37 | 147 | 64 | 67 | 0 | 0 | 389 | 704 |
| Domestic cattle | 0 | 0 | 0 | 15 | 173 | 0 | 39 | 227 |
| Domestic dog | 9 | 4 | 0 | 13 | 44 | 0 | 133 | 203 |
| *Dremomys rufigenis* | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 5 |
| *Echinosorex gymnura* | 0 | 0 | 0 | 0 | 0 | 74 | 0 | 74 |
| *Elephas maximus* | 59 | 16 | 0 | 1 | 0 | 98 | 1469 | 1643 |
| Ferret-badger species | 0 | 5 | 1 | 0 | 0 | 0 | 0 | 6 |
| Forest guard | 0 | 656 | 0 | 632 | 0 | 20 | 4463 | 5771 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Gallus gallus* | 59 | 34 | 10 | 13 | 0 | 14 | 1509 | 1639 |
| *Helarctos malayanus* | 62 | 16 | 61 | 34 | 0 | 144 | 402 | 719 |
| *Hemigalus derbyanus* | 0 | 0 | 0 | 0 | 0 | 364 | 0 | 364 |
| *Herpestes javanicus* | 0 | 0 | 0 | 0 | 0 | 0 | 33 | 33 |
| *Herpestes urva* | 28 | 100 | 43 | 9 | 0 | 39 | 211 | 430 |
| Human | 224 | 391 | 32 | 297 | 1114 | 14 | 5991 | 8063 |
| *Hystrix brachyura* | 99 | 89 | 202 | 259 | 100 | 604 | 971 | 2324 |
| *Ithaginis cruentus* | 0 | 0 | 0 | 0 | 7 | 0 | 0 | 7 |
| *Leopoldamys sabanus* | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 5 |
| *Lepus peguensis* | 1 | 0 | 0 | 0 | 0 | 0 | 21 | 22 |
| *Lophophorus impejanus* | 0 | 0 | 0 | 0 | 29 | 0 | 0 | 29 |
| *Lophura diardi* | 48 | 0 | 0 | 0 | 0 | 0 | 1723 | 1771 |
| *Lophura leucomelanos* | 0 | 60 | 0 | 13 | 33 | 0 | 0 | 106 |
| *Lophura nycthemera* | 190 | 0 | 81 | 0 | 0 | 0 | 37 | 308 |
| *Lophura rufa* | 0 | 0 | 0 | 0 | 0 | 41 | 0 | 41 |
| *Lutrogale perspicillata* | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 5 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Macaca arctoides* | 0 | 0 | 244 | 78 | 0 | 0 | 0 | 322 |
| *Macaca assamensis* | 0 | 2 | 19 | 0 | 26 | 0 | 0 | 47 |
| *Macaca leonina* | 128 | 10 | 1 | 23 | 0 | 0 | 867 | 1029 |
| *Macaca mulatta* | 0 | 0 | 17 | 2 | 0 | 0 | 0 | 19 |
| *Macaca nemestrina* | 0 | 0 | 0 | 0 | 0 | 1418 | 0 | 1418 |
| *Manis javanica* | 0 | 0 | 0 | 0 | 0 | 34 | 14 | 48 |
| Marten species | 0 | 0 | 0 | 0 | 22 | 0 | 0 | 22 |
| *Martes flavigula* | 78 | 32 | 192 | 30 | 316 | 59 | 185 | 892 |
| *Melogale personata* | 7 | 0 | 1 | 0 | 0 | 0 | 0 | 8 |
| *Moschus leucogaster* | 0 | 0 | 0 | 0 | 49 | 0 | 0 | 49 |
| Mouse-deer species | 0 | 0 | 0 | 0 | 0 | 158 | 0 | 158 |
| *Muntiacus muntjak* | 0 | 0 | 0 | 0 | 0 | 3559 | 0 | 3559 |
| *Muntiacus rooseveltorum* | 0 | 0 | 106 | 0 | 0 | 0 | 0 | 106 |
| *Muntiacus vaginalis* | 814 | 169 | 459 | 1054 | 1061 | 0 | 1581 | 5138 |
| Muntjac species | 0 | 0 | 25 | 0 | 0 | 0 | 0 | 25 |
| *Mustela kathiah* | 3 | 0 | 0 | 0 | 2 | 0 | 0 | 5 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Naemorhedus goral* | 0 | 0 | 0 | 0 | 40 | 0 | 0 | 40 |
| *Neofelis nebulosa* | 12 | 81 | 61 | 103 | 17 | 30 | 125 | 429 |
| *Paguma larvata* | 3 | 2 | 73 | 0 | 45 | 34 | 0 | 157 |
| *Panthera pardus* | 11 | 0 | 0 | 1 | 44 | 7 | 0 | 63 |
| *Panthera tigris* | 0 | 0 | 20 | 30 | 0 | 0 | 781 | 831 |
| *Paradoxurus hermaphroditus* | 278 | 110 | 89 | 86 | 0 | 79 | 444 | 1086 |
| *Pardofelis marmorata* | 21 | 36 | 46 | 61 | 0 | 21 | 18 | 203 |
| Partridge species | 0 | 0 | 0 | 0 | 7 | 0 | 0 | 7 |
| Pheasant species | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 5 |
| *Polyplectron bicalcaratum* | 0 | 88 | 11 | 39 | 0 | 0 | 0 | 138 |
| *Prionailurus bengalensis* | 106 | 78 | 39 | 159 | 126 | 153 | 533 | 1194 |
| *Prionodon linsang* | 0 | 0 | 0 | 0 | 0 | 19 | 0 | 19 |
| *Prionodon pardicolor* | 3 | 0 | 6 | 1 | 0 | 0 | 8 | 18 |
| Rat species | 365 | 130 | 0 | 3 | 37 | 320 | 360 | 1215 |
| *Rattus exulans* | 0 | 0 | 15 | 0 | 0 | 0 | 0 | 15 |
| *Rhinosciurus laticaudatus* | 0 | 0 | 0 | 0 | 0 | 12 | 0 | 12 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Rhizomys pruinosus* | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 9 |
| *Rusa unicolor* | 15 | 274 | 77 | 1 | 0 | 177 | 1950 | 2494 |
| *Semnopithecus schistaceus* | 0 | 0 | 0 | 0 | 217 | 0 | 0 | 217 |
| Squirrel species | 5 | 6 | 0 | 1 | 39 | 0 | 154 | 205 |
| *Sus scrofa* | 496 | 103 | 156 | 263 | 87 | 3559 | 2707 | 7371 |
| *Tamiops rodolphii* | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 6 |
| *Tapirus indicus* | 0 | 0 | 0 | 0 | 0 | 182 | 0 | 182 |
| *Trachypithecus obscurus* | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 10 |
| *Trachypithecus shortridgei* | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 6 |
| *Tragopan satyra* | 0 | 0 | 0 | 0 | 17 | 0 | 0 | 17 |
| *Tragulus kanchil* | 350 | 0 | 0 | 0 | 0 | 361 | 297 | 1008 |
| *Tupaia belangeri* | 25 | 9 | 0 | 0 | 0 | 0 | 0 | 34 |
| *Tupaia glis* | 0 | 0 | 0 | 0 | 0 | 117 | 0 | 117 |
| *Ursus thibetanus* | 1 | 2 | 1 | 6 | 5 | 0 | 233 | 248 |
| *Varanus salvator* | 0 | 0 | 0 | 0 | 0 | 17 | 119 | 136 |
| *Viverra megaspila* | 0 | 0 | 0 | 0 | 0 | 0 | 51 | 51 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Viverra zibetha* | 198 | 188 | 105 | 157 | 17 | 0 | 760 | 1425 |
| *Viverricula indica* | 1 | 0 | 10 | 0 | 0 | 0 | 184 | 195 |
| *Vulpes vulpes* | 0 | 0 | 0 | 0 | 28 | 0 | 0 | 28 |
| Weasel species | 0 | 0 | 0 | 0 | 17 | 0 | 0 | 17 |

**Table S4.** Biplot covariates' scores. Shown are the first two CCA axes.

| Covariates | CCA1 | CCA2 |
|---|---|---|
| Camera trap nights | 0.260 | -0.473 |
| Elevation FM | 0.226 | 0.804 |
| Mosaic areas GYRATE_AM | -0.688 | 0.191 |
| Non-forested areas GYRATE_AM | 0.646 | -0.342 |
| Protected areas GYRATE_AM | 0.704 | -0.147 |
| Shrublands and grasslands PLAND | 0.553 | 0.287 |
| Slope position FM | 0.185 | -0.353 |
| Forested areas GYRATE_AM | -0.214 | 0.117 |
| Forested areas PLAND | -0.508 | 0.216 |
| Water GYRATE_AM | -0.484 | -0.169 |
| Water PLAND | -0.544 | -0.450 |
| Y | 0.813 | 0.463 |

**Table S5.** Covariates' coefficients of the single-species GLMs. I= Intercept; CTN= Camera trap nights; E_F= Elevation FM; M_G= Mosaic areas GYRATE_AM; NF_G= Non-forested areas GYRATE_AM; PA_G= Protected areas GYRATE_AM; SG_P= Shrublands and grasslands PLAND; SP_F= Slope position FM; F_G= Forested areas GYRATE_AM; F_P= Forested areas PLAND; W_G= Water GYRATE_AM; W_P= Water PLAND; Y= Y. See Table S2 for species' codes.

| Species' code | I | CTN | E_F | M_G | NF_G | PA_G | SG_P | SP_F | F_G | F_P | W_G | W_P | Y |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 43 | -42.08 | -0.13 | 8.93 | -17.63 | -7.91 | -9.57 | -2.38 | -0.21 | 1.27 | -12.86 | -5.31 | 38.98 | 29.99 |
| 51 | -310.60 | 0.04 | 49.82 | 11.84 | -45.16 | 21.36 | 161.42 | -0.28 | 0.53 | 77.63 | -84.39 | 273.48 | 198.42 |
| 79 | -65.67 | -0.01 | -50.15 | -27.80 | -5.85 | 14.74 | 9.13 | 1.62 | -0.27 | 6.00 | -27.94 | -10.04 | 19.75 |
| 78* | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 77 | -161.95 | 0.01 | 28.76 | -8.83 | -3.16 | 1.64 | -82.75 | 0.28 | 4.80 | -43.56 | 0.28 | 89.52 | 113.13 |
| 19 | -4.09 | 0.00 | -0.63 | 0.57 | -0.31 | 0.45 | 0.68 | -0.03 | 0.45 | 0.49 | -0.10 | 0.31 | -0.58 |
| 44 | -3.42 | 0.00 | -1.09 | -0.82 | 1.12 | 1.29 | 2.39 | 0.21 | 0.99 | 0.80 | -0.68 | -0.49 | -0.09 |
| 81 | -62.46 | 0.02 | -39.29 | -0.08 | -8.71 | -0.82 | -34.26 | 1.22 | 4.33 | 2.55 | 1.01 | -1.67 | 2.38 |
| 57 | -1.82 | 0.00 | -1.74 | 0.01 | -0.21 | -0.41 | 1.00 | -0.04 | 0.43 | 0.98 | -0.02 | -0.11 | -0.27 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 40 | -573.15 | 0.00 | 11.18 | -7.74 | 13.06 | 2.90 | 2.87 | 0.58 | -0.01 | -2.29 | -1167.07 | 421.61 | 45.18 |
| 12 | -2.00 | 0.00 | -0.37 | -0.14 | 1.16 | 1.54 | 1.45 | -0.35 | 0.20 | -0.30 | 1.57 | -1.13 | -1.93 |
| 11 | -11.60 | 0.00 | -13.05 | 1.52 | 0.86 | 2.10 | 3.32 | 0.16 | 0.60 | 1.10 | -2.53 | 0.46 | -2.83 |
| 60 | -146.80 | -0.02 | 4.27 | -2.22 | -5.77 | 0.09 | -2.88 | -0.07 | 0.26 | -4.71 | -281.21 | 101.79 | 12.11 |
| 39 | -6.20 | 0.00 | -0.47 | -0.60 | 1.96 | 0.53 | 1.82 | -0.03 | -0.41 | -0.53 | 1.05 | -1.04 | -1.38 |
| 15 | -3.08 | 0.00 | -0.43 | -0.49 | 0.50 | 0.83 | 2.06 | -0.13 | 0.57 | 0.89 | -0.70 | 0.31 | -0.68 |
| 14 | -232.64 | 0.05 | 1.49 | -7.62 | -2.42 | -0.62 | -2.57 | -0.04 | 2.15 | -11.21 | 47.67 | -256.12 | 28.25 |
| 35 | -3.27 | 0.00 | -0.27 | -0.35 | -1.15 | -0.04 | 0.45 | 0.19 | 0.03 | 0.76 | 0.03 | 0.27 | 1.01 |
| 22 | -1065.35 | 0.00 | 44.88 | 1.49 | 1.27 | 0.63 | 9.79 | -0.31 | 0.96 | 7.81 | -2180.46 | 836.55 | 87.86 |
| 38 | -1.96 | 0.00 | -2.13 | 0.16 | 1.00 | 0.51 | 2.54 | -0.23 | 0.05 | 1.54 | 0.03 | -0.65 | -0.82 |
| 63 | -154.99 | 0.11 | 91.34 | 23.34 | -1.87 | -1.49 | 36.52 | -2.52 | -55.12 | 105.64 | NA | NA | 36.49 |
| 2 | -88.40 | 0.01 | -120.91 | -0.22 | 3.39 | -1.75 | 12.19 | -0.04 | 3.36 | 47.26 | -0.11 | 0.28 | 5.71 |
| 1 | -1.34 | 0.00 | -0.95 | 0.47 | 1.71 | 0.06 | 1.07 | 0.23 | -0.04 | 1.06 | -0.39 | 0.30 | -0.90 |
| 47 | -468.94 | 0.03 | 8.27 | 1.35 | -10.50 | 1.93 | -0.53 | -0.37 | 0.60 | -10.82 | 83.99 | -551.83 | 26.94 |
| 80 | -2.37 | 0.00 | -2.02 | 0.86 | 2.21 | 0.74 | 3.27 | 0.19 | 0.50 | 1.85 | 0.36 | -1.32 | -2.23 |
| 41 | -1.54 | 0.00 | -0.92 | 0.93 | 1.36 | 0.06 | 1.52 | -0.12 | 0.51 | 1.41 | -0.32 | -0.07 | -0.86 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 23 | -40.42 | 0.01 | -37.67 | 0.14 | 7.39 | 0.31 | 7.60 | 0.02 | 2.20 | 20.20 | 0.32 | -1.42 | -11.69 |
| 28 | -18.25 | 0.01 | 6.16 | -3.17 | 6.78 | 4.26 | 4.42 | 0.10 | -0.34 | -2.03 | 8.91 | -3.21 | -13.97 |
| 27 | -2.85 | 0.00 | -2.04 | 0.65 | 1.65 | 0.16 | 3.26 | -0.35 | -0.02 | 2.60 | -0.90 | -0.06 | -1.03 |
| 58 | -0.15 | 0.00 | -0.06 | 0.50 | 0.52 | -0.23 | 0.51 | -0.05 | 0.32 | 0.49 | 0.23 | -0.09 | -0.26 |
| 83 | -97.14 | -0.35 | 32.04 | -138.31 | 18.65 | -53.24 | -4.20 | -0.23 | -8.06 | 35.40 | 13.94 | -1.45 | -55.34 |
| 66 | -507.57 | 0.25 | 10.38 | -61.68 | -276.89 | 124.79 | -55.66 | 68.38 | 277.12 | -137.60 | -3.54 | 214.97 | 158.86 |
| 55 | -36.70 | 0.01 | 10.19 | -4.63 | 19.92 | -0.87 | 6.96 | 0.30 | -2.93 | 2.66 | 14.37 | 1.70 | -18.72 |
| 85 | -99.56 | -0.17 | 18.88 | -19.24 | 36.27 | -44.83 | 30.61 | 0.02 | -2.53 | 29.68 | -5.22 | 79.36 | 22.28 |
| 86 | -3.62 | 0.01 | -0.61 | 0.88 | 2.70 | 1.53 | 2.41 | -0.20 | 0.91 | 0.33 | 1.38 | -2.28 | -3.62 |
| 87 | -27.16 | 0.01 | -0.50 | -1.65 | -2.08 | 1.06 | 2.56 | -0.09 | -0.36 | 0.19 | 1.11 | -5.50 | 15.76 |
| 88 | -4.56 | 0.00 | -0.25 | -0.72 | -0.09 | 0.60 | 1.05 | -0.24 | 1.25 | 0.65 | -3.07 | 0.25 | -1.17 |
| 89 | -56.41 | 0.02 | -39.22 | -0.08 | 1.70 | -2.00 | -5.13 | 0.78 | 11.80 | 16.16 | -3.91 | 4.61 | -2.21 |
| 52 | -1223.72 | 1.02 | -120.17 | -36.36 | -21.19 | -74.39 | 314.92 | 67.90 | 72.23 | -95.92 | -396.11 | -501.92 | -878.01 |
| 73 | -14.93 | 0.01 | 2.90 | -0.86 | -6.89 | 0.50 | 1.52 | 0.11 | 0.73 | 2.13 | 1.27 | -1.64 | 4.11 |
| 72 | -19.32 | 0.01 | 0.62 | -5.60 | -2.86 | 0.05 | -1.59 | 0.33 | 0.27 | -0.54 | -6.60 | -4.81 | 5.27 |
| 71 | -1.88 | 0.00 | -1.80 | -0.06 | 0.28 | 0.65 | 0.31 | 0.22 | 0.12 | -0.15 | -0.16 | -0.56 | -0.57 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 74 | -60.09 | 0.03 | 19.45 | 8.29 | -39.50 | 12.38 | -3.55 | -0.36 | 2.82 | -18.39 | -0.13 | 41.32 | 43.89 |
| 70 | -35.53 | 0.01 | -13.95 | -0.05 | 2.40 | 0.10 | -2.08 | 0.08 | 0.81 | 6.75 | 0.56 | -0.54 | -12.39 |
| 18 | -8.63 | 0.00 | 1.66 | 0.11 | 3.03 | -0.06 | 0.55 | -0.02 | 0.05 | 0.87 | 2.07 | -0.41 | -4.32 |
| 45 | -288.51 | 0.00 | 23.17 | 60.17 | -29.09 | 6.36 | -19.26 | -0.29 | 1.29 | -21.02 | -104.09 | 325.56 | 311.93 |
| 46 | -1.47 | 0.00 | 0.51 | 0.47 | 0.52 | 0.12 | 1.37 | 0.11 | 0.23 | 1.36 | -0.51 | 0.49 | -0.16 |
| 48 | -73.99 | -0.04 | 0.54 | 1.17 | -2.75 | -0.92 | -1.39 | -0.38 | 0.16 | -0.54 | -140.96 | 45.99 | -1.47 |
| 16 | -447.82 | -0.09 | -47.53 | -93.19 | 3.21 | -3.38 | 36.77 | -0.40 | 2.91 | -5.81 | -215.73 | 363.95 | 456.97 |
| 4 | -282.15 | 0.00 | 151.20 | 1.07 | -25.65 | 14.65 | 66.73 | -0.48 | 0.71 | -7.58 | -14.64 | 2.13 | -213.34 |
| 7 | -27.72 | 0.01 | -3.36 | -0.57 | -1.07 | 1.27 | 13.03 | -0.44 | 1.10 | 8.57 | 0.33 | -0.40 | -18.11 |
| 8 | -213.37 | 0.10 | 8.34 | -8.33 | -10.53 | 1.01 | -1.03 | 0.06 | -0.16 | -5.42 | 47.72 | -255.17 | -0.31 |
| 9 | 0.64 | 0.00 | 0.04 | 0.47 | 0.78 | 0.34 | 0.46 | 0.01 | 0.12 | 0.48 | 0.26 | -0.69 | 0.25 |
| 6* | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 50 | -100.89 | -0.01 | -34.07 | -51.05 | 72.46 | -9.37 | -5.19 | 0.56 | NA | 83.46 | 34.66 | -77.59 | NA |
| 13 | -167.81 | 0.09 | 62.11 | -13.48 | -35.34 | -9.73 | -48.47 | 0.09 | 7.48 | 51.33 | 2.43 | 41.22 | 28.38 |
| 31 | -1.86 | 0.00 | -0.48 | 0.35 | 0.59 | -0.29 | 1.08 | 0.04 | 0.30 | 1.08 | -0.07 | 0.10 | 0.26 |
| 20 | -6.47 | 0.00 | 3.03 | -0.07 | -3.41 | 0.23 | 0.85 | -0.05 | -0.01 | 0.72 | 0.21 | 1.37 | -0.17 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 33 | -7.10 | 0.00 | 2.61 | 1.11 | -0.90 | 1.03 | -3.00 | 0.04 | 0.20 | -0.56 | -0.88 | 1.72 | 0.95 |
| 32 | -4.20 | 0.00 | -1.28 | -2.42 | 1.62 | 1.33 | 0.54 | -0.04 | -0.23 | 1.59 | -1.10 | 0.89 | 1.23 |
| 21 | -1.30 | 0.00 | -1.57 | 0.77 | 0.77 | 0.34 | 1.86 | 0.00 | 0.13 | 1.27 | 0.09 | -0.76 | -1.09 |
| 34 | -3.32 | 0.00 | -0.63 | -0.10 | -1.14 | -0.05 | 0.44 | -0.07 | 0.59 | 0.37 | 0.09 | -0.03 | 0.22 |
| 76 | -40.95 | -0.29 | 25.99 | 10.87 | -15.92 | 2.62 | -36.45 | 0.25 | 4.14 | -4.34 | -0.77 | -0.25 | 9.77 |
| 75 | -404.04 | -0.03 | 136.78 | 69.71 | 54.56 | -94.59 | -108.39 | 0.10 | 2.29 | -144.73 | 112.85 | 44.14 | 25.37 |
| 82 | -11.54 | 0.00 | -0.73 | 0.53 | -4.45 | -0.47 | 0.55 | 0.89 | 1.00 | -0.44 | 0.32 | 0.94 | 4.87 |
| 36 | -0.93 | 0.00 | -0.14 | 0.82 | 1.24 | -0.06 | 1.22 | -0.04 | 0.29 | 1.40 | 0.15 | -0.29 | -0.40 |
| 29 | -105.28 | 0.03 | 10.05 | 1.06 | -9.73 | 3.49 | -20.43 | 0.11 | 0.45 | 10.91 | -4.53 | 0.33 | -34.78 |
| 30 | -18.05 | 0.02 | 4.26 | -1.28 | -9.05 | -0.14 | -1.92 | -0.28 | -0.41 | 0.10 | 1.71 | -2.87 | 2.42 |
| 56 | -1.54 | 0.00 | 0.32 | 0.11 | 0.84 | 0.09 | 1.01 | 0.20 | 0.91 | 1.00 | 0.16 | -0.53 | -1.80 |
| 65 | -75.39 | -0.06 | 7.87 | 0.10 | -19.50 | -2.50 | -5.18 | -0.94 | 0.85 | -0.94 | -137.68 | 62.47 | 1.98 |
| 61 | -236.83 | -0.16 | 112.30 | -15.93 | 12.93 | -7.51 | -120.41 | -2.94 | 1.39 | -6.13 | -43.96 | -23.27 | -115.07 |
| 64 | -158.92 | 0.01 | 44.35 | -198.42 | 24.84 | 25.45 | -12.67 | -2.68 | -0.53 | 19.02 | NA | NA | 58.97 |
| 10 | -0.57 | 0.00 | -0.10 | -0.90 | 0.03 | 0.10 | 1.23 | 0.25 | -0.93 | 0.34 | 0.47 | 0.00 | -1.46 |
| 67 | -54.81 | 0.01 | 13.69 | 16.94 | 9.06 | -6.71 | -14.39 | 0.20 | 0.74 | -1.50 | 3.16 | 5.70 | 12.50 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 59 | -4.05 | 0.00 | -0.40 | -0.66 | 2.32 | -0.05 | 1.07 | 0.12 | -0.13 | 1.69 | -1.64 | 0.71 | 0.50 |
| 3 | 0.69 | 0.00 | -0.14 | 0.37 | 0.48 | -0.25 | 0.16 | -0.04 | 0.29 | 0.27 | 0.30 | -0.25 | -0.85 |
| 62* | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 17 | -95.75 | -0.01 | -9.49 | 0.99 | 22.72 | 0.18 | -7.29 | -0.04 | 1.36 | 12.16 | 2.42 | -0.27 | -46.07 |
| 69 | -118.75 | 0.08 | 43.83 | 0.74 | -31.55 | 6.25 | -9.58 | -0.49 | 6.29 | -7.66 | -5.31 | 3.53 | -48.45 |
| 68 | -36.98 | 0.01 | -8.18 | 1.59 | 3.01 | 1.86 | -4.91 | 4.01 | -4.88 | -4.07 | -0.91 | 1.91 | 25.85 |
| 84* | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 5 | -7.65 | 0.00 | 2.08 | 0.34 | 3.49 | 1.56 | -5.31 | -0.24 | -0.20 | 0.68 | 1.29 | -0.89 | -3.56 |
| 54 | -284.16 | 0.19 | -9.99 | -22.50 | -189.82 | 5.30 | 69.75 | 0.51 | -13.82 | 85.59 | 12.89 | 51.50 | 22.58 |
| 53 | -118.32 | 0.02 | -120.22 | -1.59 | 10.33 | -2.28 | -37.90 | 1.76 | 1.42 | 15.13 | 6.74 | -4.86 | -9.55 |
| 42 | -3.56 | 0.00 | -0.41 | -0.54 | 1.26 | 0.86 | 0.19 | 0.17 | 0.51 | 0.46 | -0.45 | 0.24 | 0.21 |
| 90 | -6.00 | 0.01 | 0.45 | 0.29 | 1.52 | 0.75 | 2.98 | 0.03 | 0.94 | -0.11 | 1.06 | -1.78 | -5.52 |
| 26 | -7.22 | 0.00 | 0.75 | -3.05 | -0.18 | 2.20 | 3.03 | -0.27 | 0.24 | -1.57 | 3.52 | -1.60 | -5.56 |
| 25 | -1.10 | 0.00 | -1.20 | 0.61 | 1.57 | 0.23 | 2.28 | -0.06 | -0.04 | 2.16 | -0.34 | -0.59 | -0.36 |
| 24 | -5.73 | 0.01 | 0.36 | 0.38 | 2.20 | 1.73 | 4.19 | 0.04 | 0.01 | 0.81 | 1.05 | -1.28 | -4.18 |
| 37 | -84.15 | -0.05 | -16.89 | -33.08 | -2.97 | 10.55 | 16.14 | -0.17 | -1.32 | 1.83 | -54.34 | 78.33 | 93.84 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 49 | -82.92 | 0.00 | 4.97 | -25.41 | 14.60 | -18.07 | 22.76 | -0.11 | 1.30 | 66.42 | -39.67 | 44.67 | 23.10 |
| Positive/ | 2/84 | 70/16 | 42/44 | 41/45 | 49/37 | 55/31 | 55/31 | 41/45 | 62/23 | 56/30 | 42/42 | 40/44 | 40/45 |
| Negative | (2.3/ | (81.4/ | (48.8/ | (47.7/ | (57.0/ | (64.0/ | (64.0/ | (47.7/ | (72.1/ | (65.1/ | (48.8/ | (46.5/ | (46.5/ |
| (%/%) | 97.7) | 18.6) | 51.2) | 52.3) | 43.0) | 36.0) | 36.0) | 52.3) | 26.7) | 34.9) | 48.8) | 51.2) | 52.3) |

* Single-species GLM did not converge.

**Table S6.** Proportion of protected biodiversity, per country. Only the portions of China, India and Malaysia falling within the study area have been considered in the ratio.

| Countries | % of protected biodiversity |
|---|---|
| Bangladesh | 12.23 |
| Bhutan | 46.92 |
| Cambodia | 31.25 |
| China | 13.05 |
| India | 12.93 |
| Laos | 24.27 |
| Malaysia | 13.05 |
| Myanmar | 17.55 |
| Nepal | 28.37 |
| Singapore | 6.51 |
| Taiwan | 39.77 |
| Thailand | 28.42 |
| Vietnam | 22.83 |

**Table S7.** Main systems of protected areas preserving most biodiversity. For the sake of brevity, only the first 10 are shown.

| Code | % of total biodiversity | % of protected biodiversity | Protected areas | Countries |
|---|---|---|---|---|
| 1 | 1.37 | 7.38 | Huangheshouqu Nature Reserve | China |
| | | | Manzetangshidi Nature Reserve | China |
| | | | Nianlong Nature Reserve | China |
| | | | Ruoergaishidi Nature Reserve | China |
| | | | Sanjiangyuan Nature Reserve | China |
| | | | Tiebu Nature Reserve | China |
| | | | Yanboyezeshan Nature Reserve | China |
| 2 | 1.35 | 7.27 | Baimaxueshan Nature Reserve | China |
| | | | Bumphabum Wildlife Sanctuary | Myanmar |
| | | | Hkakabo Razi National Park | Myanmar |
| | | | Hponkan Razi National Park | Myanmar |
| | | | Hukawng Valley Tiger Reserve | Myanmar |
| | | | Namdapha Wildlife Sanctuary | India |
| 3 | 0.84 | 4.52 | Khammouane Limestone (Phou Hin Poun) National Biodiversity Conservation Area | Laos |
| | | | Nakai - Nam Theun National Biodiversity Conservation Area | Laos |
| | | | Nam Chuane Conservation Area | Laos |

| | | | | |
|---|---|---|---|---|
| | | | Nam Theun Ext. National Biodiversity Conservation Area | Laos |
| | | | Vu Quang National Park | Vietnam |
| | | | Western Nghe An UNESCO-MAB Biosphere Reserve | Vietnam |
| | | | Xuan Lien Nature Reserve | Vietnam |
| 4 | 0.80 | 4.30 | Chalearm Rattanakosin National Park | Thailand |
| | | | Erawan National Park | Thailand |
| | | | Huai Kha Khaeng Wildlife Sanctuary | Thailand |
| | | | Khao Laem National Park | Thailand |
| | | | Khlong Lamngu National Park | Thailand |
| | | | Khlong Lan National Park | Thailand |
| | | | Khlong Wang Chao National Park | Thailand |
| | | | Khuen Si Nakarin National Park | Thailand |
| | | | Lansaang National Park | Thailand |
| | | | Mae Wong National Park | Thailand |
| | | | Namtok Phacharoen National Park | Thailand |
| | | | Phu Toei National Park | Thailand |
| | | | Sai Yok National Park | Thailand |
| | | | Salak Phra Wildlife Sanctuary | Thailand |
| | | | Taksin Maharat National Park | Thailand |
| | | | Thong Pha Phum National Park | Thailand |
| | | | Taninthayi Nature Reserve | Myanmar |

| | | | | |
|---|---|---|---|---|
| | | | Thungyai - Huai Kha Khaeng Wildlife Sanctuaries World Heritage Site | Thailand |
| | | | Thungyai Naresuan Wildlife Sanctuary | Thailand |
| | | | Um Phang Wildlife Sanctuary | Thailand |
| 5 | 0.59 | 3.15 | Central Cardamom Mountains Protected Forest | Myanmar |
| | | | Khlong Kruewai Chalearm Phrakiat Wildlife Sanctuary | Thailand |
| | | | Namtok Klong Kaew National Park | Thailand |
| | | | Phnom Aural Wildlife Sanctuary | Myanmar |
| | | | Phnom Samkos Wildlife Sanctuary | Myanmar |
| | | | Samlaut Multiple Use Management Area | Myanmar |
| 6 | 0.55 | 2.97 | Chu Mom Ray National Park | Vietnam |
| | | | Dong Ampham National Biodiversity Conservation Area | Laos |
| | | | Dong Houa Sao National Biodiversity Conservation Area | Laos |
| | | | Phou Kateup (Bolovens Northeast) | Laos |
| | | | Phou Kathong | Laos |
| | | | Phu Luang (Bolovens Southwest) | Laos |
| | | | Virachey National Park and ASEAN Heritage Park | Myanmar |

| | | | | |
|---|---|---|---|---|
| | | | Xe Khampho | Laos |
| | | | Xe Pian National Biodiversity Conservation Area | Laos |
| 7 | 0.53 | 2.84 | Bu Gia Map National Park | Vietnam |
| | | | Chu Prong Nature Reserve | Vietnam |
| | | | Lomphat Wildlife Sanctuary | Myanmar |
| | | | Mondulkiri Protected Forest | Myanmar |
| | | | Nsok Protected Forest | Myanmar |
| | | | Phnom Prich Wildlife Sanctuary | Myanmar |
| | | | Pnom Namlear Wildlife Sanctuary | Myanmar |
| | | | Seima Protected Forest | Myanmar |
| | | | Snoul Wildlife Sanctuary | Myanmar |
| | | | Snoul Wildlife Sanctuary | Myanmar |
| 8 | 0.50 | 2.67 | Annapurna Conservation Area | Nepal |
| | | | Gauri-Shankar Conservation Area | Nepal |
| | | | Kanchanjunga Conservation Area | Nepal |
| | | | Khangchendzonga National Park | India |
| | | | Khangchendzonga National Park World Heritage Site | India |
| | | | Langtang National Park | Nepal |
| | | | Lantang - Buffer Zone National Park - Buffer Zone | Nepal |
| | | | Makalu-Barun National Park | Nepal |
| | | | Makalu-Barun - Buffer Zone National Park - Buffer Zone | Nepal |

| | | | | |
|---|---|---|---|---|
| | | | Manaslu Conservation Area | Nepal |
| | | | Quomolangma Nature Preserve | China |
| | | | Sagarmatha National Park | Nepal |
| | | | Sagarmatha - Buffer Zone National Park - Buffer Zone | Nepal |
| | | | Sagarmatha National Park World Heritage Site | Nepal |
| 9 | 0.45 | 2.40 | Banteay Chhmar Protected Landscape | Myanmar |
| | | | Boon Trik - Yod Mon Wildlife Sanctuary | Thailand |
| | | | Dong Phayayen-Khao Yai Forest Complex World Heritage Site | Thailand |
| | | | Huai Sala Wildlife Sanctuary | Thailand |
| | | | Huai Thabthan-Had Samran Wildlife Sanctuary | Thailand |
| | | | Khao Phravihan National Park | Thailand |
| | | | Khao Yai National Park and ASEAN Heritage Park | Thailand |
| | | | Panom Dong Rak Wildlife Sanctuary | Thailand |
| | | | Phu Chong - Na Yoi National Park | Thailand |
| | | | Preah Vihear Protected Forest | Myanmar |
| | | | Preah Vihear Temple Protected Landscape | Myanmar |

| | | | | |
|---|---|---|---|---|
| | | | Ta Phraya National Park | Thailand |
| | | | Thap Lan National Park | Thailand |
| | | | Yod Dom Wildlife Sanctuary | Thailand |
| 10 | 0.44 | 2.37 | Barnadi Sanctuary | India |
| | | | Biological Corridor 1 Biological Corridor | Bhutan |
| | | | Biological Corridor 2 Biological Corridor | Bhutan |
| | | | Biological Corridor 3 Biological Corridor | Bhutan |
| | | | Biological Corridor 4 Biological Corridor | Bhutan |
| | | | Biological Corridor 5 Biological Corridor | Bhutan |
| | | | Biological Corridor 6 Biological Corridor | Bhutan |
| | | | Biological Corridor 7 Biological Corridor | Bhutan |
| | | | Biological Corridor 8 Biological Corridor | Bhutan |
| | | | Bumdeling Wildlife Sanctuary | Bhutan |
| | | | Jigme Dorji National Park | Bhutan |
| | | | Jigme Singye Wangchuck National Park | Bhutan |
| | | | Kalaktang Sanctuary | India |

| | | |
|---|---|---|
| | Khaling Wildlife Sanctuary | Bhutan |
| | Manas Sanctuary | India |
| | Manas National Park | India |
| | Manas Wildlife Sanctuary World Heritage Site | India |
| | Phipsoo Wildlife Sanctuary | Bhutan |
| | Royal Manas National Park | Bhutan |
| | Sakteng Wildlife Sanctuary | Bhutan |
| | Thrumshingla National Park | Bhutan |
| | Torsa Strict Nature Reserve | Bhutan |
| | Wangchuck Centennial Park National Park | Bhutan |

**Table S8.** Species importance. Shown are the deviance explained values of the single-species models. Species are ranked by their deviance explained values.

| Species | Deviance explained |
|---|---|
| *Viverricula indica* | 5.09E-01 |
| *Gallus gallus* | 4.80E-01 |
| *Varanus salvator* | 4.64E-01 |
| *Lophura diardi* | 4.39E-01 |
| Rat species | 4.25E-01 |
| *Bos gaurus* | 4.07E-01 |
| *Cuon alpinus* | 3.90E-01 |
| *Herpestes urva* | 3.81E-01 |
| *Viverra megaspila* | 3.78E-01 |
| *Viverra zibetha* | 3.56E-01 |
| *Rusa unicolor* | 3.48E-01 |
| Mouse-deer species | 3.37E-01 |
| *Capricornis milneedwardsii* | 3.03E-01 |
| *Herpestes javanicus* | 2.98E-01 |
| *Helarctos malayanus* | 2.80E-01 |
| *Muntiacus muntjak* | 2.73E-01 |
| *Paradoxurus hermaphroditus* | 2.73E-01 |
| *Arctonyx collaris* | 2.65E-01 |
| *Prionailurus bengalensis* | 2.36E-01 |
| *Bos javanicus* | 2.33E-01 |

| | |
|---|---|
| *Martes flavigula* | 2.29E-01 |
| *Hemigalus derbyanus* | 2.24E-01 |
| *Elephas maximus* | 2.07E-01 |
| *Lutrogale perspicillata* | 2.01E-01 |
| *Macaca leonina* | 1.97E-01 |
| *Lepus peguensis* | 1.74E-01 |
| *Macaca nemestrina* | 1.57E-01 |
| *Aonyx cinereus* | 1.50E-01 |
| *Tapirus indicus* | 1.49E-01 |
| *Canis aureus* | 1.48E-01 |
| Squirrel species | 1.29E-01 |
| *Lophura nycthemera* | 1.28E-01 |
| *Trachypithecus obscurus* | 1.24E-01 |
| *Tragulus kanchil* | 1.17E-01 |
| *Prionodon linsang* | 1.13E-01 |
| *Manis javanica* | 8.35E-02 |
| *Rhinosciurus laticaudatus* | 7.83E-02 |
| *Neofelis nebulosa* | 7.41E-02 |
| *Hystrix brachyura* | 7.41E-02 |
| *Atherurus macrourus* | 7.34E-02 |
| *Lophophorus impejanus* | 7.16E-02 |
| *Chrotogale owstoni* | 6.89E-02 |
| *Arctictis binturong* | 6.63E-02 |
| Bear species | 6.46E-02 |
| *Lophura rufa* | 6.43E-02 |

| | |
|---|---|
| *Tupaia glis* | 5.59E-02 |
| *Ithaginis cruentus* | 5.52E-02 |
| *Panthera tigris* | 4.63E-02 |
| *Paguma larvata* | 3.39E-02 |
| *Callosciurus erythraeus* | 3.21E-02 |
| Weasel species | 3.13E-02 |
| *Tupaia belangeri* | 3.03E-02 |
| *Lophura leucomelanos* | 2.99E-02 |
| *Echinosorex gymnura* | 2.85E-02 |
| *Trachypithecus shortridgei* | 2.21E-02 |
| *Vulpes vulpes* | 1.88E-02 |
| *Ursus thibetanus* | 1.57E-02 |
| *Pardofelis marmorata* | 1.48E-02 |
| *Arborophila atrogularis* | 1.15E-02 |
| *Macaca arctoides* | 9.21E-03 |
| *Leopoldamys sabanus* | 9.14E-03 |
| *Naemorhedus goral* | 8.68E-03 |
| *Macaca mulatta* | 8.43E-03 |
| *Sus scrofa* | 8.35E-03 |
| *Moschus leucogaster* | 7.50E-03 |
| *Arborophila torqueola* | 7.16E-03 |
| *Prionodon pardicolor* | 6.89E-03 |
| *Catopuma temminckii* | 6.44E-03 |
| *Panthera pardus* | 6.22E-03 |
| *Melogale personata* | 5.10E-03 |

| | |
|---|---|
| *Ailurus fulgens* | 3.90E-03 |
| *Macaca assamensis* | 3.48E-03 |
| *Semnopithecus schistaceus* | 3.29E-03 |
| *Polyplectron bicalcaratum* | 1.94E-03 |
| Marten species | 1.52E-03 |
| *Rattus exulans* | 1.27E-03 |
| *Argusianus argus* | 5.64E-04 |
| Pheasant species | 4.49E-04 |
| Partridge species | 4.37E-04 |
| *Muntiacus rooseveltorum* | 2.41E-04 |
| *Capricornis thar* | 6.42E-05 |
| *Muntiacus vaginalis* | 4.21E-05 |
| *Mustela kathiah* | 1.11E-16 |
| Ferret-badger species | 1.11E-16 |
| *Rhizomys pruinosus* | 1.11E-16 |
| *Dremomys rufigenis* | 1.11E-16 |
| Muntjac species* | NA |
| *Tamiops rodolphii** | NA |
| *Arborophila cambodiana** | NA |
| *Tragopan satyra** | NA |

* Single-species GLM did not converge.

**Table S9.** Results of the Tukey's test on the residuals of the GLM between the modelled and the empirical number of species.

| Countries | Difference | Lower | Upper | Adjusted-*p*-value |
|---|---|---|---|---|
| India-Cambodia | -2.39 | -3.06 | -1.72 | <0.0001 |
| Laos-Cambodia | -1.61 | -2.13 | -1.09 | <0.0001 |
| Myanmar-Cambodia | -1.15 | -1.73 | -0.56 | <0.0001 |
| Nepal-Cambodia | -3.26 | -4.04 | -2.49 | <0.0001 |
| Peninsular Malaysia-Cambodia | 0.93 | 0.31 | 1.54 | <0.001 |
| Thailand-Cambodia | -0.30 | -0.73 | 0.13 | 0.37 |
| Laos-India | 0.78 | 0.12 | 1.44 | <0.01 |
| Myanmar-India | 1.24 | 0.53 | 1.95 | <0.0001 |
| Nepal-India | -0.88 | -1.75 | 0.00 | <0.05 |
| Peninsular Malaysia-India | 3.32 | 2.58 | 4.05 | <0.0001 |
| Thailand-India | 2.09 | 1.50 | 2.68 | <0.0001 |
| Myanmar-Laos | 0.46 | -0.11 | 1.04 | 0.21 |
| Nepal-Laos | -1.65 | -2.42 | -0.89 | <0.0001 |
| Peninsular Malaysia-Laos | 2.54 | 1.93 | 3.15 | <0.0001 |
| Thailand-Laos | 1.31 | 0.89 | 1.73 | <0.0001 |
| Nepal-Myanmar | -2.12 | -2.93 | -1.31 | <0.0001 |
| Peninsular Malaysia-Myanmar | 2.08 | 1.41 | 2.74 | <0.0001 |
| Thailand-Myanmar | 0.85 | 0.35 | 1.35 | <0.0001 |
| Peninsular Malaysia-Nepal | 4.19 | 3.36 | 5.03 | <0.0001 |
| Thailand-Nepal | 2.96 | 2.25 | 3.67 | <0.0001 |

| | | | | |
|---|---|---|---|---|
| Thailand-Peninsular Malaysia | -1.23 | -1.76 | -0.70 | <0.0001 |

**References**


CIESIN, and CIAT. 2016. Gridded Population of the World version 4 (GPWv4): Population density grids. Socioeconomic Data and Applications Center (SEDAC), Columbia University, Palisades, NY. http://sedac.ciesin.columbia.edu/gpw.

ESA Land Cover CCI. 2017. Product user guide version 2.0. http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf.

Evans, J. S., J. Oakleaf, S. A. Cushman, and D. Theobald. 2014. An ArcGIS toolbox for surface gradient and geomorphometric modeling, version 2.0-0. Available: http://evansmurphy.wix.com/evansspatial.

Fick, S. E., and R. J. Hijmans. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37:4302-4315.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science 342:850-853.

Jarvis, A., H. I. Reuter, A. Nelson, and E. Guevara. 2008. Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT). http://srtm.csi.cgiar.org.

McGarigal, K., S. A. Cushman, and E. Ene. 2012. FRAGSTATS v4: spatial pattern analysis program for categorical and continuous maps. Computer software program produced by the authors at the University of Massachusetts, Amherst, MA. Available: http://www.umass.edu/landeco/research/fragstats/fragstats.html.

UNEP-WCMC, and IUCN. 2022. Protected Planet: The World Database on Protected Areas (WDPA), Accessed April 2022, Cambridge, UK: UNEP-WCMC and IUCN. Available at: www.protectedplanet.net.

WCS, and CIESIN. 2005. Last of The Wild data version 2, (LTW-2): global human footprint dataset (geographic). http://sedac.ciesin.columbia.edu/wildareas/.

# Chapter 3

# Multi-scale, multivariate community models improve designation of biodiversity hotspots in the Sunda Islands.

Luca Chiaverini*, David W. Macdonald*, Helen M. Bothwell, Andrew J. Hearn, Susan M. Cheyne, Iding Haidir, Luke T. B. Hunter, Żaneta Kaszta, Ewan A. Macdonald, Joanna Ross, Samuel A. Cushman

* Joint first authors

Published in Animal Conservation. (2022) 25: 660-679

https://doi.org/10.1111/acv.12771

**Abstract**

Species occur in sympatric assemblages, bound together by ecological relationships and interspecific interactions. Borneo and Sumatra host some of the richest assemblages of biota worldwide. The region, however, faces the highest global deforestation rates, which seriously threaten its unique biodiversity. We used a large camera trap dataset that recorded data for 70 terrestrial species of mammals and birds, to explore the drivers of regional species richness patterns. Using a multi-scale, multivariate modelling framework which quantified the main environmental factors associated with patterns of biodiversity, while simultaneously assessing individual relationships of each species, we determined the ecological drivers of sampled biodiversity, and their contributions to community assemblages. We then mapped predicted species richness, evaluated the effectiveness of protected areas in securing biodiversity hotspots, performed gap analysis to highlight biodiverse areas lacking protection, and compared our predictions with species richness maps produced by using IUCN range layers. Finally, we investigated the performance of each species as an indicator of sampled biodiversity. We demonstrate that biodiversity in Borneo and Sumatra is primarily affected by gradients of ecological and anthropogenic factors, and only marginally by topographic and spatial factors. In both islands, species are primarily associated with elevational gradients in vegetation and climate, leading to altitudinal zonation in niche separation as a major factor characterizing the islands' biodiversity. Species richness was highest in north-eastern Borneo and in western Sumatra. We found that most predicted biodiversity hotspots are not formally protected in either island; only 9.2% and 18.2% of the modelled species richness occurred within protected areas in Borneo and Sumatra, respectively. We highlighted that our prediction for Borneo performed better than, and differed drastically from, the IUCN species richness layer, while for Sumatra our modelled species richness layer and the IUCN one were similar, and both showed low predictive power. Our analysis suggests that common and

generalist carnivores are the most effective indicators of sampled biodiversity and have high potential as focal, umbrella or indicator species to assist multi-species vertebrate conservation planning. Understanding existing drivers and patterns of biodiversity is critical to support the development of effective community conservation strategies in this rapidly changing region.

## 1. Introduction

Species co-occurring within biological communities are affected individualistically by habitat factors and interactively by interspecific interactions (Hutchinson 1957). Nonetheless, the simultaneous distributions of multiple species can also be described by common factors (Davis et al. 2018, Haidir et al. 2018, Hearn et al. 2018b); co-occurrence patterns might reflect shared habitat requirements (Hearn et al. 2018b), be driven by biotic interactions (Ovaskainen et al. 2010), predator-prey dynamics (Coleman and Hill 2014) or be a consequence of anthropogenic pressures (Wang et al. 2015), and these factors might not operate in a mutually exclusive way (Macdonald et al. 2020). Global intensification of habitat loss necessitates a paradigmatic shift from single-species conservation and management, towards multi-species approaches (Macdonald et al. 2012, Macdonald et al. 2018a, Macdonald et al. 2020). Therefore, a fundamental ecological question, with profound practical implications, is how biological communities are assembled and what features control co-occurrence patterns.

This question is particularly relevant to the Sunda Islands of Borneo and Sumatra. Sundaland, as this region is known, is one of Earth's most important biodiversity hotspots, whose exceptional species richness includes 1,800 vertebrates, of which 701 are endemic (Myers et al. 2000). However, the biodiversity of Sundaland is seriously threatened by anthropogenic factors, including deforestation (Cushman et al. 2017) and habitat conversion (Wilcove et al. 2013), both directly linked with the accelerating economic growth of the region (Sodhi et al. 2004). Southeast Asia shows the highest rates of deforestation globally (Hansen et al. 2013), and primary vegetation in Sundaland occupies only 7.8% of its original extent (Myers et al. 2000). Additionally, deforestation in the region is expected to increase over the next decade (Cushman et al. 2017).

Our work expands upon previous studies carried out in smaller portions of the region, and focused mainly on the felid guilds (e.g., Haidir et al. (2018), Hearn et al. (2018b)), not only by assessing vaster spatial extents, but also by investigating the structure and spatial distribution of a larger, more comprehensive sample of species. A similar study was recently completed in mainland Southeast Asia (Macdonald et al. 2020) and a major focus of this paper is comparing the patterns and drivers of biodiversity in the Sunda Islands to those in mainland Southeast Asia, using comparable datasets and methods.

Macdonald et al. (2020) highlighted that, in mainland Southeast Asia, landscape compositional factors were the strongest drivers of species richness, and biodiversity was strongly affected by the extensiveness of natural vegetation, particularly forested areas, as well as anthropogenic factors like protected and developed areas. Additionally, these affected mainland Southeast Asian biodiversity at multiple spatial scales. Here, we applied a comparable multi-scale, multivariate framework (*sensu* McGarigal et al. (2016), as modified by Macdonald et al. (2020)) to assess the joint effect of landscape, anthropogenic, topographic and spatial factors on the simultaneous presence of sampled species to test the same hypotheses as in Macdonald et al. (2020) related to species co-occurrence and habitat selection in Borneo and Sumatra: (1) multi-species occurrence patterns are primarily influenced by environmental factors (e.g., forested areas, forest loss and protected areas) and only marginally by topographic and spatial factors and, (2) the aforementioned factors influence species co-occurrence at multiple spatial scales. Furthermore, Hughes (2017) and Macdonald et al. (2020) found that the modelled geographic distribution of terrestrial biodiversity in Southeast Asia drastically differs from the predictions made by using IUCN range layers; therefore we produced spatially explicit predictions of species richness in Borneo and Sumatra, expecting that (3) our species richness predictions for Borneo and Sumatra would differ from the IUCN maps, and would have higher predictive ability. In

addition, Macdonald et al. (2020) highlighted that the network of protected areas in mainland Southeast Asia preserved only a small portion of the modelled species richness; therefore, we performed gap analysis using the current protected areas network, hypothesising that (4) existing protected areas are relatively poor in location and insufficient in extent to protect the main biodiversity hotspots. Additionally, building upon the finding of Macdonald et al. (2020), we expect that (5) species richness predictions are primarily driven by common species and only marginally affected by rare species, and that (6) carnivores are good indicators of overall biodiversity, as demonstrated in other regions (Carroll et al. 2001, Dalerum et al. 2008, Macdonald et al. 2020).

## 2. Materials and methods

### 2.1.Data collection

From 2007 to 2016, our field teams conducted an extensive camera trap survey targeting wild felid populations across Borneo and Sumatra (Figure 1, Table 1). Cameras were set 1.0-2.0km apart, with two cameras per station set ~40cm above ground, and were deployed along natural ridgelines and disused logging roads allowing for greater visibility (Hearn et al. 2017, Macdonald et al. 2018a). The survey provided a rich dataset giving additional insights into meso and macro regional terrestrial biodiversity. All terrestrial mammals and birds whose species was unambiguously identified were selected for the analysis. When a species could not be clearly identified, we retained the data at a broader taxonomic level (i.e., order, family or genus). Additionally, to better assess the habitat factors driving species assemblages, we also included data related to anthropogenic disturbances captured by camera traps, such as people and domestic animals (i.e., domestic cats, domestic cattle and domestic dogs). For each species, we used the number of detections per camera trap station, reducing

overestimation bias by discounting records of the same species at the same camera station within one hour, except when animals were individually recognizable (e.g., based on the pattern of spots and stripes for felids) and when sex and/or age class were unambiguous.

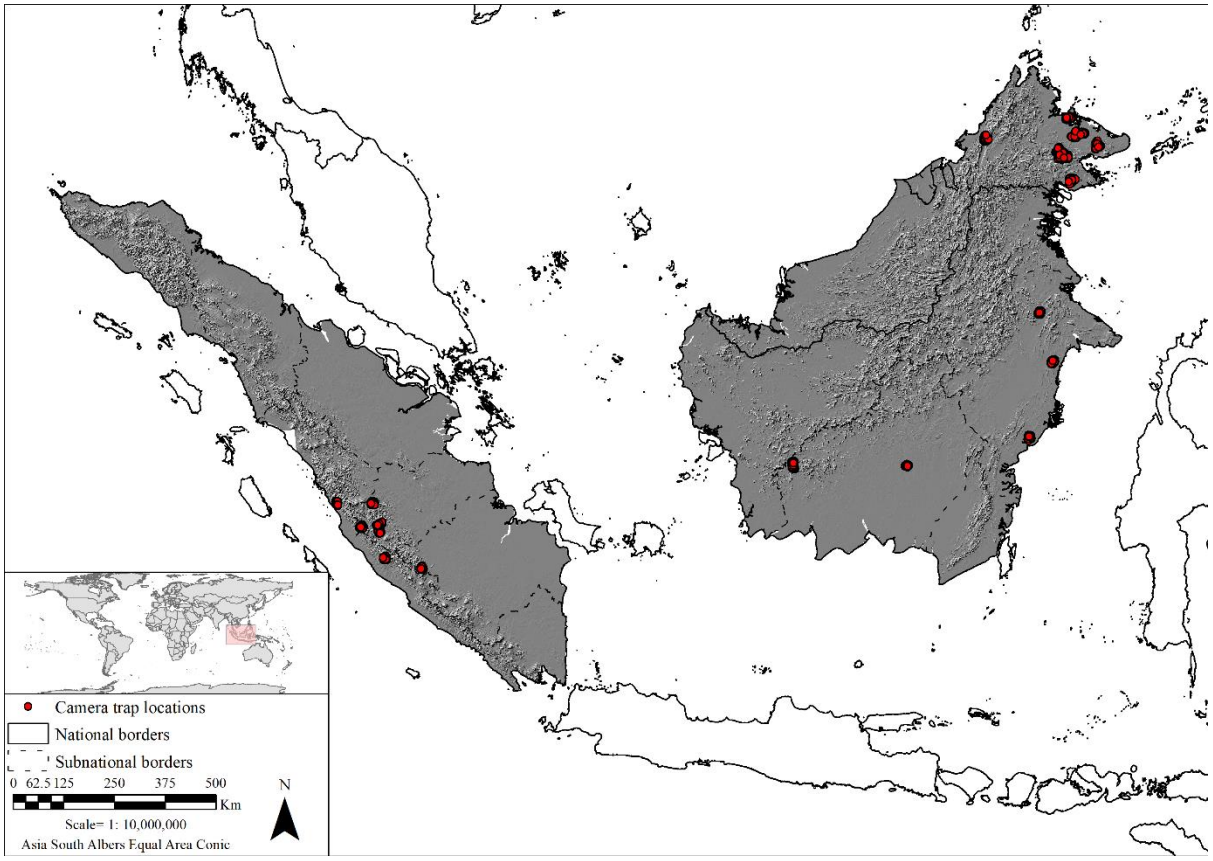



**Figure 1.** Map of the study area, including the islands of Borneo (right) and Sumatra (left), showing the camera trap locations. Each location contains one or more camera trap grids. Number and model of camera traps for each location are provided in Table 1.

**Table 1.** Sampling effort in Borneo and Sumatra. Shown are sampling locations per region, survey dates, extent of the surveyed areas, number of camera trap stations per sampling location, total number of trap nights and average number of trap nights per camera trap station, maximum and average number of species detected per sampling location, protected status of the sampling locations and camera trap models. Protected status acronyms: No (not protected), NR (Nature Reserve), PF (Production Forest), NP (National Park), Pl (plantation), CA (Conservation Area), WS (Wildlife Sanctuary), FR (Forest Reserve), WR (Wildlife Reserve). Camera trap model acronyms: CA (Cuddeback Ambush IR), Pa3 (Panthera V3), SS (Snapshot Sniper), Re5 (Reconyx HC 500), Bu (Bushnell), Re8 (Reconyx PC 800), Pa4 (Panthera V4).

| Regions | Sampling locations | Survey dates | Surveyed area ($Km^2$) | Camera trap stations | Total trap nights (mean) | Max species detected (mean) | Protected status | Camera trap model |
|---|---|---|---|---|---|---|---|---|
| Kalimantan | Bawan | 09/09/2012-26/11-2012 | 32.9 | 63 | 4,070 (64.60) | 4 (0.94) | No | CA |
| | Belantikan | 23/02/2014-17/06/2014 | 112.2 | 51 | 3,846 (75.41) | 8 (3.57) | NR/PF | CA |
| | Kutai | 12/12/2012-23/03/2013 | 88.9 | 52 | 3,269 (62.87) | 9 (3.37) | NP | CA |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Lesan | 15/07/2013-26/01/2014 | 70.8 | 73 | 10,057 (137.77) | 11 (3.03) | NR | CA |
| | Sungai Wain | 17/05/2012-08/08/2012 | 69.6 | 78 | 4,729 (60.63) | 7 (3.54) | NR | CA |
| Sabah | Crocker | 06/10/2011-07/02/2012 | 149.7 | 35 | 4,059 (115.97) | 31 (17.77) | NP | Pa3/SS/Re5/Bu |
| | Danum Palm | 15/03/2009-07/07/2009 | 7.8 | 23 | 2,214 (96.26) | 21 (8.00) | Pl | SS |
| | Danum Valley | 24/03/2012-06/10/2012 | 157.0 | 79 | 5,880 (74.43) | 25 (15.13) | CA | Pa3/Re5/Re8 |
| | IJM | 26/05/2011-18/08/2011 | 44.0 | 33 | 1,855 (56.21) | 15 (3.73) | Pl | Pa3/SS/Re5/Bu |
| | Kinabatangan | 24/07/2010-17/12/2010 | 359.5 | 68 | 4,450 (65.44) | 17 (7.81) | WS | SS/CA/Bu |
| | Malua | 09/07/2008-12/02/2009 | 102.8 | 38 | 3,867 (101.76) | 27 (14.26) | FR | SS |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Sepilok | 09/02/2011-25/05/2011 | 49.4 | 35 | 2,067 (59.06) | 17 (11.11) | FR | Pa3/SS/Re5/Bu |
| | Tabin | 18/09/2009-22/04/2010 | 144.3 | 74 | 6,877 (92.93) | 24 (15.53) | WR | SS/CA |
| | Tawau | 21/10/2012-30/12/2013 | 149.0 | 77 | 17,600 (228.57) | 33 (20.12) | NP | Re5/Re8 |
| | Ulu Segama | 24/05/2007-18/10/2007 | 60.1 | 22 | 2,847 (129.41) | 26 (18.27) | FR | SS |
| Sumatra | Bungo | 11/06/2014-11/11/2014 | 74.3 | 79 | 8,405 (106.39) | 14 (5.49) | NP | Pa4/CA |
| | Ipuh | 06/09/2015-24/12/2015 | 63.6 | 79 | 6,276 (79.44) | 17 (6.49) | NP | CA |
| | Kambang | 01/04/2015-08/07/2015 | 54.4 | 149 | 12,599 (84.56) | 9 (2.81) | NP | CA |
| | Linggau | 10/01/2016-24/03/2016 | 47.5 | 120 | 7,269 (60.58) | 11 (3.55) | NP | CA |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Muara Hemat | 17/10/2014-05/02/2015 | 72.0 | 164 | 13,466 (82.11) | 13 (4.83) | NP | CA |
| | RKE | 25/04/2015-10/08/2015 | 91.4 | 72 | 5,586 (77.58) | 18 (4.42) | NP | CA |
| | Sipurak | 24/11/2014-08/04/2015 | 80.0 | 80 | 7,227 (90.34) | 18 (6.25) | NP | CA |
| Tot | | 24/05/2007-24/03/2016 | 2,081.2 | 1,544 | 138,515 (89.71) | 33 (7.15) | | |

### 2.2.Habitat covariates

We selected a preliminary set of 35 covariates considered important for regional terrestrial biodiversity (Hearn et al. 2018b, Macdonald et al. 2018a), including nineteen landscape, four anthropogenic, four topographic and eight spatial covariates (Table 2). Additionally, we included camera effort as a predictor variable, defined as the cumulative number of nights in which camera traps were active. Similar to Macdonald et al. (2020), we kept the distinct covariates' grouping (i.e., environmental, defined as the combination of landscape and anthropogenic covariates, topographic, spatial and camera effort), to evaluate the contribution of each of them to sampled species detection counts.

Species-habitat relationships are known to be scale-dependent (Cushman and McGarigal 2002, 2004b, McGarigal et al. 2016). To investigate the scales at which sampled biodiversity responded to habitat features, we calculated each covariate at eight spatial scales, by averaging the pixel values of the original raster layers using circular buffers of 250m, 500m, 1,000m, 2,000m, 4,000m, 6,000m, 8,000m and 10,000m radii, centred on each camera trap location. Scales were chosen to cover a wide range of behavioural processes, describing different ways in which sampled species could interact with environmental features of the landscape, from fine-scale resource selection within home ranges, to broad-scale selection of home ranges (Simcharoen et al. 2014, Hearn et al. 2018a).

To produce more biologically informative variables, we calculated class- and landscape-level metrics in FRAGSTATS (McGarigal et al. 2012) for discrete layers, and focal statistics in ArcMap v10.6.1 (ESRI 2018) for continuous layers, using for both the software circular moving windows of radii corresponding to the aforementioned spatial scales. At the class level, defined as every patch type in the landscape (e.g., every class in a land-cover layer), we calculated PLAND, representing the percentage of the landscape occupied by a specific class,

and GYRATE_AM, providing a measure of landscape continuity. At the landscape level, defined as the combination of all classes in the patch mosaic (e.g., the entire land-cover layer, comprising all cover classes simultaneously), we computed aggregation index (AI), edge density (ED), contrast-weighted edge density (CWED), patch density (PD) and largest patch index (LPI). These landscape metrics were selected following past research showing they are particularly strong indicators of species occurrence patterns in complex landscapes (Grand et al. 2004, Chambers et al. 2016, Macdonald et al. 2020). For continuous layers we calculated focal mean (FM) and standard deviation (SD). These resulted in a total of 67 covariates (Table 2).

**Table 2.** Preliminary set of covariates and corresponding representative scales for biodiversity in Borneo and Sumatra. Additionally, the description of the primary set of covariates selected for Borneo and Sumatra biodiversity models is provided. For the analyses, landscape and anthropogenic covariates were merged into the macro-group of environmental covariates. We calculated class- and landscape-level metrics in FRAGSTATS (McGarigal et al. 2012) for discrete layers, and focal statistics in ArcMap v10.6.1 (ESRI 2018) for continuous layers. PLAND, represents the percentage of the landscape occupied by a specific class, GYRATE_AM is a measure of landscape continuity. At the landscape level, we computed aggregation index (AI), edge density (ED), contrast-weighted edge density (CWED), patch density (PD) and largest patch index (LPI). For continuous layers we calculated focal mean (FM) and standard deviation (SD). All raster layers were first re-projected to Asia South Albers Equal Area Conic projection in ArcMap v10.6.1, by applying nearest neighbour re-sampling technique for discrete layers and bilinear interpolation re-sampling technique for continuous layers. Then, by applying the same re-sampling techniques, all raster layers were re-sampled to 250m resolution.

| | | | | Borneo | Sumatra |
|---|---|---|---|---|---|
| Class | Covariate | Source | Metric | Best scale (m) | Best scale (m) |
| Landscape | Landcover | Miettinen et al. (2012) | AI | 4,000⁑ | 4,000⁑ |
| | | | CWED | 2,000⁑ | 4,000⁑ |
| | | | ED | 2,000⁑ | 4,000⁑ |

| | | | | |
|---|---|---|---|---|
| | | PD | 4,000⁑ | 2,000⁑ |
| Lowland forests | Miettinen et al. (2012) | PLAND | 4,000⁑ | 8,000⁑ |
| | | GYRATE_AM | 8,000 | 10,000 |
| Lowland mosaic areas | Miettinen et al. (2012) | PLAND | 4,000⁑ | 6,000⁑ |
| | | GYRATE_AM | 1,000⁑ | 6,000⁑ |
| Lower-montane forests | Miettinen et al. (2012) | PLAND | 6,000⁑ | 4,000⁑ |
| | | GYRATE_AM | 6,000⁑ | 2,000⁑ |
| Lowland open areas | Miettinen et al. (2012) | PLAND | 2,000⁑ | 6,000† |
| | | GYRATE_AM | 2,000 | 8,000 |
| Mangroves | Miettinen et al. (2012) | PLAND | 10,000 | NA‡ |
| | | GYRATE_AM | 4,000⁑ | NA‡ |
| Montane mosaic areas | Miettinen et al. (2012) | PLAND | NA‡ | 2,000⁑ |
| | | GYRATE_AM | NA‡ | 2,000 |
| Montane open areas | Miettinen et al. (2012) | PLAND | NA‡ | 8,000⁑ |
| | | GYRATE_AM | NA‡ | 8,000† |
| Palm plantations | Miettinen et al. (2012) | PLAND | 2,000 | 10,000⁑ |

| | | | | |
|---|---|---|---|---|
| | | GYRATE_AM | 4,000⁑ | 10,000 |
| Peat swamp forests | Miettinen et al. (2012) | PLAND | NA‡ | NA‡ |
| | | GYRATE_AM | NA‡ | NA‡ |
| Plantation and regrowth | Miettinen et al. (2012) | PLAND | 10,000⁑ | 2,000⁑ |
| | | GYRATE_AM | 500⁑ | 1,000 |
| Upper-montane forests | Miettinen et al. (2012) | PLAND | NA‡ | 4,000 |
| | | GYRATE_AM | NA‡ | 4,000⁑ |
| Urban areas | Miettinen et al. (2012) | PLAND | NA‡ | NA‡ |
| | | GYRATE_AM | NA‡ | NA‡ |
| Water | Miettinen et al. (2012) | PLAND | 6,000⁑ | NA‡ |
| | | GYRATE_AM | 4,000† | NA‡ |
| Continuous tree canopy cover | Hansen et al. (2013) | FM | 4,000⁑ | 4,000⁑ |
| | | SD | 2,000⁑ | 4,000⁑ |
| Tree canopy cover reclassified | Hansen et al. (2013) | AI | 2,000⁑ | 4,000⁑ |
| | | CWED | 2,000⁑ | 4,000⁑ |
| | | ED | 2,000⁑ | 4,000⁑ |

| | | | | | |
|---|---|---|---|---|---|
| | | | LPI | 4,000‡ | 4,000‡ |
| | | | PD | 4,000 | 4,000‡ |
| | Non-forested areas (0-20%) | Hansen et al. (2013) | PLAND | 4,000‡ | 2,000‡ |
| | | | GYRATE_AM | 1,000‡ | 2,000‡ |
| | Open forests (20-40%) | Hansen et al. (2013) | PLAND | 6,000‡ | 6,000‡ |
| | | | GYRATE_AM | 2,000† | 4,000† |
| | Closed forests (>40%) | Hansen et al. (2013) | PLAND | 4,000‡ | 4,000‡ |
| | | | GYRATE_AM | 6,000‡ | 6,000† |
| Anthropogenic | Human footprint | WCS and CIESIN (2005) | FM | 10,000 | 10,000‡ |
| | | | SD | 2,000 | 10,000 |
| | Human population density | CIESIN and CIAT (2016) | FM | 6,000‡ | 10,000‡ |
| | | | SD | 6,000† | 8,000† |
| | Forest loss | Hansen et al. (2013) | PLAND | 10,000 | 4,000 |
| | | | GYRATE_AM | 8,000‡ | 2,000‡ |
| | Protected areas | UNEP-WCMC and IUCN (2022) | PLAND | 4,000 | 10,000‡ |
| | | | GYRATE_AM | 4,000‡ | 8,000 |

| | | | | | |
|---|---|---|---|---|---|
| Topographic | Elevation | Jarvis et al. (2008) | FM | 250 | 250 |
| | | | SD | 8,000⁑ | 10,000⁑ |
| | Roughness | Jarvis et al. (2008), Evans et al. (2014) | FM | 4,000⁑ | 10,000⁑ |
| | | | SD | 6,000⁑ | 10,000 |
| | Slope position | Jarvis et al. (2008), Evans et al. (2014) | FM | 8,000 | 10,000 |
| | | | SD | 4,000⁑ | 10,000⁑ |
| | Compound topographic index | Jarvis et al. (2008), Evans et al. (2014) | FM | 2,000⁑ | 500 |
| | | | SD | 4,000 | 4,000 |
| Spatial | X | NA | NA | NA | NA |
| | Y | NA | NA | NA⁑ | NA⁑ |
| | X*Y | NA | NA | NA⁑ | NA⁑ |
| | $X^2$ | NA | NA | NA⁑ | NA⁑ |
| | $Y^2$ | NA | NA | NA⁑ | NA⁑ |
| | $X^2$*Y | NA | NA | NA⁑ | NA⁑ |
| | X*$Y^2$ | NA | NA | NA⁑ | NA⁑ |
| | $X^2$*$Y^2$ | NA | NA | NA | NA⁑ |

| | | | | | |
|---|---|---|---|---|---|
| Camera effort | Camera trap nights | NA | NA | NA | NA |

‡ Covariates excluded from the model because occurring in <10% of camera trap stations.

⁜ Covariates excluded from the model after the multicollinearity analysis.

† Covariates excluded from the model after the forward selection analysis.

### 2.3.Scale optimisation

Since interspecific relationships among species assemblages are different between Borneo and Sumatra, we implemented the modelling workflow previously applied by Macdonald et al. (2020), and described in the next paragraphs, independently for the two islands. We removed all poorly sampled covariates (i.e., those occurring at <10% of camera stations) to avoid assessing unrepresentative habitat features, since zero-inflation of explanatory variables can negatively affect parameter estimation (Martin et al. 2005).

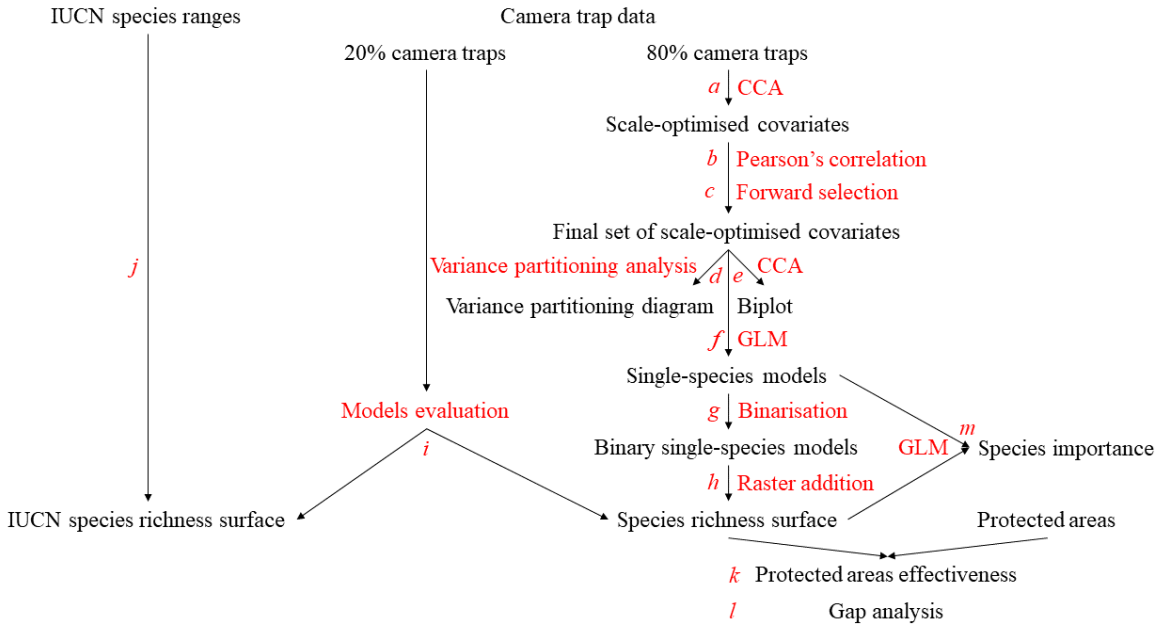



**Figure 2.** Workflow diagram of the methods applied to map the hotspots of terrestrial vertebrate biodiversity in Borneo and Sumatra using camera trap species detection data. Using 80% of camera traps, we first performed canonical correspondence analysis (CCA) independently for each covariate, for each scale, using the number of detections of the sampled species as response variable, to assess the most representative spatial scales for sampled biodiversity (*a*). Then, we calculated Pearson's correlation between all remaining

covariates, and removed those highly correlated at $|r| \geq 0.7$ (*b*) and we obtained the final set of covariates by applying forward selection, retaining covariates at $p < 0.001$ (*c*). We performed variance partitioning to evaluate the variance explained by different groups of covariates (*d*), and performed CCA between the final set of scale-optimised covariates and the number of detections of the sampled species, producing biplots to evaluate relationships of sampled species with covariates (*e*). We then produced single-species generalised linear models (GLMs), independently for each species and predicted these across both islands (*f*), converted the predictions to binary by considering pixels with values <1 as absences, and as presences otherwise (*g*), and produced a species richness layer by adding the single-species binary models (*h*). We evaluated the performance of the species richness surface by using the 20% of camera traps held out for validation and calculating *Nagelkerke-pseudo-*$R^2$ (*i*). We also produced a layer of species richness by adding IUCN range layers for the sampled species (*j*), and evaluated the IUCN species richness surface by using the same method used to evaluate the modelled one (*i*). Finally, we evaluated the effectiveness of the current system of protected areas to preserve modelled species richness (*k*), and performed gap analysis evaluating the most biodiverse areas lacking formal protection (*l*). Finally, we evaluated the importance of each species as an indicator to represent the distribution of the overall sampled biodiversity (*m*).

To identify representative scales for sampled biodiversity, we performed Canonical Correspondence Analysis (CCA; Figure 2 (*a*)) (ter Braak 1986, ter Braak and Prentice 1988, McGarigal et al. 2000) on randomly sampled 80% of the camera trap stations (we used the remaining 20% to evaluate the ability of the species richness surface to correctly predict observed richness – see below), independently at each scale for each covariate, using the *cca* function in the *vegan* package (Oksanen et al. 2018) in R v3.5.1 (R Core Team 2018). CCA is

a multivariate modelling technique that accounts for unimodal distribution of response variables, which are expected in species presence patterns across broad environmental gradients (Austin et al. 1994). CCA allowed us to model species-habitat relationships simultaneously for all sampled species and, at the same time, to account for the individual contribution of each species to the modelled relationships (Macdonald et al. 2020). For each covariate, at each scale, we performed CCA using the number of detections of the sampled species as response variables, and we identified the optimal scale by selecting the scale whose CCA showed the highest canonical *eigenvalue* (Borcard et al. 1992). We then checked multicollinearity by calculating Pearson's correlation coefficient between all pairs of scale-optimised covariates (Figure 2 (*b*)). When two covariates were highly correlated ($|r| \geq 0.7$), we removed the covariate whose univariate CCA had the lower adjusted-$R^2$ (Guisan and Zimmermann 2000). The final set of covariates was selected by performing forward selection on the previously selected 80% of camera trap stations, using the *ordistep* function in the *vegan* package (Figure 2 (*c*)), independently on each group of covariates (i.e., environmental, topographic and spatial) and, within the multivariate context of the CCA, simultaneously for all sampled species, retaining only those covariates significant at $p < 0.001$ within each group (Cushman and McGarigal 2002, 2004a).

We performed variance partitioning analysis (Figure 2 (*d*)) to investigate the variance explained in the multivariate detection counts of sampled species by each group of covariates (individual contribution), as well as by different combinations of groups of covariates (joint contribution) (Borcard et al. 1992, Cushman and McGarigal 2004a, b), by using the number of detections of the sampled species as response variables and applying the *varpart* function in the *vegan* package.

### 2.4.Species richness and model performance

We performed CCA between the number of detections of the sampled species at the previously selected 80% of camera trap stations and the final set of scale-optimised covariates (Figure 2 (*e*)) to assess their contribution to the sampled species detections, and interpreted the biplot to evaluate the interactions of each species with specific habitat factors (Macdonald et al. 2020). CCA biplots are graphical representations of the multivariate results that simultaneously show both the sampled species and the influence of the covariates, which are shown as vectors pointing in the direction of increasing value and with length proportional to their influence on predicted species composition and detection. The vectors' lengths (i.e., the influence of the covariates on the species) are given by the *eigenvalues* calculated in the CCA. Species are represented in the biplot by dots, and distances from the origin along the dot-product vector of each covariate are proportional to the correlation between that species' number of detections and the value of that predictor variable.

To predict species richness, we first performed single-species generalised linear models (GLMs) by using the package *stats* in R v3.5.1, independently for each species (Figure 2 (*f*), Tables S1 and S2), using the standardised final set of scale-optimised covariates from the CCA as predictor variables, and accommodating the continuous discrete response variables (i.e., species' detections) by adopting a quasi-Poisson distribution (Vincent and Haworth 1983). We opted for a quasi-Poisson distribution given that the data for several species were overdispersed and did not meet the Poisson distribution's assumption of equal mean and variance (Figures S1 and S2).

We then used the parameter estimates from the single-species GLMs in combination with the covariate layers (Table 2) for Borneo and Sumatra to predict species counts across the entire extent of both islands. The single-species predictions were then converted to binary

form, treating <1 value pixels as absences, and ≥1 value pixels as presences (Figure 2 (*g*)). Finally, the single-species binary surfaces were combined into a species richness layer by addition (Figure 2 (*h*); Grand et al. (2004), Macdonald et al. (2020)).

Since the camera trap surveys did not sample the full gradients of all the chosen predictor variables over the entire study area (e.g., most of the camera trap locations occurred within protected areas, surveys did not cover the most remote areas occurring at high elevations and camera traps were mainly deployed in easily accessible areas like along disused logging roads; Figure 1), to avoid producing unreliable models by making predictions on areas that were ecologically too different from the sampling locations, we masked out the most different areas by producing a Mahalanobis distance surface. The Mahalanobis distance is the distance between two points in a multivariate space, that we defined by using the layers of the environmental factors that most likely determined a geographically biased camera trap survey: tree canopy cover, protected areas, human footprint, elevation, roughness, annual maximum temperature, annual minimum temperature and annual precipitations. The climatic covariates were derived from PRISM Climate Group (2016), while the landscape, anthropogenic and topographic covariates' sources were the same as the ones used to model species richness (Table 2). The analysis, performed by using the package *dismo* (Hijmans et al. 2017) in R v3.5.1, provided a continuous surface of the distances in the multidimensional space of each pixel from those where the camera traps were deployed. We then binarised the continuous surface by calculating the $20^{th}$ percentile of the distribution of the Mahalanobis distances and excluded all pixels with distance values equal to or higher than that value from our predictive surfaces.

We produced the single-species models by using the previously randomly selected 80% of camera trap stations, keeping the remaining 20% to evaluate the ability of the species richness surface to predict observed richness by adopting the method applied by Macdonald

et al. (2020): we computed the number of species predicted by our species richness layer at the camera trap stations used for validation, and performed GLM with Poisson distribution between the predicted and the empirical number of species, and then calculated the *Nagelkerke-pseudo-$R^2$* (Nagelkerke (1991); Figure 2 (*i*)). The *Nagelkerke-pseudo-$R^2$* is an index ranging from 0 to 1 that can be used to compare the performances of different models. Additionally, we assessed the performances of the single-species GLMs by calculating their residual deviances, which indicate how well the response variables (i.e., species' detections) are predicted when the predictor variables are included in the models. The lower the residual deviance, the better the single-species GLM explains the species detection data.

We compared the multi-species predictive surface with a species richness surface produced by combining the IUCN geographic range layers of the sampled species (Figure 2 (*j*)) (IUCN 2019) in ArcMap v10.6.1. To produce the IUCN surface, we used the polygons of the extent of occurrence in which the species were considered extant and resident. We evaluated the performance of the IUCN surface by calculating the *Nagelkerke-pseudo-$R^2$* between the empirical number of species at the camera stations held out for validation and the number of species predicted by the IUCN surface (Figure 2 (*i*)).

### 2.5.Protected areas effectiveness

We assessed the effectiveness of the current system of protected areas network by calculating the ratio between the cumulative number of species predicted within protected areas as defined by the World Database on Protected Areas (WDPA; UNEP-WCMC and IUCN (2022)) relative to the cumulative number of species predicted in the entire islands (Grand et al. 2004, Macdonald et al. 2020), independently for Borneo and Sumatra. Hence,

we evaluated the performances of each protected area in preserving predicted species richness (Figure 2 (*k*)).

### 2.6.Gap analysis

Additionally, we reclassified the species richness surface into low, medium and high priority areas for conservation, based on the percentiles distribution (>50$^{th}$, >70$^{th}$ and >90$^{th}$, respectively (Macdonald et al. 2018a)). Then, we identified the most important biodiversity hotspots not formally protected by calculating the ratio between the cumulative number of species predicted in the most biodiverse areas (>90$^{th}$ percentile) relative to the cumulative number of species predicted in the entire islands, independently for Borneo and Sumatra (Figure 2 (*l*)).

### 2.7.Indicator species effectiveness

To assess the importance of each species as an indicator of sampled biodiversity, we randomly sampled 1% of the pixels from the multi-species surface and from each single-species presence-absence surface and then performed GLMs independently for each species, in which the predictor variable was represented by the modelled number of species, while the response variables were represented by each species' presence-absence (Figure 2 (*m*)). The lower the residual deviance of the model, the more important that species was for explaining overall terrestrial vertebrate biodiversity (Macdonald et al. 2020).

## 3. Results

### 3.1.Data collection

Camera traps were set in 15 sampling locations in Borneo (5 in the Indonesian provinces of Kalimantan and 10 in the Malaysian state of Sabah) and in 7 sampling locations in Sumatra, each covering an average area of 91.63km$^2$ (±14.97 SE) and yielding a total of 1,544 camera trap stations (801 in Borneo, of which 317 in Kalimantan and 484 in Sabah, and 743 in Sumatra). We achieved a combined sampling effort of 138,515 trap nights, with 77,687 trap nights in Borneo (mean= 96.99±2.12 SE) and 60,828 in Sumatra (mean= 81.87±0.72 SE; Table 1).

Species were recorded at all 22 sampling locations, including 747 camera trap stations in Borneo (93.26%) and 662 in Sumatra (89.10%). The 54 camera trap stations that failed to detect any species in Borneo were all in Kalimantan. The maximum number of species detected at a single camera trap station was 11 in Kalimantan, 33 in Sabah and 18 in Sumatra (Table 1). Overall, we sampled 58 species, including 47 mammals and 11 terrestrial birds. Additionally, we sampled 10 mammalian and 2 avian groups for which we retained data at broader taxonomic levels. Not all the species were detected in the two islands; in Borneo we sampled 40 mammals and 8 terrestrial birds, representing respectively the 18.0% and 1.9% of the total number of mammal and birds occurring on the island, while in Sumatra we sampled 26 mammals and 5 terrestrial birds, representing respectively the 12.9% and 0.9% of the total number of mammals and birds occurring on the island. Total detections for sampled species are reported in Table S3. Moreover, we detected 4 types of anthropogenic disturbance: domestic cats, domestic cattle, domestic dogs and humans (Table S3). These anthropogenic elements were included only in the multivariate analyses as part of the response variable alongside the sampled species, to assess the most representative spatial scales, the covariates to be included in the final models and to produce the biplots. We included these

anthropogenic elements in order to evaluate their contribution to the species-habitat relationships, as well as to assess their multivariate relationships with sampled species, but they were not used to produce the GLMs and the species richness surface, because we only wanted to predict wild species richness patterns.

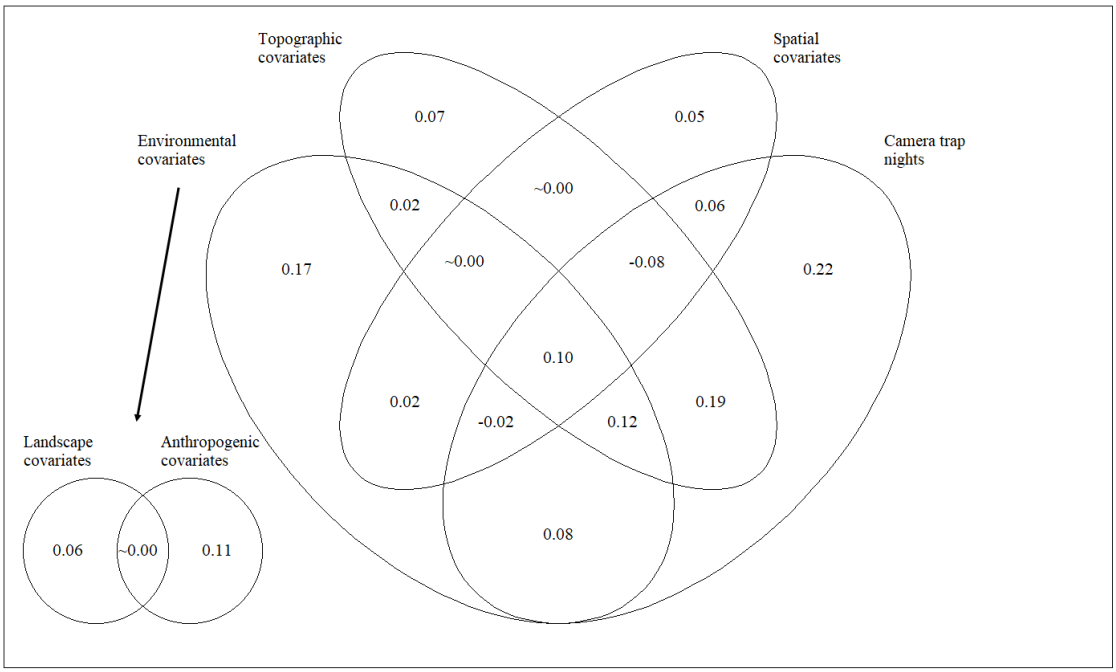
### 3.2.Scale optimisation

Ten covariates in Borneo and eight in Sumatra occurred at <10% of camera trap stations and were excluded due to lack of predictive power. Scale optimisation revealed that, on average, vertebrates were associated with medium (2-6km) and broad (8-10km) scale patterns. Four covariates in Borneo (8.16%) and three in Sumatra (5.88%) were identified as optimal at fine scales (≤1km). Thirty-seven covariates in Borneo (75.51%) and thirty in Sumatra (58.82%) were selected at medium scales, while eight covariates in Borneo (16.33%) and eighteen in Sumatra (35.29%) were selected at broad scales (Table 2).

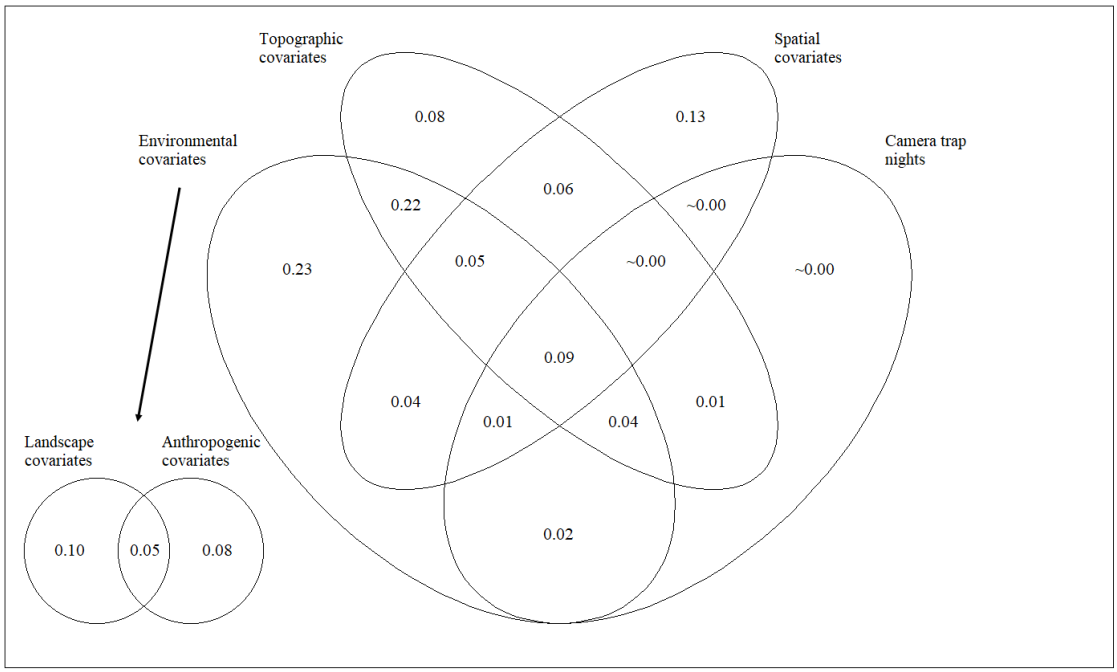
We excluded forty highly correlated covariates from Borneo and thirty-nine from Sumatra. Following forward selection, we removed three covariates from Borneo and five from Sumatra, obtaining a final set of fifteen and sixteen covariates, respectively (Table 2).

The final sets of covariates explained 23.9% and 10.4% of absolute variance of the detection data in Borneo and Sumatra, respectively. Variance partitioning revealed that, in Borneo, species detection was most strongly predicted by camera effort (22% individual contribution of the single group of covariates and 67% joint contribution of the group of covariates with other groups of covariates), followed by environmental covariates (17% individually and 49% jointly). Variance explained by environmental covariates was in turn shared between landscape covariates (6% both individually and jointly) and anthropogenic covariates (11% both individually and jointly). Topographic and spatial covariates'

independent contributions were comparatively lower (7% and 5%, respectively), but their shared contributions were more substantial (42% and 13%, respectively; Figure 3a). In Sumatra, environmental covariates explained the greatest proportion of relative variance (23% individually and 70% jointly). Variance explained by environmental covariates was shared between landscape covariates (10% individually and 15% jointly) and anthropogenic covariates (8% individually and 13% jointly). Comparatively, topographic and spatial covariates explained less variance (8% individually and 55% jointly, and 13% individually and 38% jointly, respectively; Figure 3b).




**(a)**




**(b)**

**Figure 3.** Variance partitioning analyses for Borneo (a) and for Sumatra (b), showing the proportion of variance in the detections of the sampled species explained in the Canonical Correspondence Analyses by different subsets of covariates. Shown are the relative values. Absolute variance explained for Borneo= 0.239. Absolute residual variance for Borneo= 0.761. Absolute variance explained for Sumatra= 0.104. Absolute residual variance for Sumatra= 0.896.

### 3.3.Species richness and model performance

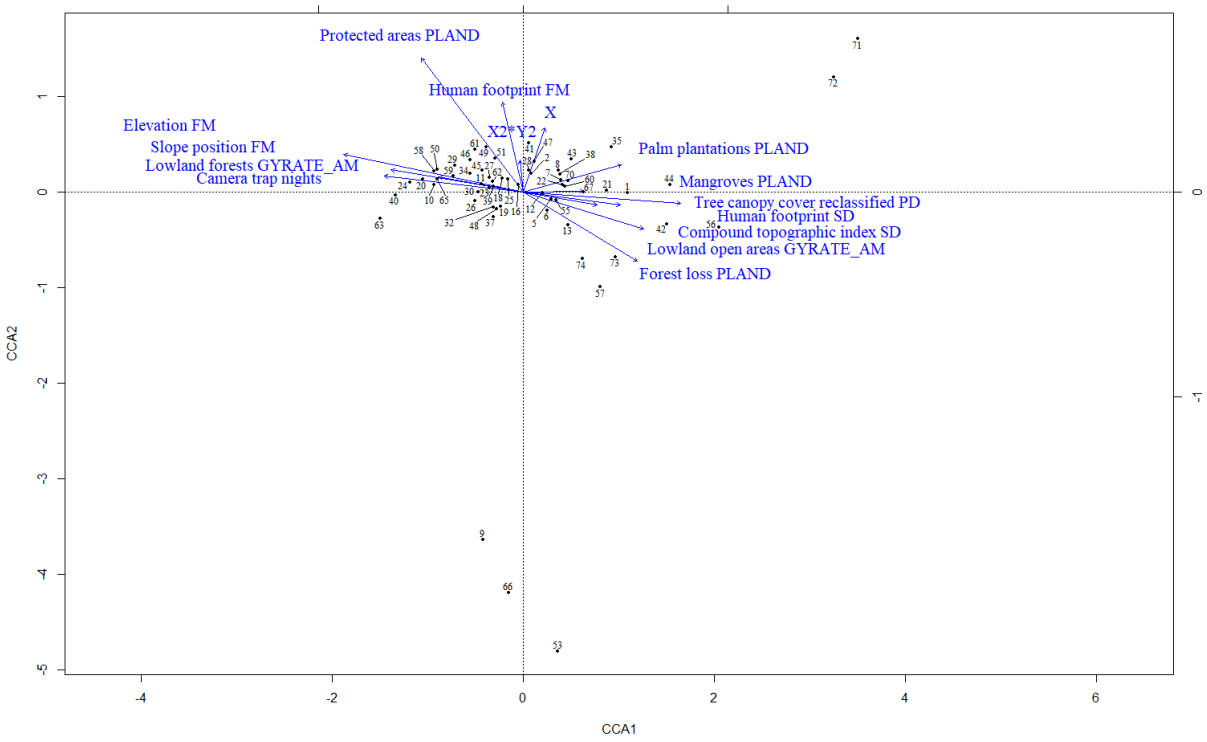
CCA performed between the number of detections of the sampled species and the final set of scale-optimised covariates confirmed that species-habitat relationships were primarily affected by environmental and topographic covariates. In Borneo, elevation, patch density of

tree cover, slope position and lowland open areas, were the main factors positively affecting species detections. However, anthropogenic factors, like protected areas and forest loss, were also influential in affecting species detections (Figure 4a, Table S4). Similarly, in Sumatra species detection patterns were most strongly associated with environmental factors like upper-montane forests, plantation and regrowth and lowland open areas, but also with topographic (elevation), anthropogenic (forest loss) and spatial (longitude) covariates (Figure 4b, Table S5).

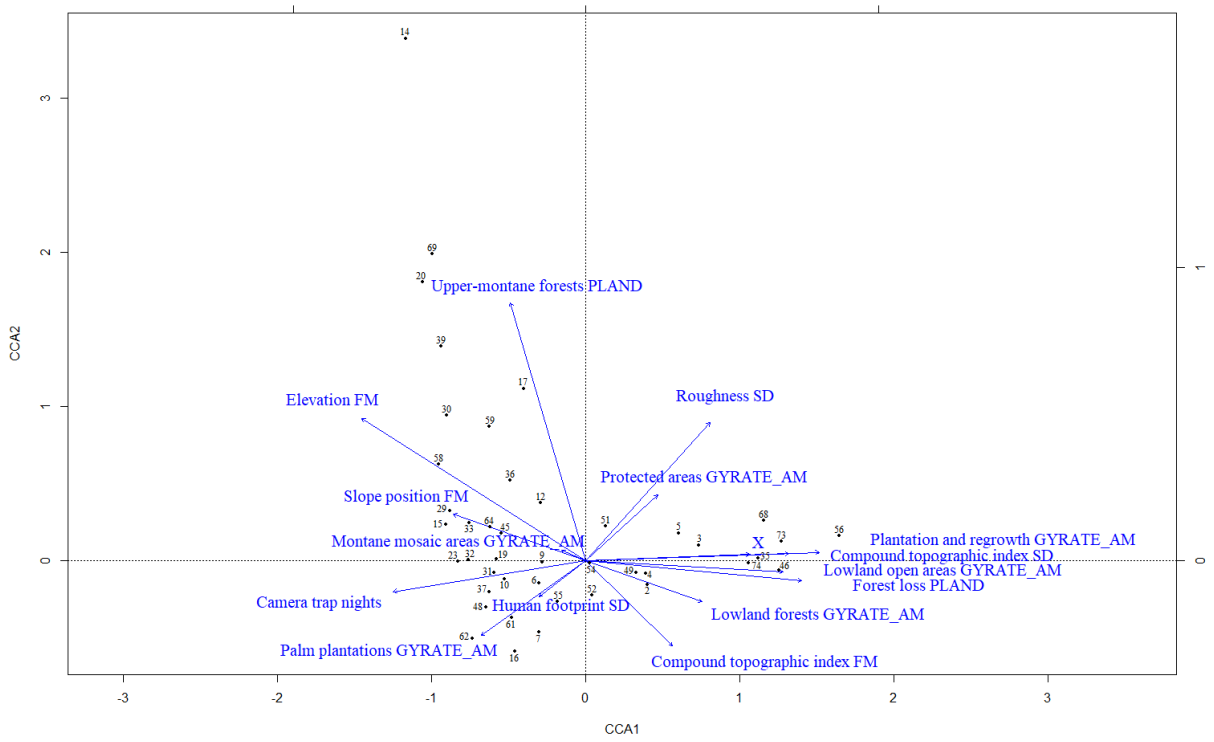
Analysis of the CCA biplots confirmed the positive association of many species with forests and elevation gradients. In Borneo, the majority of species appeared to be relative habitat generalists, as revealed by clustering at the centre of the biplot (Figure 4a). Three carnivore species (Hose's civet (*Diplogale hosei*), Bornean ferret badger (*Melogale everetti*) and masked palm civet (*Paguma larvata*)) showed strong affinity for high elevations and mountainsides (as indicated by slope position). Endangered flagship species were positively associated with anthropogenic factors: Asian elephant (*Elephas maximus*) was associated with palm plantations, while banteng (*Bos javanicus*) and Bornean orangutan (*Pongo pygmaeus*) occurred in strong association with forest loss.

In Sumatra, elevation was an important driver of many species including, from the most associated with high areas to the least, mainland serow (*Capricornis sumatraensis*), Salvadori's pheasant (*Lophura inornata*), masked palm civet (*Paguma larvata*), yellow-throated marten (*Martes flavigula*), civet species, Sunda clouded leopard (*Neofelis diardi*), partridge and pheasant species, dhole (*Cuon alpinus*) and sambar (*Rusa unicolor*; Figure 4b). Generalist species, including Sunda leopard cat (*Prionailurus javanensis*), Malay crested fireback (*Lophura rufa*), long-tailed porcupine (*Trichys fasciculata*) and common long-tailed macaque (*Macaca fascicularis*), were positively associated with lowland open areas and forest loss. Additionally, the critically endangered Sunda pangolin (*Manis javanica*) was

positively associated with palm plantations and human footprint. Overall, fewer species were clustered in the centre of the biplot, revealing a greater degree of specialisation compared to Borneo.




**(a)**




**(b)**

**Figure 4.** Biplots resulting from the Canonical Correspondence Analyses between the number of detections of the sampled species and the final sets of scale-optimised covariates, illustrating the relationships between species sampled in Borneo (a) and in Sumatra (b), and habitat covariates. See Table 2 for covariate description and Table S3 for species codes.

Combining by addition the predicted occurrence surfaces produced for each sampled species based on the results of the single-species GLMs (Tables S1 and S2), we predicted species richness across Borneo and Sumatra (Figure 5), compared it to the IUCN surfaces (Figure 6) and evaluated the explanatory and predictive power of the models. For Borneo, analysis of the residual deviances of the single-species models revealed that they had strong explanatory power, even though the models for a few species revealed high residual deviances (mean residual deviance= 1,365.0, median residual deviance= 767.8; Figure S3).

Additionally, our species richness surface showed strong positive relationships with empirical species richness based on the independent holdout validation data (*Nagelkerke-pseudo-$R^2$*= 0.72, $p < 0.0001$), while the IUCN surface performed poorly (*Nagelkerke-pseudo-$R^2$*= 0.06, $p = 0.003$), and showed a contrasting geographic distribution of biodiversity (Figure 6).

For Sumatra, the residual deviances revealed strong explanatory power of the single-species models (mean residual deviance= 564.6, median residual deviance= 251.1; Figure S4). However, our species richness surface and the IUCN prediction were much more similar with respect to the geographic distribution of biodiversity hotspots than in Borneo (Figure 6). Importantly, in Sumatra neither our species richness surface nor the IUCN prediction were highly predictive of observed species richness based on the holdout validation data (*Nagelkerke-pseudo-$R^2$*= 0.007, $p = 0.34$ for our species richness surface and *Nagelkerke-pseudo-$R^2$*= 0.002, $p = 0.59$ for the IUCN prediction).

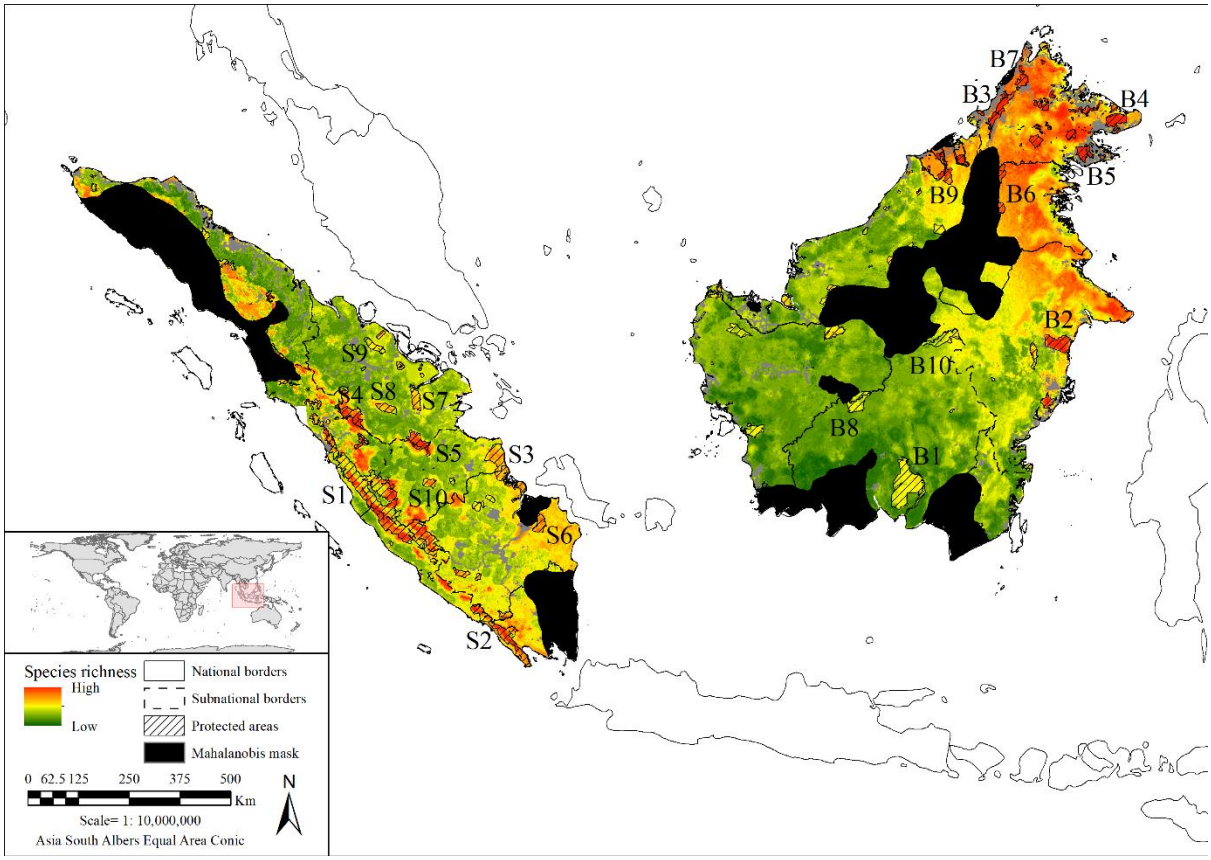



**Figure 5.** Predicted species richness in Borneo (right) and Sumatra (left), showing protected areas overlaid in black hash marks. Species richness predictions were based on single-species generalised linear models between the sampled species' number of detections and the scale-optimised covariates found important to describe overall sampled biodiversity geographic patterns in a Canonical Correspondence Analysis. Grey areas represent high human footprint. Black areas represent the most distinct regions from the sampled locations, as highlighted by a Mahalanobis distance surface, which were excluded from predictions. Numbers represent codes of the main systems of protected areas, as reported in Tables S6 and S7.

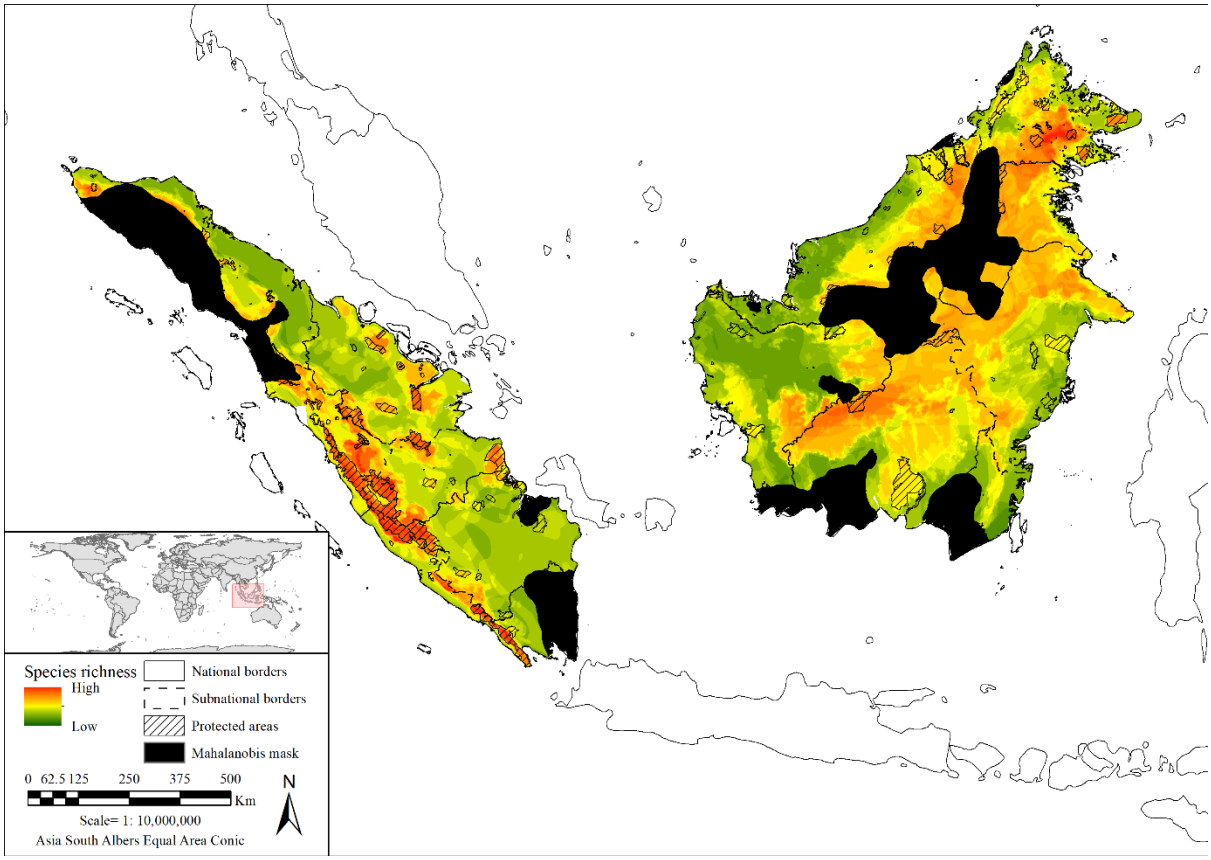



**Figure 6.** Predicted species richness in Borneo (right) and Sumatra (left) as obtained by summing the IUCN geographic range layers of the sampled species. Protected areas are shown in black hash marks. Black areas represent the most distinct regions from the sampled locations, as highlighted by a Mahalanobis distance surface, which were excluded from the predictions.

### 3.4.Protected areas effectiveness

The evaluation of the effectiveness of protected areas revealed that they preserved a relatively small proportion of biodiversity; only 9.2% and 18.2% of the cumulative modelled species richness occurred within protected areas in Borneo and Sumatra, respectively. In both islands, a small number of protected areas provided for most of the protected biodiversity. In Borneo, Sebangau, Kutai and Crocker Range National Parks represented the most important

protected areas, harbouring 1.20%, 0.82% and 0.63% of the overall biodiversity, and 12.94%, 8.91% and 6.84% of the protected biodiversity, respectively (Table S6). The aforementioned most important protected areas for Bornean biodiversity occurred in Kalimantan, the Indonesian portion of Borneo, and in Sabah, in the Malaysian side of the island. In Sumatra, Kerinci Seblat, Bukit Barisan Selatan and Berbak Sembilang National Parks protected 8.17%, 1.99% and 1.44% of the overall biodiversity, and 44.98%, 10.95% and 7.91% of the protected biodiversity, respectively (Table S7).

### 3.5.Gap analysis

The reclassification of the predictive surfaces into low, medium and high biodiverse areas allowed us to highlight the most critical areas for conservation (Figure 7). In Borneo, the Malaysian state of Sabah and the Indonesian provinces of North and East Kalimantan were the regions where most of the biodiversity hotspots occurred. However, most of them fell outside the current system of protected areas (Figure 7, Table S8). In Sabah, a critically important biodiverse area of more than 5,500km$^2$ fell entirely outside protected areas (Table S8). In North and East Kalimantan, at least five distinguished biodiversity hotspots have been highlighted, and none of them was protected (Figure 7, Table S8).

In Sumatra, the most important predicted biodiversity hotspots coincided with protected areas in the western mountainous regions. Our results showed that in Sumatra there seems to have been a more severe pattern of habitat loss which resulted in biodiversity being more associated with, and constrained by, protected areas than it is in Borneo. However, we also highlighted important hotspots outside protected areas (Figure 7, Table S8). Of these, the most important ones surrounded and connected existing protected areas, like the satellite hotspots connecting the Kerinci Seblat National Park and the Bukit Barisan Selatan National

Park. Additionally, extensive areas of medium and low biodiversity occurred in the southern parts of the island, and these currently almost completely lack protection.

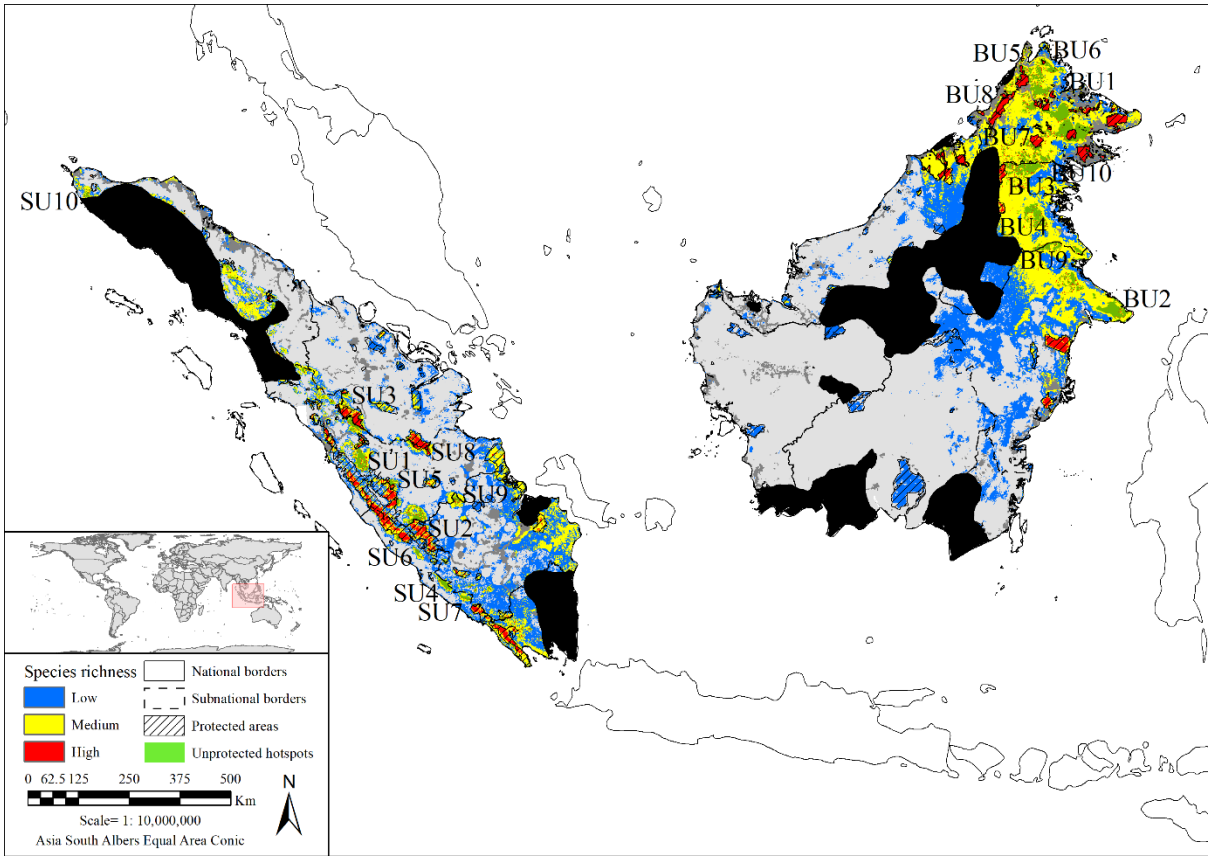



**Figure 7.** Predicted species richness in Borneo (right) and Sumatra (left), derived from the continuous species richness surface shown and described in Figure 5, and reclassified into low, medium and high priority areas based on the percentiles distribution of the species richness values (>50$^{th}$, >70$^{th}$ and >90$^{th}$, respectively). Highlighted are the main biodiversity hotspots falling outside the current system of protected areas. Protected areas are overlaid in black hash marks. Black areas represent the most distinct regions from the sampled locations, as highlighted by a Mahalanobis distance surface, which were excluded from the predictions. Numbers represent codes of the main unprotected hotspots, as reported in Table S8.

### 3.6.Indicator species effectiveness

Indicator species effectiveness analysis revealed that lesser oriental chevrotain (*Tragulus kanchil*) and Sunda leopard cat (*Prionailurus javanensis*) were the best indicators of biodiversity in Borneo and Sumatra, respectively. Surprisingly, rat species were the second best indicator in both islands. In Borneo, two civet species (banded civet (*Hemigalus derbyanus*) and Malay civet (*Viverra tangalunga*)) were among the top indicator species. Moonrat (*Echinosorex gymnura*), Sunda stink-badger (*Mydaus javanensis*), thick-spined porcupine (*Hystrix crassispinis*) and short-tailed mongoose (*Herpestes brachyurus*) were also strong indicators of biodiversity, along with Sabah partridge (*Arborophila graydoni*) and other partridge species (Table S9). In Sumatra, similar to mainland Southeast Asia (Macdonald et al. 2020), various carnivores were among the top indicators, including members of the Felidae (Sunda leopard cat (*Prionailurus javanensis*) and Sunda clouded leopard (*Neofelis diardi*)), Viverridae (masked palm civet (*Paguma larvata*)) and Canidae (dhole (*Cuon alpinus*)) families. Other important indicators of Sumatran biodiversity included bronze-tailed peacock-pheasant (*Polyplectron chalcurum*), moonrat (*Echinosorex gymnura*), bearded pig (*Sus barbatus*) and mainland serow (*Capricornis sumatraensis*; Table S9).

## 4. Discussion

This study utilised the largest camera trap survey ever carried out in Borneo and Sumatra to (1) assess multivariate habitat relationships of terrestrial meso and mega vertebrate communities, (2) quantify the relative contributions of different factors driving community assemblages, (3) evaluate the level of biodiversity protection, (4) highlight areas where conservation actions should be prioritised, (5) compare rigorously modelled species richness

patterns with more anecdotal knowledge about species distribution, and (6) investigate efficacy of different species as biodiversity indicators. The study has been conducted within a multivariate context, meaning that we accounted for the community-wide responses to habitat factors, while still considering the individual preferences of sampled species. Our results provide the most comprehensive assessment to date of the factors affecting terrestrial vertebrate biodiversity and its spatial patterns across Borneo and Sumatra.

### 4.1.Multi-scale selection of habitat factors

As hypothesised, we found that communities in Borneo and Sumatra respond to habitat factors at different spatial scales. Specifically, sampled species responded to habitat features primarily at medium (2-6km) to broad (8-10km) scales, confirming previous findings on mainland Southeast Asian biodiversity (Macdonald et al. 2020). Our findings corroborate general patterns seen in mainland Southeast Asia (Macdonald et al. 2020), demonstrating, in particular, the importance of extensive intact forest for maintaining high species richness. We also found that anthropogenic factors affect species at broad scales, as found by Hearn et al. (2018b) and Macdonald et al. (2020), confirming the broad negative effects that anthropogenic disturbance, particularly forest loss, has on biodiversity in Southeast Asia and in the Sunda Islands. Borneo and Sumatra have both faced intensive deforestation in the last 50 years, at least, mainly due to conversion of forests into agricultural land, particularly palm oil plantations (Ocampo-Peñuela et al. 2020, Singh and Yan 2021). Here, we highlighted that the detrimental effects of habitat loss affect species not only in proximity of the sources of disturbance, but extend far beyond them. Elevation was the only covariate selected at the finest scale in both islands, consistent with the findings reported by Hearn et al. (2018b), likely reflecting the islands' highly heterogeneous topography. Interestingly, Macdonald et al.

(2020) found elevation to be most related to biodiversity in mainland Southeast Asian at very broad scales, likely reflecting the more extensive impact that altitudinal factors have on biodiversity in wider landscapes.

### 4.2.Environmental predictors of community assemblages

Variance in the multivariate detection counts of sampled species explained for Borneo was comparable with other studies using CCA with high species diversity (Palmer 1993, ter Braak and Verdonschot 1995, Ohmann and Spies 1998, Macdonald et al. 2020), while the same variance explained for Sumatra, although lower, was in the 10-50% range of variance explained typically reported for CCA (Palmer 1993).

As shown by the variance partitioning analysis, environmental factors were major drivers of the Sunda Islands' multivariate species assemblages, explaining the greatest proportion of total variance for Sumatra and the second greatest proportion of total variance for Borneo, reflecting that landscape composition and configuration, along with anthropogenic impacts on the landscape, strongly drive species assemblage patterns. Topographic and spatial covariates were overall less influential. However, CCA and relative biplots highlighted that topographic factors, and especially elevation, were nonetheless fundamental drivers of habitat selection for several species, both in Borneo and Sumatra.

In Borneo, carnivores were associated with the elevation gradient, supporting previous findings by Mathai et al. (2016) and Mathai et al. (2019), predicting that three endemic Bornean carnivores (Bornean ferret badger (*Melogale everetti*), Hose's civet (*Diplogale hosei*) and Borneo bay cat (*Catopuma badia*)), along with masked palm civet (*Paguma larvata*), tend to be associated with higher elevations. Similarly, Hearn et al. (2018b) found strong relationships with higher elevations and Sunda clouded leopard (*Neofelis diardi*)

occurrence in Sabah. The negative impact of human footprint and forest loss was particularly evident for the critically endangered Bornean orangutan (*Pongo pygmaeus*) and the endangered banteng (*Bos javanicus*). Bornean orangutan (*Pongo pygmaeus*) is known to mainly occur in primary forests and pristine forest habitats (Felton et al. 2003, Knop et al. 2004), while banteng (*Bos javanicus*) in Borneo prefers forest habitats as well as open grassland, mature timber plantations and abandoned cultivated lands (Melletti and Burton 2014). However, we highlighted that in Borneo these species were associated with lowland habitats that have experienced extensive forest loss in recent decades, and their association with deforested areas is indicative of the fact that they persist only in areas with high degree of anthropogenic disturbance. This likely reflects their obligate association with lowland rainforest ecosystems which have had rapid and extensive impacts from deforestation and forest degradation. Recent studies have suggested that both these species might be more tolerant of deforestation than was previously thought. For example, Ancrenaz et al. (2010), Meijaard et al. (2010) and Davies et al. (2017), found that Bornean orangutan (*Pongo pygmaeus*) has a degree of tolerance towards disturbed forests and, as discussed above, it is known that banteng (*Bos javanicus*) can be found in disturbed habitats. However, the persistence of these species in areas of heavy human footprint might reflect an extinction debt of large recent changes which has yet to result in equilibrium levels of population decline and local extinction. Therefore, our results indicate that the ecological niches of the most endangered species are currently strongly limited to areas of their original ranges that have experienced the largest impacts of forest loss, and highlight their vulnerability to habitat loss in regions most vulnerable to deforestation. We highlighted also the striking association between Asian elephant (*Elephas maximus*) and palm plantations. Even though the species is commonly found in diverse habitats ranging from tropical evergreen and semi-evergreen forests to cultivated and secondary forests (Choudhury 1999), this association also reflects

that deforestation and land use changes are concentrated at lower elevations and flatter terrain, which are both critical habitat elements for the species (Cushman et al. 2017). Therefore, current Asian elephant (*Elephas maximus*) occurrence patterns likely indicate niche displacement from optimal lowland forests, which are mostly converted to oil palm and industrial timber plantations.

In Sumatra, we observed a greater range of ecological specialization among sampled species, with higher elevation forests emerging as particularly important for biodiversity. A group of species including Sunda clouded leopard (*Neofelis diardi*) showed clear selection for high elevations and upper-montane forests, with predicted occurrence concentrated in the western mountainous region of Sumatra, in agreement with habitat suitability models for Sunda clouded leopard (*Neofelis diardi*) (Macdonald et al. 2018a) and for overall biodiversity (Catullo et al. 2008). More generalist species were associated with deforested lowland areas and this, coupled with the absence of specialists in these habitats, demonstrated the low suitability of disturbed areas for specialised species. Additionally, as highlighted in Borneo, in Sumatra we also found endangered flagship species to be associated with disturbed habitats. Specifically, we found Sunda pangolin (*Manis javanica*) to be associated with oil palm plantations and areas of higher human footprint, raising serious concerns about the conservation status of this critically endangered species, as this likely reflects the increased human impacts in the highly productive and most accessible lowland areas that provide critical habitat for this and other endangered species.

During the last 50 years, at least, Borneo and Sumatra have faced some of the highest rates of deforestation globally, with more than 30% of their original forests converted mostly to agricultural lands and palm oil plantations. Deforestation has, in turn, had significant impacts on the species occurring in the Sunda Islands, including direct loss of biodiversity, increased human-wildlife conflicts, as well as effects on the availability of suitable habitats,

such as habitat loss and fragmentation. Our results suggest that anthropogenic impacts in Borneo and Sumatra have been so rapid and extensive that there is a disequilibrium between the distribution of many species and their habitat needs (e.g., a *lag effect* (Kaszta et al. 2020)). Hence, the association of several endangered flagship species, such as Bornean orangutan (*Pongo pygmaeus*) and Asian elephant (*Elephas maximus*), with anthropogenic factors and disturbed habitat, reflects the degree of perturbation of the Sunda Islands, and must be a red flag about the conservation of their delicate communities, which are currently surviving in sink habitats – environments that would normally be unfavourable or of lower quality for them. In situations when source habitats are lost or degraded, sink habitats might become the last resort for species. However, conservation in sink habitats is challenging due to their limited resources and unsuitable habitat conditions (Dias 1996). Sink habitats can determine declining of populations if they cannot support the survival and reproductive requirements of individuals. Limited resources, as well as unsuitable environmental conditions, might cause higher mortality rates, lower reproductive success, and decrease population size over time.

### 4.3.Indicator species

In both islands, rat species were the second best indicator of biodiversity. However, the challenge of identifying individual rat species from camera trap images limited our analysis to the order level. We acknowledge this limitation and note that it is very unlikely that all the rat species have the same importance as indicators. In addition, we believe that this result is influenced by the large number of generalist species represented in this group. We discourage the use of this group of species for management indications, and believe that future research is needed to clarify the best indicators among rodents.

In Borneo, Viverridae were crucial indicators, with two species (banded civet (*Hemigalus derbyanus*) and Malay civet (*Viverra tangalunga*)) occurring among the best indicator species, corroborating previous findings on the pivotal role of this family for Bornean (Ross et al. 2017) and Southeast Asian (Macdonald et al. 2020) biodiversity.

In Sumatra, we found carnivores to be the strongest overall indicators of species richness. Members from the main carnivore families were among the best indicators, including Felidae (Sunda leopard cat (*Prionailurus javanensis*) and Sunda clouded leopard (*Neofelis diardi*)), Viverridae (masked palm civet (*Paguma larvata*)) and Canidae (dhole (*Cuon alpinus*)). Niche differentiation theory (Hardin 1960) states that if two or more species compete, niche differentiation is a likely outcome to avoid extinction. Accordingly, and similarly to what found in mainland Southeast Asia (Macdonald et al. 2020), we found a single member from each family playing an important role as indicator; only the most common members appeared to be good indicators, while more specialised and rarer species emerged as weaker candidates.

Mammalian carnivores have often been proposed as umbrella species given their large habitat requirements and high mobility (Noss et al. 1996, Carroll et al. 2001, Dalerum et al. 2008). Despite the appeal of this proposition, the assumption that ensuring habitat requirements for one species would automatically preserve sympatric species has been contested (e.g., Cushman et al. (2010)). However, Carroll et al. (2001), Roberge and Angelstam (2004), Cushman and Landguth (2012) and Penjor et al. (2021) demonstrated that conservation strategies based on combinations of multiple umbrella species cover broader ranges of habitat gradients, and can effectively ensure the protection of a higher fraction of sympatric species. Our analysis confirms the importance of carnivores as proxies of biodiversity and, in particular, of the more common and less specialised ones, such as Sunda stink-badger (*Mydaus javanensis*) in Borneo and Sunda leopard cat (*Prionailurus javanensis*)

in Sumatra. Additionally, in a region as rich in endemism as the Sunda Islands, considering more than one carnivore within an umbrella grouping is critical to maximise protection of the regional biodiversity.

### 4.4.Importance of empirical predictions

Biodiversity richness patterns, protected area effectiveness and gap analysis are often assessed using species' inferred distribution layers or other proxies based on coarse-scale, non-empirical data, such as IUCN range maps. It is critical, therefore, to assess how they perform relative to optimised empirically-based models. For Borneo, our species richness surface provided a much more reliable prediction of biodiversity hotspots than did the IUCN surface. The IUCN surface, in addition to having much lower *Nagelkerke-pseudo-$R^2$*, showed very different patterns of predicted species richness, with biodiversity spread more evenly over the central regions of Borneo. Worryingly, Hughes (2017) and Macdonald et al. (2020) also found the IUCN surface of the mainland Southeast Asian biodiversity to drastically differ from the prediction produced with empirical data, with the latter also performing much better than the former. Since several studies use IUCN range layers to provide conservation recommendations, we urge a thorough investigation of factors leading to discrepancies between the expert-based maps and empirical data-based assessments of the sort we have explored.

For Sumatra, our predictive surface and the IUCN map showed similar distribution of biodiversity hotspots, and both had low *Nagelkerke-pseudo-$R^2$*, revealing weak predictive power, which was surprising given the strong explanatory power of the single-species models. We believe that the poor predictive power is due to the disequilibrium between species distributions and their habitat needs. In a perturbed ecological system where species

are associated with habitat conditions because of anthropogenic pressure and not because of ecological associations, species-habitat relationships are difficult to define with precision, leading to weaker predictions (Guisan and Thuiller 2005).

Additionally, a potential alternative, but not mutually exclusive, reason for the weak predictive power in Sumatra, is the spatial clustering of our camera trap surveys. In contrast to Borneo, where sampling locations occurred across a wider set of environmental gradients, in Sumatra camera traps have been deployed only in portions of the western mountainous region. In addition, both in Borneo and Sumatra, camera trap locations were mainly clustered within protected areas, providing a further potential source of spatial bias. However, by sampling within protected areas, we likely provided a more realistic view on species richness patters, reflecting the expected species-habitat relationships that would occur in undisturbed contexts, without accounting for adverse effects of human disturbance, poaching and other factors not associated with measured habitat factors. However, the camera trap surveys could have still been inadequate to represent the range of environmental conditions of the islands, which is known to affect the ability to correctly identify species-habitat relationships (Chiaverini et al. 2021). Therefore, we have minimised the potential errors arising from the spatially biased survey by masking the areas showing the strongest ecological differences from the sampling locations in the multivariate environment of the Mahalanobis distance surface. Future surveys focused on unsampled areas across the full range of ecological gradients are of vital importance to better assess the distribution of biodiversity in Borneo and Sumatra.

### 4.5.Conservation implications

Northern Borneo, specifically the Malaysian state of Sabah, the Indonesian provinces of North and East Kalimantan, and Brunei, are predicted to harbour the highest species richness in Borneo. Interestingly, Raes et al. (2009) found the same regions to be critical hotspots also for botanical species richness. These regions' high species richness is associated with moderate to higher elevations, rough topography and numerous remaining patches of primary forest. While only a few protected areas occur in this region, their critical importance is clear. In Sabah, where heavy logging followed by land conversion to oil palm plantations has altered the landscape, the few protected areas are among those with the highest values of protected biodiversity, such as Crocker Range National Park, Tabin Wildlife Reserve – Mount Hatton Protection Forest Reserve and Ulu Kalumpang Protection Forest Reserve – Tawau Hills Park. Also, Sabah harbours the highest levels of species richness outside of protected areas, with vast unprotected lands that we identified as a biodiversity hotspot. To assess the effectiveness of protected areas and perform gap analysis, however, we evaluated protected areas as those defined by the World Database on Protected Areas (WDPA; UNEP-WCMC and IUCN (2022)) that, in Sabah, includes only areas assigned to Class 1 protection (Dudley et al. 2010). However, several other areas assigned from Class 2 to 5 occur in Sabah and, even though they are excluded from the WDPA, they cannot be deforested. In addition, even though we only evaluated protected areas, it has to be highlighted the important role that alternative conservation strategies such as sustainable forestry practices have on preserving regional biodiversity (Fuller et al. 2004). Additionally, northern and eastern Kalimantan, regions almost completely lacking protection, harbour other critical biodiversity hotspots. Worryingly, this is the area where Indonesia's new capital will be developed (Teo et al. 2020). Hence, extending the current system of protected areas and the network of connectivity corridors in northern Borneo is critical for preserving the biodiversity of the

island (Hearn et al. 2019, Kaszta et al. 2019). It is fundamental to highlight that, even though we did not assess the central highlands of Borneo due to their ecological differences from the areas that we sampled, which might have resulted in producing unreliable prediction, those areas have previously been found fundamental for Bornean biodiversity, particularly for carnivores (Mathai et al. 2016, Macdonald et al. 2018b), and future surveys at higher elevations are important for covering a broader altitudinal range.

In Sumatra, the western mountainous regions harbour the highest levels of species richness. However, we have to be cautious in asserting strong conclusions based on our results given the low predictive power of our species richness surface. A similar pattern characterised by large patches of suitable habitat along the western mountains, and smaller, fragmented patches occurring throughout the central and eastern Sumatran lowlands, was highlighted by Macdonald et al. (2018a) for Sunda clouded leopard (*Neofelis diardi*). This pattern appears to be driven by anthropogenic factors, with habitat loss and deforestation predominating in eastern portions of the island and at lower elevations, and protected areas concentrated in the western highlands. Forest loss and subsequent conversion to oil palm plantations are the main factors negatively affecting species occurrence by driving habitat loss and fragmentation. Conversely, protected areas are a vital factor contributing positively to species richness; the main biodiversity hotspots in Sumatra occur within protected areas, such as Kerinci Seblat and Bukit Barisan Selatan National Parks. Together, these protected areas harbour more than 50% of predicted Sumatran biodiversity. However, in our analysis we excluded the northern portion of the island due to its ecological differences from the locations that we sampled. In northern Sumatra, Gunung Leuser National Park occurs, which is a critically important protected area and, even though in our work we did not assess its contribution in harbouring regional biodiversity, future studies at higher latitudes are strongly needed. Extending the current network of protected areas in Sumatra is critical, especially

along the western mountains where we highlighted the most important unprotected hotspots. Establishing additional protected areas in the southern lowlands is of the essence where forest loss is a clear and imminent threat to biodiversity.

### 4.6.Scope and limitations

Even though the camera trap survey we used in the current work likely represents the most intense effort ever carried out in Borneo and Sumatra to collect camera trap data for meso and macro terrestrial fauna, a number of limitations to our study have to be acknowledged.

First, the camera trap survey did not sample the full gradients of all the environmental factors chosen to model species richness. Ideally, sampling strategies should be designed to identify the full range of gradients that are most influential to species' habitat selection processes, rather than being biased towards a fraction of the full spectrum of environmental conditions (Wessels et al. 1998). Failing to do so might result in models with erroneous parameter estimates and lower predictive performances (Chiaverini et al. 2021). However, in our case, it was impossible to implement a randomly stratified survey given the extent of the study area and the inaccessibility of some regions. Therefore, we produced a Mahalanobis distance surface in order to mask the regions of the study area that were more different from the sampled ones, reducing in this way the risk of producing unreliable predictions of species richness.

Additionally, camera trap grids were unevenly distributed between Borneo and Sumatra, with the latter showing a much more clustered and less representative distribution than the former. This can clearly be seen in the difference of the absolute variance explained by data between Borneo (23.9%) and Sumatra (10.4%). Although the Mahalanobis distance surface

ensured masking the areas ecologically more different from the sampled ones, these differences suggest that the model for Borneo was much more reliable than the model for Sumatra, and more attention should be paid with Sumatra's model when deriving conservation and policy indications, especially for areas that are further away from sampling locations. Similar cautions should to be paid when using the model for Borneo too; however the much more representative survey design implemented in Borneo suggests that the model was stronger and more reliable when projected outside the sampled areas.

Another potential source of bias was represented by the imperfect and varying detection of sampled species (Sollmann et al. 2013). Most of the camera trap locations occurred within protected areas, where environmental conditions drastically differ from unprotected areas, where anthropic pressures might significantly reduce the detectability of several species (Lessa et al. 2017). In addition, deploying camera traps mainly along natural ridgelines and disused logging roads, might have determined an overestimation bias of species that are more likely to use these areas (e.g., felids), leading to underestimating smaller, rarer and more cryptic species (Kolowski and Forrester 2017). However, having modelled the Mahalanobis distance surface accounting also for the influence of protected areas and human footprint, we minimised this geographic bias. Additionally, the same survey protocol was adopted for all the camera trap grids across Borneo and Sumatra, and this consistency minimised the variations in detection probabilities across sampling locations, reducing the potential bias in the species-habitat relationships' estimates.

Although we adopted some of the most precise and reliable land-use layers available for Sunda Islands, coupled with the richest camera trap dataset ever collected in the area, it is fundamental to acknowledge that the land-use layer has been produced more than 10 years ago, and that the camera trap survey has been carried out between 16 and 7 years ago. Considering the magnitude and the speed of the deforestation in Borneo and Sumatra, we

have to acknowledge that our models might fail to accurately represent the current distribution of suitable habitat for sampled biodiversity. We are confident that the modelling framework that we applied accurately described the species-habitat relationships between Sundaic terrestrial vertebrate communities and their habitat; however the ongoing habitat loss occurring in Borneo and Sumatra would require up-to-date layers accurately describing the land-use of the region, in order to produce timely and punctual predictions of biodiversity distribution.

## 5. References


Ancrenaz, M., L. Ambu, I. Sunjoto, E. Ahmad, K. Manokaran, E. Meijaard, and I. Lackman. 2010. Recent Surveys in the Forests of Ulu Segama Malua, Sabah, Malaysia, Show That Orang-utans (P. p. morio) Can Be Maintained in Slightly Logged Forests. Plos One **5**.

Austin, M. P., A. O. Nicholls, M. D. Doherty, and J. A. Meyers. 1994. Determining Species Response Functions to an Environmental Gradient by Means of a Beta-Function. Journal of Vegetation Science **5**:215-228.

Borcard, D., P. Legendre, and P. Drapeau. 1992. Partialling out the Spatial Component of Ecological Variation. Ecology **73**:1045-1055.

Carroll, C., R. F. Noss, and P. C. Paquet. 2001. Carnivores as focal species for conservation planning in the Rocky Mountain region. Ecological Applications **11**:961-980.

Catullo, G., M. Masi, A. Falcucci, L. Maiorano, C. Rondinini, and L. Boitani. 2008. A gap analysis of Southeast Asian mammals based on habitat suitability models. Biological Conservation **141**:2730-2744.

Chambers, C. L., S. A. Cushman, A. Medina-Fitoria, J. Martinez-Fonseca, and M. Chavez-Velasquez. 2016. Influences of scale on bat habitat relationships in a forested landscape in Nicaragua. Landscape Ecology **31**:1299-1318.

Chiaverini, L., H. Y. Wan, B. Hahn, A. Cilimburg, T. N. Wasserman, and S. A. Cushman. 2021. Effects of non-representative sampling design on multi-scale habitat models: flammulated owls in the Rocky Mountains. Ecological Modelling **450**.

Choudhury, A. 1999. Status and conservation of the Asian Elephant Elephas maximus in north-eastern India. Mammal Review **29**:141-173.

CIESIN, and CIAT. 2016. Gridded Population of the World version 4 (GPWv4): Population density grids. Socioeconomic Data and Applications Center (SEDAC), Columbia University, Palisades, NY. http://sedac.ciesin.columbia.edu/gpw.

Coleman, B. T., and R. A. Hill. 2014. Living in a landscape of fear: the impact of predation, resource availability and habitat structure on primate range use. Animal Behaviour **88**:165-173.

Cushman, S. A., and E. L. Landguth. 2012. Multi-taxa population connectivity in the Northern Rocky Mountains. Ecological Modelling **231**:101-112.

Cushman, S. A., E. Macdonald, E. Landguth, Y. Malhi, and D. Macdonald. 2017. Multiple-scale prediction of forest loss risk across Borneo. Landscape Ecology **32**:1581-1598.

Cushman, S. A., and K. McGarigal. 2002. Hierarchical, multi-scale decomposition of species-environment relationships. Landscape Ecology **17**:637-646.

Cushman, S. A., and K. McGarigal. 2004a. Hierarchical analysis of forest bird species-environment relationships in the Oregon Coast Range. Ecological Applications **14**:1090-1105.

Cushman, S. A., and K. McGarigal. 2004b. Patterns in the species-environment relationship depend on both scale and choice of response variables. Oikos **105**:117-124.

Cushman, S. A., K. S. McKelvey, B. R. Noon, and K. McGarigal. 2010. Use of Abundance of One Species as a Surrogate for Abundance of Others. Conservation Biology **24**:830-840.

Dalerum, F., M. J. Somers, K. E. Kunkel, and E. Z. Cameron. 2008. The potential for large carnivores to act as biodiversity surrogates in southern Africa. Biodiversity and Conservation **17**:2939-2949.

Davies, A. B., M. Ancrenaz, F. Oram, and G. P. Asner. 2017. Canopy structure drives orangutan habitat selection in disturbed Bornean forests. Proceedings of the National Academy of Sciences of the United States of America **114**:8307-8312.

Davis, C. L., L. N. Rich, Z. J. Farris, M. J. Kelly, M. S. Di Bitetti, Y. Di Blanco, S. Albanesi, M. S. Farhadinia, N. Gholikhani, S. Hamel, B. J. Harmsen, C. Wultsch, M. D. Kane, Q. Martins, A. J. Murphy, R. Steenweg, S. Sunarto, A. Taktehrani, K. Thapa, J. M. Tucker, J. Whittington, F. A. Widodo, N. G. Yoccoz, and D. A. W. Miller. 2018. Ecological correlates of the spatial co-occurrence of sympatric mammalian carnivores worldwide. Ecology Letters **21**:1401-1412.

Dias, P. C. 1996. Sources and sinks in population biology. Trends in Ecology & Evolution **11**:326-330.

Dudley, N., J. D. Parrish, K. H. Redford, and S. Stolton. 2010. The revised IUCN protected area management categories: the debate and ways forward. Oryx **44**:485-490.

ESRI. 2018. ArcGIS Desktop: Release 10.6.1. Environmental System Research Institute. Redlands, CA, USA.

Evans, J. S., J. Oakleaf, S. A. Cushman, and D. Theobald. 2014. An ArcGIS toolbox for surface gradient and geomorphometric modeling, version 2.0-0. Available: http://evansmurphy.wix.com/evansspatial.

Felton, A. M., L. M. Engstrom, A. Felton, and C. D. Knott. 2003. Orangutan population density, forest structure and fruit availability in hand-logged and unlogged peat swamp forests in West Kalimantan, Indonesia. Biological Conservation **114**:91-101.

Fuller, D. O., T. C. Jessup, and A. Salim. 2004. Loss of forest cover in Kalimantan, Indonesia, since the 1997-1998 El Nino. Conservation Biology **18**:249-254.

Grand, J., J. Buonaccorsi, S. A. Cushman, C. R. Griffin, and M. C. Neel. 2004. A multiscale landscape approach to predicting bird and moth rarity hotspots, in a threatened pitch pine-scrub oak community. Conservation Biology **18**:1063-1077.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters **8**:993-1009.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling **135**:147-186.

Haidir, I. A., D. W. Macdonald, and M. Linkie. 2018. Assessing the spatiotemporal interactions of mesopredators in Sumatra's tropical rainforest. Plos One **13**.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science **342**:850-853.

Hardin, G. 1960. Competitive Exclusion Principle. Science **131**:1292-1297.

Hearn, A. J., S. A. Cushman, B. Goossens, E. Macdonald, J. Ross, L. T. B. Hunter, N. K. Abram, and D. W. Macdonald. 2018a. Evaluating scenarios of landscape change for Sunda clouded leopard connectivity in a human dominated landscape. Biological Conservation **222**:232-240.

Hearn, A. J., S. A. Cushman, B. Goossens, J. Ross, E. A. Macdonald, L. T. B. Hunter, and D. W. Macdonald. 2019. Predicting connectivity, population size and genetic diversity of Sunda clouded leopards across Sabah, Borneo. Landscape Ecology **34**:275-290.

Hearn, A. J., S. A. Cushman, J. Ross, B. Goossens, L. T. B. Hunter, and D. W. Macdonald. 2018b. Spatio-temporal ecology of sympatric felids on Borneo. Evidence for resource partitioning? Plos One **13**.

Hearn, A. J., J. Ross, H. Bernard, S. A. Bakar, B. Goossens, L. T. B. Hunter, and D. W. Macdonald. 2017. Responses of Sunda clouded leopard Neofelis diardi population density to anthropogenic disturbance: Refining estimates of its conservation status in Sabah. Oryx:1-11.

Hijmans, R. J., S. Phillips, J. Leathwick, and J. Elith. 2017. dismo: Species distribution modeling.

Hughes, A. C. 2017. Mapping priorities for conservation in Southeast Asia. Biological Conservation **209**:395-405.

Hutchinson, G. E. 1957. Population Studies - Animal Ecology and Demography - Concluding Remarks. Cold Spring Harbor Symposia on Quantitative Biology **22**:415-427.

IUCN. 2019. The IUCN Red List of Threatened Species. Version 2019-5. https://www.iucnredlist.org. Downloaded on 17 May 2019.

Jarvis, A., H. I. Reuter, A. Nelson, and E. Guevara. 2008. Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT). http://srtm.csi.cgiar.org.

Kaszta, Z., S. A. Cushman, A. J. Hearn, D. Burnham, E. A. Macdonald, B. Goossens, S. K. S. S. Nathan, and D. W. Macdonald. 2019. Integrating Sunda clouded leopard (Neofelis diardi) conservation into development and restoration planning in Sabah (Borneo). Biological Conservation **235**:63-76.

Kaszta, Z., S. A. Cushman, S. Htun, H. Naing, D. Burnham, and D. W. Macdonald. 2020. Simulating the impact of Belt and Road initiative and other major developments in Myanmar on an ambassador felid, the clouded leopard, Neofelis nebulosa. Landscape Ecology **35**:727-746.

Knop, E., P. I. Ward, and S. A. Wich. 2004. A comparison of orang-utan density in a logged and unlogged forest on Sumatra. Biological Conservation **120**:183-188.

Kolowski, J. M., and T. D. Forrester. 2017. Camera trap placement and the potential for bias due to trails and other features. Plos One **12**.

Lessa, I. C. M., A. C. Ferreguetti, M. Kajin, C. R. Dickman, and H. G. Bergallo. 2017. You can't run but you can hide: the negative influence of human presence on mid-sized mammals on an Atlantic island. Journal of Coastal Conservation **21**:829-836.

Macdonald, D. W., H. M. Bothwell, A. J. Hearn, S. M. Cheyne, I. Haidir, L. T. B. Hunter, A. Kaszta, M. Linkie, E. A. Macdonald, J. Ross, and S. A. Cushman. 2018a. Multi-scale habitat selection modeling identifies threats and conservation opportunities for the Sunda clouded leopard (Neofelis diardi). Biological Conservation **227**:92-103.

Macdonald, D. W., D. Burnham, A. E. Hinks, and R. Wrangham. 2012. A problem shared is a problem reduced: seeking efficiency in the conservation of felids and primates. Folia Primatol (Basel) **83**:171-215.

Macdonald, D. W., L. Chiaverini, H. M. Bothwell, Z. Kaszta, E. Ash, G. Bolongon, O. E. Can, A. Campos-Arceiz, P. Channa, G. R. Clements, A. J. Hearn, L. Hedges, S. Htun, J. F. Kamler, E. A. Macdonald, J. Moore, H. Naing, M. Onuma, A. Rasphone, D. M. Rayan, J. Ross, P. Singh, C. K. W. Tan, J. Wadey, B. P. Yadav, and S. A. Cushman. 2020. Predicting biodiversity richness in rapidly changing landscapes: climate, low human pressure or protection as salvation? Biodiversity and Conservation.

Macdonald, E. A., S. A. Cushman, E. L. Landguth, A. J. Hearn, Y. Malhi, and D. W. Macdonald. 2018b. Simulating impacts of rapid forest loss on population size, connectivity and genetic diversity of Sunda clouded leopards (Neofelis diardi) in Borneo. Plos One **13**.

Martin, T. G., B. A. Wintle, J. R. Rhodes, P. M. Kuhnert, S. A. Field, S. J. Low-Choy, A. J. Tyre, and H. P. Possingham. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecology Letters **8**:1235-1246.

Mathai, J., J. W. Duckworth, E. Meijaard, G. Fredriksson, J. Hon, A. Sebastian, M. Ancrenaz, A. J. Hearn, J. Ross, S. Cheyne, A. Wilting, and B. C. Consortium. 2016. Carnivore conservation planning on Borneo: identifying key carnivore landscapes, research priorities and conservation interventions. Raffles Bulletin of Zoology:186-217.

Mathai, J., J. Niedballa, V. Radchuk, R. Sollmann, I. Heckmann, J. Brodie, M. Struebig, A. J. Hearn, J. Ross, D. W. Macdonald, J. Hon, and A. Wilting. 2019. Identifying refuges for Borneo's elusive Hose's civet. Global Ecology and Conservation **17**:e00531.

McGarigal, K., S. Cushman, and S. G. Stafford. 2000. Multivariate statistics for wildlife and ecology research. Pages 1 online resource (xiii, 283 pages). Springer,, New York.

McGarigal, K., S. A. Cushman, and E. Ene. 2012. FRAGSTATS v4: spatial pattern analysis program for categorical and continuous maps. Computer software program produced by the authors at the University of Massachusetts, Amherst, MA. Available: http://www.umass.edu/landeco/research/fragstats/fragstats.html.

McGarigal, K., H. Y. Wan, K. A. Zeller, B. C. Timm, and S. A. Cushman. 2016. Multi-scale habitat selection modeling: a review and outlook. Landscape Ecology **31**:1161-1175.

Meijaard, E., G. Albar, Nardiyono, Y. Rayadin, M. Ancrenaz, and S. Spehar. 2010. Unexpected Ecological Resilience in Bornean Orangutans and Implications for Pulp and Paper Plantation Management. Plos One **5**.

Melletti, M., and J. Burton. 2014. Ecology, Evolution and Behaviour of Wild Cattle Implications for Conservation Introduction. Ecology, Evolution and Behaviour of Wild Cattle: Implications for Conservation:1-6.

Miettinen, J., C. H. Shi, W. J. Tan, and S. C. Liew. 2012. 2010 land cover map of insular Southeast Asia in 250-m spatial resolution. Remote Sensing Letters **3**:11-20.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. Nature **403**:853-858.

Nagelkerke, N. J. D. 1991. A Note on a General Definition of the Coefficient of Determination. Biometrika **78**:691-692.

Noss, R. F., H. B. Quigley, M. G. Hornocker, T. Merrill, and P. C. Paquet. 1996. Conservation biology and carnivore conservation in the Rocky Mountains. Conservation Biology **10**:949-963.

Ocampo-Peñuela, N., J. Garcia-Ulloa, I. Kornecki, C. D. Philipson, and J. Ghazoul. 2020. Impacts of Four Decades of Forest Loss on Vertebrate Functional Habitat on Borneo. Frontiers in Forests and Global Change **3**.

Ohmann, J. L., and T. A. Spies. 1998. Regional gradient analysis and spatial pattern of woody plant communities of Oregon forests. Ecological Monographs **68**:151-182.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. Henry, H. Stevens, E. Szoecs, and H. Wagner. 2018. Vegan: community ecology package. R package version 2.4-6. http://CRAN.R-project.org/package=vegan.

Ovaskainen, O., J. Hottola, and J. Siitonen. 2010. Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. Ecology **91**:2514-2521.

Palmer, M. W. 1993. Putting Things in Even Better Order - the Advantages of Canonical Correspondence-Analysis. Ecology **74**:2215-2230.

Penjor, U., Z. Kaszta, D. W. Macdonald, and S. A. Cushman. 2021. Prioritizing areas for conservation outside the existing protected area network in Bhutan: the use of multi-species, multi-scale habitat suitability models. Landscape Ecology **36**:1281-1309.

PRISM Climate Group. 2016. Oregon State University. http://prism.oregonstate.edu.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raes, N., M. C. Roos, J. W. F. Slik, E. E. van Loon, and H. ter Steege. 2009. Botanical richness and endemicity patterns of Borneo derived from species distribution models. Ecography **32**:180-192.

Roberge, J. M., and P. Angelstam. 2004. Usefulness of the umbrella species concept as a conservation tool. Conservation Biology **18**:76-85.

Ross, J., A. J. Hearn, and D. W. Macdonald. 2017. The Bornean carnivore community: lessons from a little-known guild. Biology and Conservation of Musteloids:326-339.

Simcharoen, A., T. Savini, G. A. Gale, S. Simcharoen, S. Duangchantrasiri, S. Pakpien, and J. L. D. Smith. 2014. Female tiger Panthera tigris home range size and prey abundance: important metrics for management. Oryx **48**:370-377.

Singh, M., and S. H. Yan. 2021. Spatial-temporal variations in deforestation hotspots in Sumatra and Kalimantan from 2001-2018. Ecology and Evolution **11**:7302-7314.

Sodhi, N. S., L. P. Koh, B. W. Brook, and P. K. L. Ng. 2004. Southeast Asian biodiversity: an impending disaster. Trends in Ecology & Evolution **19**:654-660.

Sollmann, R., A. Mohamed, H. Samejima, and A. Wilting. 2013. Risky business or simple solution - Relative abundance indices from camera-trapping. Biological Conservation **159**:405-412.

Teo, H. C., A. M. Lechner, S. Sagala, and A. Campos-Arceiz. 2020. Environmental Impacts of Planned Capitals and Lessons for Indonesia's New Capital. Land **9**.

ter Braak, C. J. F. 1986. Canonical Correspondence-Analysis - a New Eigenvector Technique for Multivariate Direct Gradient Analysis. Ecology **67**:1167-1179.

ter Braak, C. J. F., and I. C. Prentice. 1988. A Theory of Gradient Analysis. Advances in Ecological Research **18**:271-317.

ter Braak, C. J. F., and P. F. M. Verdonschot. 1995. Canonical Correspondence-Analysis and Related Multivariate Methods in Aquatic Ecology. Aquatic Sciences **57**:255-289.

UNEP-WCMC, and IUCN. 2022. Protected Planet: The World Database on Protected Areas (WDPA), Accessed April 2022, Cambridge, UK: UNEP-WCMC and IUCN. Available at: www.protectedplanet.net.

Vincent, P. J., and J. M. Haworth. 1983. Poisson Regression-Models of Species Abundance. Journal of Biogeography **10**:153-160.

Wang, Y. W., M. L. Allen, and C. C. Wilmers. 2015. Mesopredator spatial and temporal responses to large predators and human development in the Santa Cruz Mountains of California. Biological Conservation **190**:23-33.

WCS, and CIESIN. 2005. Last of The Wild data version 2, (LTW-2): global human footprint dataset (geographic). http://sedac.ciesin.columbia.edu/wildareas/.

Wessels, K. J., A. S. Van Jaarsveld, J. D. Grimbeek, and M. J. Van der Linde. 1998. An evaluation of the gradsect biological survey method. Biodiversity and Conservation **7**:1093-1121.

Wilcove, D. S., X. Giam, D. P. Edwards, B. Fisher, and L. P. Koh. 2013. Navjot's nightmare revisited: logging, agriculture, and biodiversity in Southeast Asia. Trends in Ecology & Evolution **28**:531-540.

# Supplementary material for Chapter 3

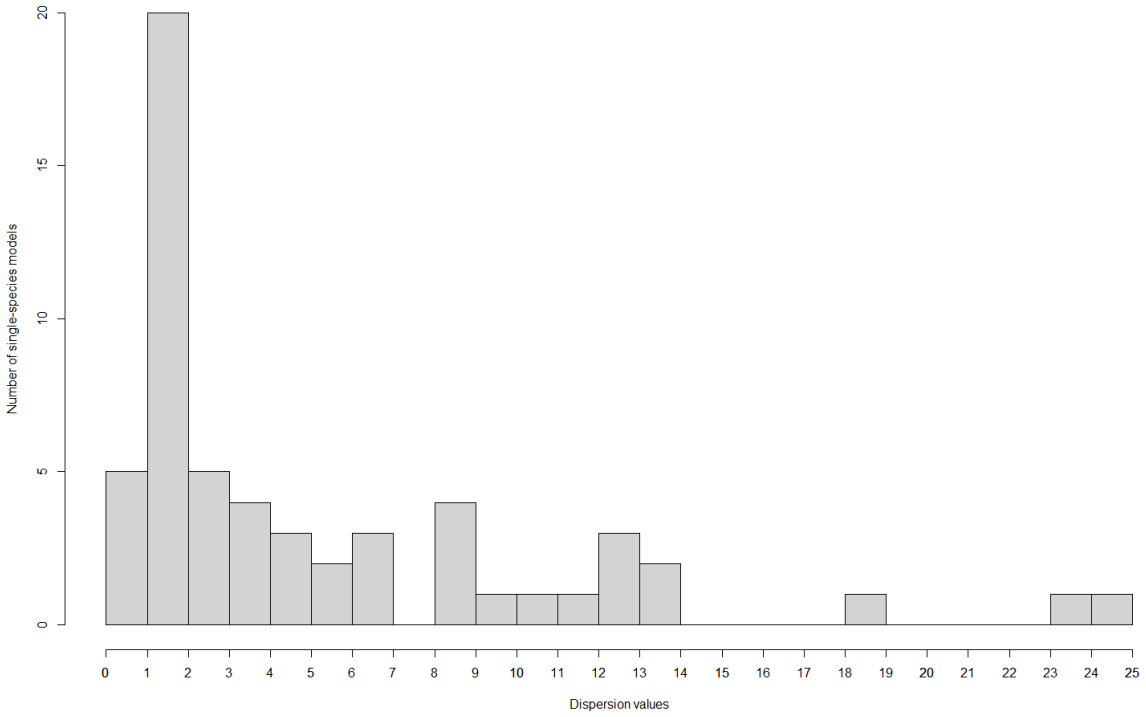



**Figure S1.** Histograms showing the distribution of the dispersion values in relation to the number of single-species GLMs with Poisson distribution, for Borneo.

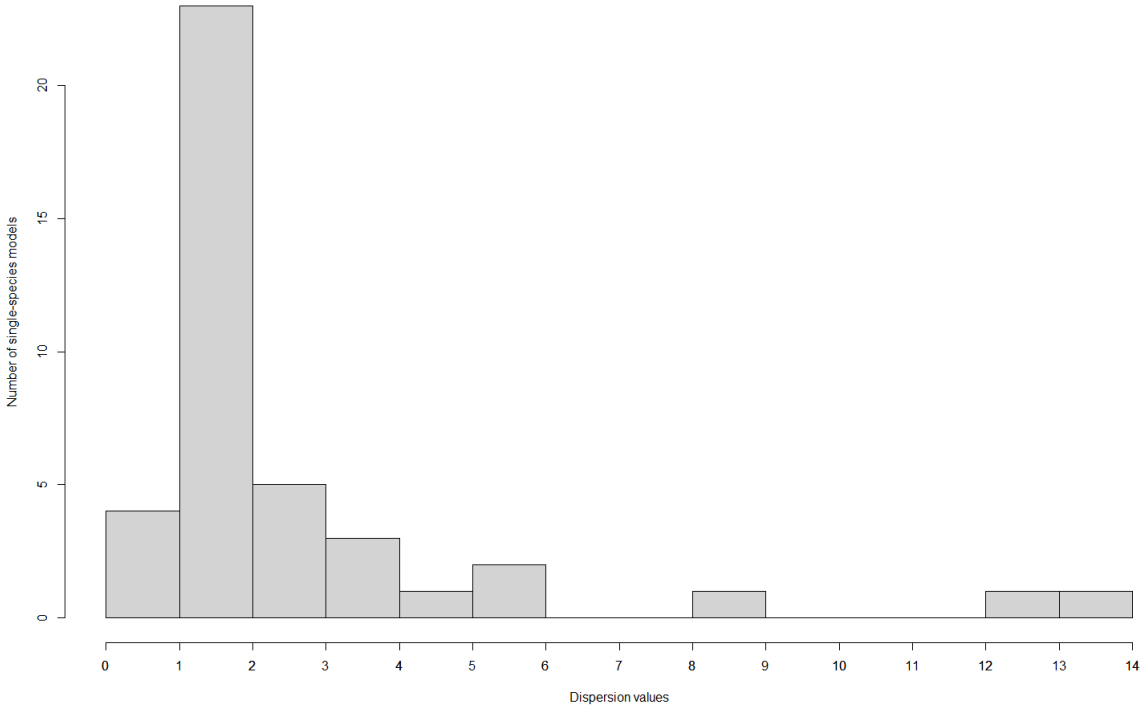



**Figure S2.** Histograms showing the distribution of the dispersion values in relation to the number of single-species GLMs with Poisson distribution, for Sumatra.

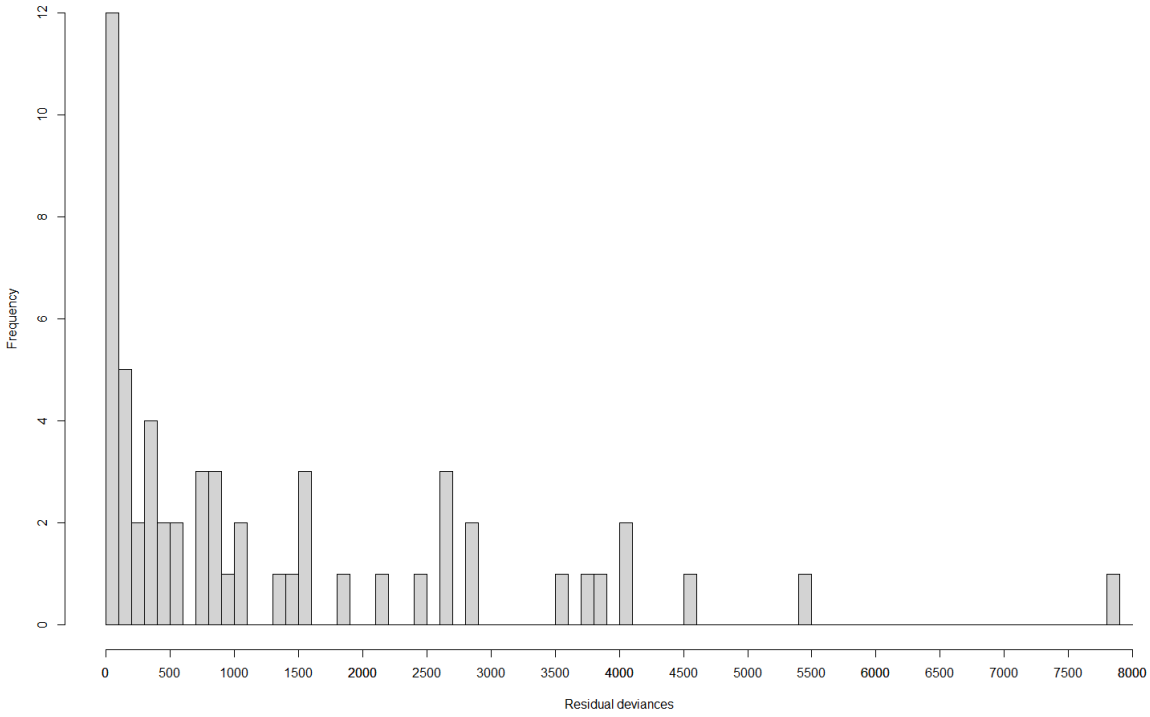



**Figure S3.** Histograms showing the distribution of the residual deviances of the single-species GLMs of species sampled in Borneo.

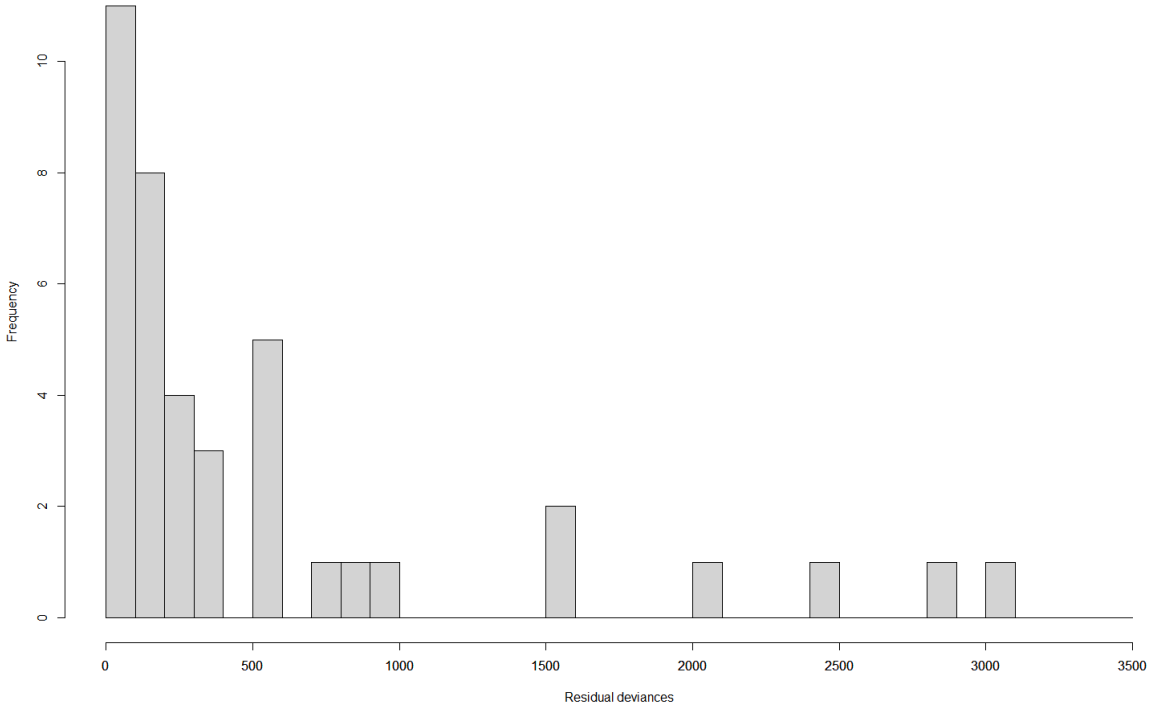



**Figure S4.** Histograms showing the distribution of the residual deviances of the single-species GLMs of species sampled in Sumatra.

**Table S1.** Covariates' coefficients of the single-species GLMs for Borneo. Covariates acronyms: I (intercept), TN (camera trap nights), CTI_SD (compound topographic index SD 4,000m), E_FM (elevation FM 250m), HF_FM (human footprint FM 10,000m), HF_SD (human footprint SD 2,000m), LF_GY (lowland forests GYRATE_AM 8,000m), LO_GY (lowland open areas GYRATE_AM 2,000m), FL_PL (forest loss PLAND 10,000m), M_PL (mangroves PLAND 10,000m), PP_PL (palm plantations PLAND 2,000m), PA_PL (protected areas PLAND 4,000m), SP_FM (slope position FM 8,000m), TR_PD (tree canopy cover reclassified PD 4,000m). See Table S3 for species' codes.

| Species' code | I | TN | CTI_SD | E_FM | HF_FM | HF_SD | LF_GY | LO_GY | FL_PL | M_PL | PP_PL | PA_PL | SP_FM | TR_PD | X | $X^2Y^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -23.74 | 0.01 | -0.34 | -2.57 | -0.05 | 0.02 | 0.17 | 0.21 | -0.63 | -44.33 | -0.50 | 0.56 | -0.15 | 0.27 | 3.78 | 2.89 |
| 2 | -3.41 | 0.00 | 0.20 | -0.02 | 1.01 | -0.31 | -0.39 | -0.21 | -0.45 | -0.37 | -0.67 | -0.38 | 0.18 | -0.21 | 5.82 | 2.62 |
| 5 | 1.26 | 0.01 | 0.06 | 0.05 | 0.07 | 0.16 | 0.29 | 0.02 | -0.19 | -0.28 | -0.18 | 0.26 | -0.05 | 0.43 | 0.85 | 0.34 |
| 6 | 0.10 | 0.00 | 0.27 | -0.62 | 0.21 | -0.08 | 0.46 | -0.73 | -0.32 | 0.15 | -0.87 | 0.36 | 0.00 | 0.44 | 0.42 | 0.30 |
| 7 | -1.04 | 0.00 | 0.23 | -0.55 | 0.58 | 0.09 | 0.35 | -0.50 | -0.49 | -0.13 | -0.55 | -0.15 | -0.06 | -0.43 | 3.53 | 1.46 |
| 8 | -1.44 | 0.00 | 0.30 | -1.97 | 1.31 | -0.51 | 0.06 | -0.30 | -1.34 | -0.29 | -0.88 | 0.64 | 0.06 | 0.33 | 0.39 | 1.07 |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 9 | -1.46 | 0.00 | 0.31 | 1.04 | -0.62 | -0.14 | 0.13 | -0.05 | -0.41 | 0.16 | -0.38 | -0.96 | -0.08 | -0.12 | -0.22 | -0.85 |
| 10 | -31.44 | 0.01 | 0.00 | 0.79 | 0.32 | 0.08 | 0.23 | 0.06 | -0.46 | -69.08 | -0.85 | 1.36 | -0.03 | 0.60 | -1.21 | -1.01 |
| 11 | -0.27 | 0.00 | 0.07 | 0.25 | 1.30 | 0.00 | 1.18 | -0.36 | -0.42 | -0.09 | -0.76 | -0.52 | -0.12 | -0.11 | 1.35 | -0.36 |
| 12 | -0.69 | 0.01 | -0.25 | -0.85 | 0.27 | 0.19 | 0.24 | -0.49 | -0.22 | -0.61 | -0.53 | 0.20 | -0.04 | -0.11 | 1.02 | 0.93 |
| 13 | -14.50 | 0.01 | -0.57 | -2.64 | 0.48 | -0.07 | 0.07 | -1.33 | 0.87 | -17.26 | 0.39 | 0.58 | -0.38 | -5.48 | 1.02 | 2.07 |
| 16 | -4.02 | 0.00 | 0.34 | -0.22 | -0.23 | 0.03 | -0.56 | -0.87 | -1.37 | -0.31 | -0.25 | -0.51 | 0.15 | -0.88 | 2.01 | 1.10 |
| 18 | -107.76 | 0.00 | -0.01 | -0.23 | 0.21 | 0.94 | 2.48 | -21.88 | -1.57 | -19.34 | -272.16 | -0.99 | 0.82 | -0.34 | 3.70 | 2.00 |
| 19 | -4.48 | 0.00 | -0.44 | -0.56 | 1.30 | -0.45 | 0.33 | 0.46 | -1.17 | -0.37 | -1.05 | -0.08 | 0.07 | -0.22 | 0.41 | 1.24 |
| 20 | -31.12 | 0.01 | 0.32 | 1.20 | -0.32 | -0.18 | -0.06 | -22.98 | 0.36 | -30.62 | -0.24 | 0.60 | 0.11 | -0.58 | 3.97 | 1.51 |
| 21 | -3.05 | 0.01 | 0.18 | -2.05 | 0.09 | -0.02 | -0.11 | 0.05 | -0.33 | -0.72 | 0.11 | -0.25 | -0.17 | -0.55 | 3.21 | 2.24 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 22 | -17.40 | 0.01 | -0.29 | -0.56 | -1.29 | 0.55 | 0.27 | -22.43 | 0.49 | 0.64 | -4.35 | 1.92 | -0.51 | 0.27 | 0.93 | 1.07 |
| 23 | -1.11 | 0.00 | 0.36 | 0.44 | 0.60 | -0.41 | 0.04 | 0.03 | -0.11 | -0.37 | -0.26 | -0.48 | 0.06 | -0.68 | 2.80 | 0.96 |
| 24 | -107.82 | 0.00 | 0.34 | 1.26 | 2.00 | 0.66 | 1.60 | -22.05 | -0.10 | -33.14 | -259.05 | -1.15 | 0.41 | -2.25 | 0.56 | 1.06 |
| 25 | -0.75 | 0.01 | 0.25 | 0.40 | -0.07 | -0.04 | 0.32 | 0.13 | -0.23 | -0.05 | 0.06 | 0.11 | -0.05 | -0.12 | 3.82 | 1.50 |
| 26 | -4.17 | 0.01 | 0.53 | 0.54 | -0.29 | -0.11 | 0.17 | -0.99 | -0.59 | -0.66 | 0.17 | -0.13 | 0.09 | -0.48 | 1.71 | 0.92 |
| 27 | -3.44 | 0.01 | 0.29 | 0.28 | 0.37 | 0.29 | 0.81 | 0.19 | -0.53 | -1.05 | -0.18 | -0.17 | -0.02 | -0.05 | 2.80 | 2.01 |
| 28 | -4.61 | 0.01 | -0.01 | -0.59 | -0.22 | -0.12 | 0.15 | 0.34 | -0.76 | -0.68 | 0.51 | 1.19 | 0.00 | -0.07 | 1.11 | 1.05 |
| 29 | -29.78 | 0.00 | 0.34 | 0.47 | 1.19 | 0.26 | 0.73 | -23.27 | -0.10 | -26.01 | -2.48 | 1.28 | 0.28 | 1.03 | 2.52 | 2.06 |
| 30 | -2.15 | 0.00 | 0.38 | 0.95 | -0.60 | 0.16 | 0.58 | -0.42 | -0.21 | 0.33 | -1.01 | 1.13 | -0.04 | 1.28 | -0.18 | -0.41 |
| 32 | -8.31 | 0.00 | 1.07 | 1.36 | -0.58 | 0.09 | 0.65 | 0.25 | 0.50 | -6.90 | 0.21 | -0.13 | 0.06 | -0.48 | 6.52 | 2.04 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 34 | -73.24 | 0.01 | -0.33 | 0.30 | 0.97 | -0.81 | -0.81 | -9.42 | 1.07 | -0.50 | -213.92 | 3.99 | -0.96 | -0.66 | -3.97 | -3.23 |
| 35 | -2.37 | 0.01 | -0.51 | -1.47 | 0.92 | -0.07 | 1.09 | -0.17 | -0.96 | -0.33 | 0.12 | 0.40 | -0.48 | 0.56 | 1.63 | 2.02 |
| 37 | -1.02 | 0.00 | 0.17 | 0.10 | 0.55 | 0.06 | 0.46 | -0.06 | -0.11 | -0.49 | -0.27 | 0.23 | 0.02 | -0.24 | -0.22 | -0.50 |
| 38 | -10.36 | 0.01 | 0.04 | 0.29 | -0.01 | 0.20 | -0.47 | 0.12 | -0.08 | 0.02 | -0.28 | 0.63 | -0.15 | 0.17 | 14.98 | 7.38 |
| 39 | -2.51 | 0.01 | 0.17 | 0.42 | -0.08 | 0.29 | 0.00 | -0.61 | 0.00 | -0.35 | 0.07 | 0.72 | -0.18 | 0.34 | 0.13 | 0.41 |
| 40 | -444.79 | 0.38 | 47.13 | 45.50 | -145.59 | 63.19 | 133.53 | 4.12 | -95.15 | 97.56 | 24.09 | -50.53 | 9.51 | -58.58 | 213.82 | 179.54 |
| 41 | -27.08 | 0.03 | 0.27 | 0.10 | 0.94 | 0.08 | 0.09 | -4.17 | -0.39 | -9.17 | 1.36 | 4.12 | -0.03 | 3.59 | 19.69 | 11.10 |
| 42 | -33.15 | -0.02 | -0.45 | -0.90 | -0.78 | 1.41 | -1.20 | 0.90 | -3.34 | -63.16 | -0.21 | -1.51 | -0.23 | -0.25 | 4.42 | 3.12 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 43 | -11.18 | 0.01 | -0.70 | -1.85 | -0.52 | -0.33 | -1.08 | -1.41 | -1.85 | -0.22 | -0.09 | 0.23 | -0.05 | -0.03 | 10.21 | 5.98 |
| 44 | -16.12 | 0.00 | -0.45 | -0.19 | -3.02 | 1.38 | -0.75 | 1.48 | -0.61 | 1.84 | -14.31 | 6.16 | -0.72 | 1.62 | 8.67 | 4.88 |
| 45 | -1.22 | 0.00 | 0.04 | 0.22 | 0.66 | -0.12 | -0.30 | -0.18 | -0.86 | -0.35 | -0.01 | -0.26 | -0.02 | -0.49 | 2.43 | 1.16 |
| 46 | -3.74 | 0.00 | 0.13 | -0.26 | 2.14 | 0.71 | -0.14 | -1.09 | -2.29 | -0.03 | -0.10 | 2.44 | 0.30 | 0.36 | -2.88 | -0.02 |
| 47 | -3.85 | 0.01 | 0.21 | -0.83 | 0.06 | -0.55 | -1.36 | -0.32 | -0.65 | -0.74 | 0.25 | 0.03 | 0.19 | -0.72 | 5.35 | 2.87 |
| 48 | -0.28 | 0.00 | -0.10 | 0.07 | 0.10 | 0.13 | -0.15 | -0.01 | -0.46 | -0.77 | 0.14 | 0.29 | 0.02 | -0.26 | 0.00 | 0.20 |
| 49 | -3.23 | 0.01 | -0.19 | 0.08 | 0.20 | -0.55 | -0.65 | -0.78 | -0.38 | 0.13 | -0.16 | 1.18 | 0.12 | 0.82 | 0.58 | 0.85 |
| 50 | -32.00 | 0.01 | -0.04 | 0.63 | 1.68 | 0.21 | -0.10 | -0.18 | -0.82 | -63.02 | 0.59 | 1.23 | -0.15 | -0.43 | -1.70 | -0.21 |
| 51 | -4.77 | 0.01 | -0.19 | 0.20 | 0.76 | 0.30 | -0.74 | -0.87 | -1.14 | 0.10 | -1.38 | 0.47 | 0.14 | 0.77 | 4.49 | 2.41 |
| 53 | -68.60 | -0.01 | 0.96 | -2.79 | -8.22 | -3.53 | -3.88 | 1.80 | 1.48 | -10.53 | 6.12 | -33.31 | 1.37 | 0.58 | -3.39 | -5.57 |
| 55 | 0.82 | 0.01 | -0.02 | -0.39 | 0.77 | 0.03 | -0.02 | 0.13 | -0.32 | -0.28 | 0.00 | -0.21 | -0.04 | -0.24 | 0.62 | 0.63 |
| 56 | -4.67 | 0.02 | -0.19 | -4.96 | -1.52 | 0.18 | -0.24 | 0.31 | -0.26 | 0.35 | -0.22 | 0.68 | 0.02 | 0.44 | 2.05 | 2.06 |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 57 | -2.10 | 0.01 | 0.02 | -0.87 | -0.40 | -0.18 | -0.53 | 0.02 | 0.30 | -0.29 | -0.38 | 0.28 | -0.04 | -0.01 | 0.83 | 0.78 |
| 58 | -90.00 | 0.01 | 0.45 | 0.97 | 2.00 | 0.35 | 0.87 | -21.64 | -1.02 | -30.19 | -192.46 | 3.08 | -0.60 | 0.43 | -2.57 | -0.40 |
| 59 | -23.13 | 0.02 | 1.09 | 1.60 | -3.75 | -0.09 | 0.05 | -0.28 | 0.06 | -4.26 | 0.01 | 1.63 | -0.14 | 1.79 | 25.98 | 12.25 |
| 60 | -13.81 | 0.00 | 0.89 | -0.68 | 1.14 | -0.66 | -1.73 | -1.16 | -1.72 | -1.94 | -0.54 | 0.05 | 0.05 | -0.13 | 18.47 | 8.42 |
| 61 | -128.64 | 0.01 | -0.07 | -0.79 | 0.09 | 0.60 | -1.84 | -26.28 | -4.39 | -101.34 | -225.19 | 1.05 | 0.74 | 4.33 | 0.04 | 0.73 |
| 62 | -1.20 | 0.01 | 0.26 | 0.14 | 0.95 | -0.31 | 0.62 | -0.12 | -0.53 | -0.25 | -0.45 | -0.13 | -0.25 | -0.02 | 0.99 | 0.77 |
| 63 | -328.23 | -1.12 | -114.03 | -32.94 | -158.16 | -63.37 | 11.37 | 28.20 | 83.28 | 124.73 | 87.52 | -30.45 | 42.75 | -283.00 | 197.73 | 165.80 |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 65 | -134.18 | 0.00 | 0.63 | 1.25 | -0.05 | -0.04 | 0.18 | -25.59 | -0.77 | -66.10 | -304.03 | 1.56 | 0.02 | 1.66 | -0.85 | -0.76 |
| 66 | -1933.14 | 6.44 | 86.87 | -1073.04 | 31.65 | -69.36 | 227.51 | 3.60 | 18.38 | 285.47 | 250.06 | 429.47 | 84.94 | -220.39 | -863.34 | -808.50 |
| 67 | -0.76 | 0.00 | 0.27 | -0.83 | 0.29 | -0.17 | -0.32 | 0.06 | -0.84 | -0.88 | 0.15 | -0.08 | -0.12 | -0.04 | 1.43 | 0.47 |
| 70 | -17.89 | 0.00 | 0.57 | -1.43 | 0.16 | 0.55 | -1.24 | -23.50 | -4.35 | -1.07 | -0.23 | -2.62 | -0.20 | -1.41 | 4.50 | 1.81 |

**Table S2.** Covariates' coefficients of the single-species GLMs for Sumatra. Covariates acronyms: I (intercept), TN (camera trap nights), CTI_FM (compound topographic index FM 500m), CTI_SD (compound topographic index SD 4,000m), E_FM (elevation FM 250m), HF_SD (human footprint SD 10,000m), LF_GY (lowland forests GYRATE_AM 10,000m), LO_GY (lowland open areas GYRATE_AM 8,000m), FL_PL (forest loss PLAND 4,000m), MM_GY (montane mosaic areas GYRATE_AM 2,000m), PP_GY (palm plantations GYRATE_AM 10,000m), PR_GY (plantation and regrowth GYRATE_AM 1,000m), PA_GY (protected areas GYRATE_AM 8,000m), R_SD (roughness SD 10,000m), SP_FM (slope position FM 10,000m), UF_PL (upper-montane forests PLAND 4,000m). See Table S3 for species' codes.

| Species' code | I | TN | CTI_FM | CTI_SD | E_FM | HF_SD | LF_GY | LO_GY | FL_PL | MM_GY | PP_GY | PR_GY | PA_GY | R_SD | SP_FM | UF_PL | X |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | -39.25 | 0.03 | 0.52 | -1.19 | 3.24 | -3.81 | 25.15 | 0.34 | 8.28 | -4.04 | -7.64 | -5.57 | -2.11 | 8.37 | -3.28 | -16.65 | -5.27 |
| 3 | -0.97 | 0.01 | -0.07 | 0.08 | -0.17 | 0.48 | 0.47 | 0.68 | -0.27 | -0.08 | -0.19 | 0.34 | 0.05 | -0.39 | -0.11 | 0.60 | 1.31 |
| 4 | -1.69 | 0.01 | -0.05 | -0.03 | -0.50 | -0.33 | -0.09 | 0.14 | 0.21 | 0.34 | 0.04 | 0.82 | 0.19 | -0.30 | -0.09 | 0.16 | 0.00 |
| 5 | -1.76 | 0.01 | -0.23 | -0.51 | -0.16 | 0.16 | 0.61 | 0.65 | -0.74 | -0.09 | 0.11 | 0.43 | 0.35 | -0.14 | -0.22 | 0.74 | 1.93 |
| 6 | -2.27 | 0.00 | 0.83 | -0.99 | 1.17 | -0.30 | 0.01 | -0.01 | -0.26 | 0.15 | 0.18 | 0.39 | -0.49 | 0.41 | 0.22 | -2.40 | 0.50 |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | -3.93 | 0.02 | 0.32 | 0.00 | 0.25 | -0.15 | 0.00 | 0.61 | -0.03 | -0.01 | 0.67 | 0.06 | 0.26 | 0.06 | 0.29 | -0.60 | -0.07 |
| 9 | -179.41 | 1.24 | 10.96 | 26.42 | -18.26 | 13.74 | 38.12 | 5.15 | -22.93 | 20.76 | 19.67 | 2.72 | 6.12 | 2.76 | 1.83 | 23.41 | -13.33 |
| 10 | -1.10 | 0.02 | 0.02 | -0.34 | 0.28 | 0.04 | -0.01 | 0.12 | -0.18 | -0.08 | 0.12 | -0.08 | -0.06 | 0.11 | 0.10 | -0.86 | 0.22 |
| 12 | -3.74 | 0.03 | 0.04 | -0.19 | 0.72 | -0.05 | -0.54 | 0.61 | 0.38 | -0.19 | -0.05 | 0.43 | 0.92 | -0.95 | 0.06 | -0.06 | -0.19 |
| 14 | -3.76 | 0.00 | -0.19 | 0.02 | 1.10 | 0.21 | 0.71 | -1.40 | 1.33 | 0.09 | -1.77 | -0.89 | -0.39 | 0.45 | -0.12 | -0.03 | -0.35 |
| 15 | -3.08 | 0.01 | -0.03 | 0.13 | 0.27 | -0.24 | 0.14 | 0.20 | -0.88 | -0.54 | 0.27 | -0.76 | 0.21 | 0.09 | 0.22 | -0.39 | -0.45 |
| 16 | -15.65 | 0.06 | 1.75 | -2.81 | -1.05 | -0.61 | -3.77 | 0.62 | -0.45 | 0.13 | -0.34 | -0.95 | -0.70 | -0.25 | -0.85 | -9.61 | 2.51 |
| 17 | -13.82 | 0.09 | 0.39 | -1.15 | 0.85 | 0.42 | 0.67 | -0.70 | -2.89 | 0.43 | 1.02 | 2.17 | -0.74 | 1.38 | 0.00 | 0.35 | 4.52 |
| 19 | -7.51 | 0.04 | 0.24 | -0.28 | 0.59 | -0.24 | 0.04 | -0.99 | -0.21 | -0.21 | -0.14 | 1.01 | 0.68 | 0.02 | 0.09 | -0.69 | 0.83 |
| 20 | -6.05 | 0.03 | 0.97 | -0.33 | 2.25 | 0.58 | 1.61 | -1.22 | 0.73 | 0.17 | -1.24 | -0.15 | -0.70 | 0.31 | 0.00 | -0.19 | 0.30 |
| 23 | -6.84 | 0.02 | 0.32 | -0.50 | 0.88 | -1.42 | 0.98 | 0.22 | -0.03 | -1.07 | 0.57 | -0.60 | 0.74 | 1.17 | -0.24 | -2.61 | -1.73 |
| 29 | -6.56 | 0.03 | 0.16 | 0.40 | 1.18 | 0.11 | 0.56 | 0.30 | -2.98 | 0.02 | 0.45 | 0.60 | -0.38 | 0.80 | -0.02 | -1.24 | 0.99 |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 30 | -1.72 | -0.01 | -0.52 | 0.19 | 1.43 | 0.40 | 0.67 | 0.44 | 0.16 | 0.00 | -0.02 | -1.22 | -0.34 | -0.38 | 0.20 | -0.21 | -0.32 |
| 31 | -3.08 | 0.01 | -0.17 | 0.22 | 1.47 | 0.57 | 0.39 | 0.66 | -0.14 | -0.10 | 0.26 | -0.39 | 0.29 | -0.59 | 0.04 | -1.80 | 0.22 |
| 32 | -5.23 | 0.02 | 0.08 | -0.42 | -0.19 | -0.76 | -0.80 | 0.12 | 0.24 | -0.12 | -0.01 | -0.85 | 1.24 | -0.45 | -0.19 | -0.31 | -0.41 |
| 33 | -2.91 | 0.01 | 0.02 | 0.44 | 0.73 | 0.29 | 0.53 | 0.30 | -0.35 | 0.25 | 0.25 | -0.62 | -0.14 | 0.00 | 0.20 | -0.17 | -0.11 |
| 35 | -25.83 | -0.01 | 0.01 | 0.13 | -1.59 | -0.38 | -0.68 | -0.54 | 0.24 | 0.84 | -49.21 | -0.41 | 2.89 | -0.66 | -0.17 | -4.55 | 0.57 |
| 36 | -4.20 | 0.00 | -0.37 | -0.15 | 0.49 | -1.62 | 1.56 | -0.01 | 0.13 | 0.70 | -0.84 | -0.10 | -0.55 | 1.05 | -0.10 | -0.08 | -0.36 |
| 37 | -2.04 | 0.01 | -0.11 | -0.23 | 0.53 | 0.17 | 0.28 | 0.18 | 0.12 | 0.18 | 0.05 | -0.86 | -0.06 | -0.22 | 0.08 | -0.77 | 0.15 |
| 39 | -5.26 | 0.03 | 0.01 | 0.48 | 0.75 | -0.10 | 0.23 | 0.06 | 0.26 | -0.12 | -0.55 | -0.82 | 0.36 | -0.13 | 0.24 | -0.17 | 0.02 |
| 45 | -3.37 | 0.02 | -0.07 | 0.25 | 0.97 | -0.46 | 0.82 | -0.32 | 0.50 | 0.18 | -0.27 | -0.39 | 0.34 | 0.22 | 0.22 | -0.60 | 0.27 |
| 46 | -1267.89 | 9.28 | -124.10 | 34.01 | -395.25 | 59.69 | 136.06 | 10.72 | -18.16 | 62.44 | 81.37 | 70.68 | -48.35 | -54.75 | 32.79 | 309.05 | 157.25 |
| 48 | -1.79 | 0.02 | 0.05 | -0.35 | -0.29 | 0.38 | -0.09 | -0.35 | 0.04 | 0.14 | 0.02 | -0.45 | -0.09 | 0.05 | 0.15 | -0.25 | 0.39 |
| 49 | -4.90 | 0.01 | -0.04 | -0.44 | 0.60 | -0.47 | -1.10 | 0.38 | 0.73 | -0.17 | -0.05 | -0.04 | 0.53 | -0.91 | -0.39 | -1.11 | 1.29 |
| 51 | -2.15 | -0.01 | 1.18 | -1.23 | 1.44 | -0.49 | 0.76 | -0.42 | 0.23 | 0.04 | 0.16 | 0.54 | -0.84 | 0.73 | -0.29 | -1.07 | 0.58 |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 52 | -7.36 | -0.02 | 0.52 | 0.76 | -0.27 | -0.74 | -1.09 | -0.45 | -1.47 | -9.94 | 0.30 | 2.31 | 0.60 | -1.31 | -0.90 | -0.32 | -0.68 |
| 54 | -8.74 | 0.06 | -0.98 | 0.47 | -1.08 | 0.35 | -0.20 | -1.05 | -0.41 | -0.45 | -0.05 | 0.68 | -0.32 | 0.35 | 0.31 | -0.33 | 1.10 |
| 55 | -0.43 | 0.02 | -0.15 | 0.12 | -0.55 | 0.02 | -0.45 | -0.13 | 0.16 | -0.21 | -0.11 | 0.03 | 0.24 | -0.60 | -0.01 | -0.20 | 0.18 |
| 56 | -47.34 | 0.02 | -0.22 | 3.97 | -5.24 | 2.60 | 0.06 | -0.46 | -2.24 | -11.22 | -47.41 | 3.67 | 0.96 | -2.90 | 0.04 | -42.55 | 0.52 |
| 58 | -43.87 | 0.01 | -0.59 | 4.78 | 3.05 | -2.88 | -3.83 | -29.77 | -1.49 | -10.55 | -1.05 | 4.33 | -4.75 | 1.29 | -2.93 | -0.42 | 8.40 |
| 59 | -15.23 | -0.02 | -0.21 | -0.73 | -1.06 | -1.26 | 3.57 | -0.96 | -2.59 | 1.88 | 4.44 | -1.57 | 12.68 | 1.36 | 0.99 | 0.83 | 3.53 |
| 61 | -162.32 | 0.08 | -0.01 | -2.90 | 3.50 | -2.42 | 4.58 | -1.42 | 0.51 | -11.86 | -2.94 | 1.52 | -3.97 | 3.60 | 0.58 | -508.19 | 0.95 |
| 62 | -1.53 | 0.01 | 0.24 | -0.66 | -0.02 | 0.35 | 0.62 | -0.52 | -0.47 | 0.12 | -0.09 | -0.85 | -0.44 | 0.20 | 0.45 | -1.63 | 0.25 |
| 64 | -25.15 | -0.03 | 0.67 | -0.81 | -1.47 | 2.17 | 0.65 | -0.84 | -5.35 | 0.60 | -47.02 | 0.69 | -2.40 | -0.20 | 1.68 | 1.11 | 0.61 |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 68 | -56.83 | 0.54 | 0.52 | 0.67 | 3.83 | -5.29 | -0.09 | 1.27 | 1.97 | 15.99 | NA | 3.32 | 0.83 | 2.31 | -0.44 | 0.00 | 1.16 |
| 69 | -18.72 | 0.00 | 0.83 | 0.41 | 1.54 | 0.06 | 1.18 | -10.89 | -3.92 | -0.18 | 0.00 | 1.60 | 0.25 | 1.16 | 0.08 | 0.03 | 0.97 |

**Table S3.** Species sampled in Borneo and Sumatra with relative identification codes and detections.

| Scientific name | Common name | Class | Order | Family | Genus | Code | Borneo Detections | Sumatra Detections |
|---|---|---|---|---|---|---|---|---|
| *Elephas maximus* | Asian elephant | Mammalia | Proboscidea | Elephantidae | Elephas | 1 | 456 | 0 |
| *Echinosorex gymnura* | Moonrat | Mammalia | Eulipotyphla | Erinaceidae | Echinosorex | 2 | 825 | 22 |
| Pig species | NA | Mammalia | Cetartiodactyla | Suidae | NA | 3 | 0 | 1,909 |
| *Sus scrofa* | Wild boar | Mammalia | Cetartiodactyla | Suidae | Sus | 4 | NA | 459 |
| *Sus barbatus* | Bearded pig | Mammalia | Cetartiodactyla | Suidae | Sus | 5 | 7,836 | 1,456 |
| Mouse-deer species | NA | Mammalia | Cetartiodactyla | Tragulidae | NA | 6 | 2,364 | 186 |
| *Tragulus kanchil* | Lesser oriental chevrotain | Mammalia | Cetartiodactyla | Tragulidae | Tragulus | 7 | 2,310 | 91 |
| *Tragulus napu* | Greater oriental chevrotain | Mammalia | Cetartiodactyla | Tragulidae | Tragulus | 8 | 1,646 | 0 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Muntjac species | NA | Mammalia | Cetartiodactyla | Cervidae | Muntiacus | 9 | 865 | 10 |
| *Muntiacus muntjak* | Southern red muntjac | Mammalia | Cetartiodactyla | Cervidae | Muntiacus | 10 | 3,817 | 1,293 |
| *Muntiacus atherodes* | Bornean yellow muntjac | Mammalia | Cetartiodactyla | Cervidae | Muntiacus | 11 | 5,178 | NA |
| *Rusa unicolor* | Sambar | Mammalia | Cetartiodactyla | Cervidae | Rusa | 12 | 2,276 | 253 |
| *Bos javanicus* | Banteng | Mammalia | Cetartiodactyla | Bovidae | Bos | 13 | 106 | NA |
| *Capricornis sumatraensis* | Mainland serow | Mammalia | Cetartiodactyla | Bovidae | Capricornis | 14 | NA | 114 |
| *Tapirus indicus* | Malay tapir | Mammalia | Perissodactyla | Tapiridae | Tapirus | 15 | NA | 323 |
| *Manis javanica* | Sunda pangolin | Mammalia | Pholidota | Manidae | Manis | 16 | 120 | 13 |
| Civet species | NA | Mammalia | Carnivora | Viverridae | NA | 17 | 0 | 16 |
| *Arctogalidia trivirgata* | Small-toothed palm civet | Mammalia | Carnivora | Viverridae | Arctogalidia | 18 | 6 | 0 |
| *Arctictis binturong* | Binturong | Mammalia | Carnivora | Viverridae | Arctictis | 19 | 55 | 31 |
| *Paguma larvata* | Masked palm civet | Mammalia | Carnivora | Viverridae | Paguma | 20 | 215 | 121 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Paradoxurus hermaphroditus* | Common palm civet | Mammalia | Carnivora | Viverridae | Paradoxurus | 21 | 495 | 0 |
| *Cynogale bennettii* | Otter civet | Mammalia | Carnivora | Viverridae | Cynogale | 22 | 37 | 0 |
| *Hemigalus derbyanus* | Banded civet | Mammalia | Carnivora | Viverridae | Hemigalus | 23 | 1,804 | 30 |
| *Diplogale hosei* | Hose's civet | Mammalia | Carnivora | Viverridae | Diplogale | 24 | 211 | NA |
| *Viverra tangalunga* | Malay civet | Mammalia | Carnivora | Viverridae | Viverra | 25 | 5,614 | 0 |
| Mongoose species | NA | Mammalia | Carnivora | Herpestidae | NA | 26 | 255 | 0 |
| *Herpestes brachyurus* | Short-tailed mongoose | Mammalia | Carnivora | Herpestidae | Herpestes | 27 | 680 | 0 |
| *Herpestes semitorquatus* | Collared mongoose | Mammalia | Carnivora | Herpestidae | Herpestes | 28 | 144 | 0 |
| *Prionodon linsang* | Banded linsang | Mammalia | Carnivora | Prionodontidae | Prionodon | 29 | 99 | 82 |
| *Neofelis diardi* | Sunda clouded leopard | Mammalia | Carnivora | Felidae | Neofelis | 30 | 546 | 237 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Panthera tigris* | Tiger | Mammalia | Carnivora | Felidae | Panthera | 31 | NA | 289 |
| *Pardofelis marmorata* | Marbled cat | Mammalia | Carnivora | Felidae | Pardofelis | 32 | 190 | 51 |
| *Catopuma temminckii* | Asiatic golden cat | Mammalia | Carnivora | Felidae | Catopuma | 33 | NA | 156 |
| *Catopuma badia* | Borneo bay cat | Mammalia | Carnivora | Felidae | Catopuma | 34 | 46 | NA |
| *Prionailurus javanensis* | Sunda leopard cat | Mammalia | Carnivora | Felidae | Prionailurus | 35 | 1,907 | 31 |
| *Cuon alpinus* | Dhole | Mammalia | Carnivora | Canidae | Cuon | 36 | NA | 51 |
| *Helarctos malayanus* | Sun bear | Mammalia | Carnivora | Ursidae | Helarctos | 37 | 901 | 501 |
| *Mydaus javanensis* | Sunda stink-badger | Mammalia | Carnivora | Mephitidae | Mydaus | 38 | 413 | NA |
| *Martes flavigula* | Yellow-throated marten | Mammalia | Carnivora | Mustelidae | Martes | 39 | 290 | 104 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Melogale everetti* | Bornean ferret badger | Mammalia | Carnivora | Mustelidae | Melogale | 40 | 8 | NA |
| *Mustela nudipes* | Malay weasel | Mammalia | Carnivora | Mustelidae | Mustela | 41 | 29 | 0 |
| Otter species | NA | Mammalia | Carnivora | Mustelidae | NA | 42 | 11 | 0 |
| *Aonyx cinereus* | Asian small-clawed otter | Mammalia | Carnivora | Mustelidae | Aonyx | 43 | 19 | 0 |
| *Lutrogale perspicillata* | Smooth-coated otter | Mammalia | Carnivora | Mustelidae | Lutrogale | 44 | 16 | 0 |
| Rat species | NA | Mammalia | Rodentia | NA | NA | 45 | 2,009 | 172 |
| *Trichys fasciculata* | Long-tailed porcupine | Mammalia | Rodentia | Hystricidae | Trichys | 46 | 740 | 15 |
| *Hystrix crassispinis* | Thick-spined porcupine | Mammalia | Rodentia | Hystricidae | Hystrix | 47 | 529 | NA |
| *Hystrix brachyura* | Malayan porcupine | Mammalia | Rodentia | Hystricidae | Hystrix | 48 | 1,885 | 772 |
| Squirrel species | NA | Mammalia | Rodentia | Sciuridae | NA | 49 | 325 | 28 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Rheithrosciurus macrotis* | Tufted ground squirrel | Mammalia | Rodentia | Sciuridae | Rheithrosciurus | 50 | 223 | NA |
| Treeshrew species | NA | Mammalia | Scandentia | NA | NA | 51 | 521 | 64 |
| Langur species | NA | Mammalia | Primates | Cercopithecidae | NA | 52 | 0 | 6 |
| *Presbytis rubicunda* | Red langur | Mammalia | Primates | Cercopithecidae | Presbytis | 53 | 11 | NA |
| *Presbytis melalophos* | Yellow-handed mitered langur | Mammalia | Primates | Cercopithecidae | Presbytis | 54 | NA | 36 |
| *Macaca nemestrina* | Southern pig-tailed macaque | Mammalia | Primates | Cercopithecidae | Macaca | 55 | 4,698 | 1,994 |
| *Macaca fascicularis* | Common long-tailed macaque | Mammalia | Primates | Cercopithecidae | Macaca | 56 | 965 | 64 |
| *Pongo pygmaeus* | Bornean orangutan | Mammalia | Primates | Hominidae | Pongo | 57 | 329 | NA |
| Pheasant species | NA | Aves | Galliformes | Phasianidae | NA | 58 | 12 | 10 |
| Partridge species | NA | Aves | Galliformes | Phasianidae | NA | 59 | 19 | 9 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Arborophila graydoni* | Sabah partridge | Aves | Galliformes | Phasianidae | Arborophila | 60 | 41 | NA |
| *Rollulus rouloul* | Crested partridge | Aves | Galliformes | Phasianidae | Rollulus | 61 | 144 | 13 |
| *Argusianus argus* | Great argus | Aves | Galliformes | Phasianidae | Argusianus | 62 | 1,956 | 1,466 |
| *Haematortyx sanguiniceps* | Crimson-headed partridge | Aves | Galliformes | Phasianidae | Haematortyx | 63 | 11 | NA |
| *Polyplectron chalcurum* | Bronze-tailed peacock-pheasant | Aves | Galliformes | Phasianidae | Polyplectron | 64 | NA | 46 |
| *Lophura bulweri* | Bulwer's pheasant | Aves | Galliformes | Phasianidae | Lophura | 65 | 424 | NA |
| *Lophura pyronota* | Bornean crestless fireback | Aves | Galliformes | Phasianidae | Lophura | 66 | 8 | NA |
| *Lophura ignita* | Bornean crested fireback | Aves | Galliformes | Phasianidae | Lophura | 67 | 1,235 | NA |
| *Lophura rufa* | Malay crested fireback | Aves | Galliformes | Phasianidae | Lophura | 68 | NA | 19 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Lophura inornata* | Salvadori's pheasant | Aves | Galliformes | Phasianidae | Lophura | 69 | NA | 20 |
| *Carpococcyx radiceus* | Bornean ground-cuckoo | Aves | Cuculiformes | Cuculidae | Carpococcyx | 70 | 27 | NA |
| Domestic cats | NA | NA | NA | NA | NA | 71 | 26 | 0 |
| Domestic cattle | NA | NA | NA | NA | NA | 72 | 20 | 0 |
| Domestic dogs | NA | NA | NA | NA | NA | 73 | 28 | 110 |
| Humans | NA | NA | NA | NA | NA | 74 | 60 | 1,668 |

**Table S4.** Borneo biplot covariates' scores. Shown are the first two CCA axes.

| Covariates | CCA1 | CCA2 |
|---|---|---|
| Camera trap nights | -0.680 | 0.080 |
| Compound topographic index SD 4,000 | 0.359 | -0.062 |
| Elevation FM 250 | -0.879 | 0.185 |
| Human footprint FM 10,000 | -0.103 | 0.441 |
| Human footprint SD 2,000 | 0.472 | -0.063 |
| Lowland forests GYRATE_AM 8,000 | -0.415 | 0.075 |
| Lowland open areas GYRATE_AM 2,000 | 0.586 | -0.178 |
| Forest loss PLAND 10,000 | 0.556 | -0.337 |
| Mangroves PLAND 10,000 | 0.294 | 0.006 |
| Palm plantations PLAND 2,000 | 0.479 | 0.133 |
| Protected areas PLAND 4,000 | -0.500 | 0.656 |
| Slope position FM 8,000 | -0.646 | 0.109 |
| Tree canopy cover reclassified PD 4,000 | 0.768 | -0.053 |
| X | 0.105 | 0.316 |
| $X^2*Y^2$ | -0.018 | 0.157 |

**Table S5.** Sumatra biplot covariates' scores. Shown are the first two CCA axes.

| Covariates | CCA1 | CCA2 |
|---|---|---|
| Camera trap nights | -0.658 | -0.106 |
| Compound topographic index FM 500 | 0.296 | -0.292 |
| Compound topographic index SD 4,000 | 0.560 | 0.021 |
| Elevation FM 250 | -0.766 | 0.486 |
| Human footprint SD 10,000 | -0.161 | -0.123 |
| Lowland forests GYRATE_AM 10,000 | 0.396 | -0.141 |
| Lowland open areas GYRATE_AM 8,000 | 0.674 | -0.038 |
| Forest loss PLAND 4,000 | 0.738 | -0.069 |
| Montane mosaic areas GYRATE_AM 2,000 | -0.083 | 0.034 |
| Palm plantations GYRATE_AM 10,000 | -0.359 | -0.256 |
| Plantation and regrowth GYRATE_AM 1,000 | 0.797 | 0.028 |
| Protected areas GYRATE_AM 8,000 | 0.246 | 0.223 |
| Roughness SD 10,000 | 0.423 | 0.472 |
| Slope position FM 10,000 | -0.453 | 0.158 |
| Upper-montane forests PLAND 4,000 | -0.259 | 0.881 |
| X | 0.694 | 0.023 |

**Table S6.** Most efficient systems of protected areas in preserving predicted species richness in Borneo.

| Code | Protected areas | % of total biodiversity | % of protected biodiversity |
|---|---|---|---|
| B1 | Sebangau National Park | 1.20 | 12.94 |
| B2 | Kutai National Park | 0.82 | 8.91 |
| B3 | Crocker Range National Park | 0.63 | 6.84 |
| B4 | Tabin Wildlife Reserve - Mount Hatton Protection Forest Reserve | 0.61 | 6.62 |
| B5 | Ulu Kalumpang Protection Forest Reserve - Tawau Hill Park | 0.40 | 4.33 |
| B6 | Kayan Mentarang National Park | 0.35 | 3.76 |
| B7 | Kinabalu Park | 0.33 | 3.62 |
| B8 | Bukit Baka - Bukit Raya National Park | 0.33 | 3.53 |
| B9 | Gunung Mulu National Park – Labi Hills Forest Reserve | 0.32 | 3.45 |
| B10 | Bukit Sapat Hawung Nature Reserve | 0.31 | 3.32 |

**Table S7.** Most efficient systems of protected areas in preserving predicted species richness in Sumatra.

| Code | Protected areas | % of total biodiversity | % of protected biodiversity |
|---|---|---|---|
| S1 | Kerinci Seblat National Park | 8.17 | 44.98 |
| S2 | Bukit Barisan Selatan National Park | 1.99 | 10.95 |
| S3 | Berbak Sembilang National Park | 1.44 | 7.91 |
| S4 | Bukit Rimbang Bukit Baling Wildlife Reserve | 0.98 | 5.41 |
| S5 | Bukit Tiga Puluh National Park | 0.94 | 5.16 |
| S6 | Padang Sugihan Wildlife Reserve | 0.45 | 2.50 |
| S7 | Kerumutan Wildlife Reserve | 0.43 | 2.34 |
| S8 | Tesso Nilo National Park | 0.36 | 1.98 |
| S9 | Giam Siak Kecil Wildlife Reserve | 0.27 | 1.50 |
| S10 | Bukit Dua Belas National Park | 0.26 | 1.42 |

**Table S8.** Most biodiverse areas falling outside the current system of protected areas in Borneo and Sumatra, and relative amount of cumulative species richness that they harbour.

| Borneo | | | Sumatra | | |
|---|---|---|---|---|---|
| Code | Area ($km^2$) | % of total biodiversity | Code | Area ($km^2$) | % of total biodiversity |
| BU1 | 5,631 | 2.43 | SU1 | 843 | 0.58 |
| BU2 | 1,813 | 0.78 | SU2 | 550 | 0.39 |
| BU3 | 1,716 | 0.73 | SU3 | 484 | 0.35 |
| BU4 | 1,284 | 0.54 | SU4 | 426 | 0.30 |
| BU5 | 1,155 | 0.49 | SU5 | 199 | 0.14 |
| BU6 | 538 | 0.23 | SU6 | 175 | 0.12 |
| BU7 | 511 | 0.22 | SU7 | 166 | 0.12 |
| BU8 | 373 | 0.16 | SU8 | 169 | 0.12 |
| BU9 | 365 | 0.15 | SU9 | 178 | 0.12 |
| BU10 | 324 | 0.14 | SU10 | 138 | 0.09 |

**Table S9.** Indicator species for Bornean and Sumatran biodiversity. Shown are the residual deviance values of the single-species models. Species are ranked by their residual deviance values.

| Borneo | | Sumatra | |
|---|---|---|---|
| Species | Residual deviance | Species | Residual deviance |
| *Tragulus kanchil* | 1.32E+05 | *Prionailurus javanensis* | 4.93E+04 |
| Rat species | 1.34E+05 | Rat species | 5.34E+04 |
| *Echinosorex gymnura* | 1.43E+05 | *Polyplectron chalcurum* | 5.53E+04 |
| *Hemigalus derbyanus* | 1.50E+05 | *Paguma larvata* | 6.50E+04 |
| *Mydaus javanensis* | 1.51E+05 | *Neofelis diardi* | 6.51E+04 |
| *Hystrix crassispinis* | 1.64E+05 | *Echinosorex gymnura* | 6.57E+04 |
| *Viverra tangalunga* | 1.66E+05 | *Cuon alpinus* | 6.58E+04 |
| *Arborophila graydoni* | 1.71E+05 | Partridge species | 6.60E+04 |
| Partridge species | 1.89E+05 | *Sus barbatus* | 6.62E+04 |
| *Herpestes brachyurus* | 2.19E+05 | *Capricornis sumatraensis* | 6.68E+04 |
| *Lophura ignita* | 2.23E+05 | *Helarctos malayanus* | 6.78E+04 |
| *Argusianus argus* | 2.33E+05 | *Tapirus indicus* | 6.84E+04 |
| *Muntiacus atherodes* | 2.35E+05 | *Hemigalus derbyanus* | 7.05E+04 |
| *Paradoxurus hermaphroditus* | 2.43E+05 | *Martes flavigula* | 7.05E+04 |

| | | | |
|---|---|---|---|
| *Pardofelis marmorata* | 2.48E+05 | *Argusianus argus* | 7.15E+04 |
| Treeshrew species | 2.55E+05 | *Macaca nemestrina* | 7.18E+04 |
| *Carpococcyx radiceus* | 2.62E+05 | *Panthera tigris* | 7.25E+04 |
| *Aonyx cinereus* | 2.73E+05 | *Catopuma temminckii* | 7.28E+04 |
| *Rusa unicolor* | 2.97E+05 | *Prionodon linsang* | 7.58E+04 |
| *Mustela nudipes* | 3.05E+05 | Pig species | 7.83E+04 |
| Mouse-deer species | 3.06E+05 | *Macaca fascicularis* | 7.83E+04 |
| *Melogale everetti* | 3.08E+05 | Muntjac species | 7.94E+04 |
| *Prionailurus javanensis* | 3.14E+05 | Civet species | 8.00E+04 |
| *Manis javanica* | 3.14E+05 | Mouse-deer species | 8.03E+04 |
| *Helarctos malayanus* | 3.18E+05 | *Sus scrofa* | 8.22E+04 |
| Mongoose species | 3.26E+05 | *Manis javanica* | 8.30E+04 |
| *Lutrogale perspicillata* | 3.29E+05 | *Arctictis binturong* | 8.30E+04 |
| Squirrel species | 3.30E+05 | Treeshrew species | 8.30E+04 |
| *Sus barbatus* | 3.33E+05 | *Presbytis melalophos* | 8.32E+04 |
| *Hystrix brachyura* | 3.33E+05 | *Trichys fasciculata* | 8.35E+04 |
| *Herpestes semitorquatus* | 3.34E+05 | *Pardofelis marmorata* | 8.36E+04 |
| *Martes flavigula* | 3.36E+05 | *Rusa unicolor* | 8.37E+04 |
| *Neofelis diardi* | 3.36E+05 | Squirrel species | 8.38E+04 |
| Muntjac species | 3.36E+05 | *Hystrix brachyura* | 8.38E+04 |
| *Cynogale bennettii* | 3.40E+05 | Pheasant species | 8.39E+04 |

| | | | |
|---|---|---|---|
| *Haematortyx sanguiniceps* | 3.40E+05 | Langur species | 8.42E+04 |
| *Macaca fascicularis* | 3.44E+05 | *Rollulus rouloul* | 8.43E+04 |
| *Macaca nemestrina* | 3.49E+05 | *Lophura inornata* | 8.43E+04 |
| *Tragulus napu* | 3.50E+05 | *Tragulus kanchil* | 8.43E+04 |
| *Lophura pyronota* | 3.51E+05 | *Muntiacus muntjak* | 8.43E+04 |
| *Arctictis binturong* | 3.51E+05 | *Lophura rufa* | 8.43E+04 |
| *Pongo pygmaeus* | 3.52E+05 | | |
| *Prionodon linsang* | 3.52E+05 | | |
| *Presbytis rubicunda* | 3.52E+05 | | |
| *Trichys fasciculata* | 3.52E+05 | | |
| *Paguma larvata* | 3.52E+05 | | |
| *Bos javanicus* | 3.52E+05 | | |
| *Catopuma badia* | 3.52E+05 | | |
| *Diplogale hosei* | 3.52E+05 | | |
| *Arctogalidia trivirgata* | 3.52E+05 | | |
| Otter species | 3.52E+05 | | |
| *Muntiacus muntjak* | 3.52E+05 | | |
| *Elephas maximus* | 3.52E+05 | | |
| *Rheithrosciurus macrotis* | 3.52E+05 | | |
| Pheasant species | 3.52E+05 | | |
| *Rollulus rouloul* | 3.52E+05 | | |
| *Lophura bulweri* | 3.52E+05 | | |

# Chapter 4

# Not seeing the forest for the trees: generalised linear model out-performs random forest in species distribution modelling for Southeast Asian felids.

Luca Chiaverini, David W. Macdonald, Andrew J. Hearn, Żaneta Kaszta, Eric Ash, Helen M. Bothwell, Özgün Emre Can, Phan Channa, Gopalasamy Reuben Clements, Iding Achmad Haidir, Pyae Phyoe Kyaw, Jonathan H. Moore, Akchousanh Rasphone, Cedric Kai Wei Tan, Samuel A. Cushman

Published in Ecological Informatics. (2023) 75: 102026

https://doi.org/10.1016/j.ecoinf.2023.102026

**Abstract**

Species Distribution Models (SDMs) are a powerful tool to derive habitat suitability predictions relating species occurrence data with habitat features. Two of the most frequently applied algorithms to model species-habitat relationships are Generalised Linear Models (GLM) and Random Forest (RF). The former is a parametric regression model providing functional models with direct interpretability. The latter is a machine learning non-parametric algorithm, more tolerant than other approaches in its assumptions, which has often been shown to outperform parametric algorithms. Other approaches have been developed to produce robust SDMs, like training data bootstrapping and spatial scale optimisation. Using felids presence-absence data from three study regions in Southeast Asia (mainland, Borneo and Sumatra), we tested the performances of SDMs by implementing four modelling frameworks: GLM and RF with bootstrapped and non-bootstrapped training data. With Mantel and ANOVA tests we explored how the four combinations of algorithms and bootstrapping influenced SDMs and their predictive performances. Additionally, we tested how scale-optimisation responded to species' size, taxonomic associations (species and genus), study area and algorithm. We found that choice of algorithm had strong effect in determining the differences between SDMs' spatial predictions, while bootstrapping had no effect. Additionally, algorithm followed by study area and species, were the main factors driving differences in the spatial scales identified. SDMs trained with GLM showed higher predictive performance, however, ANOVA tests revealed that algorithm had significant effect only in explaining the variance observed in sensitivity and specificity and, when interacting with bootstrapping, in Percent Correctly Classified (PCC). Bootstrapping significantly explained the variance in specificity, PCC and True Skills Statistics (TSS). Our results suggest that there are systematic differences in the scales identified and in the predictions produced by GLM vs. RF, but that neither approach was consistently better than the other.

The divergent predictions and inconsistent predictive abilities suggest that analysts should not assume machine learning is inherently superior and should test multiple methods. Our results have strong implications for SDM development, revealing the inconsistencies introduced by the choice of algorithm on scale optimisation, with GLM selecting broader scales than RF.

## 1. Introduction

Species distribution models (SDMs) are a powerful tool to derive spatially-explicit predictions of habitat suitability, by statistically relating species occurrence data at known locations with environmental features of those locations (Guisan and Thuiller 2005, Elith and Leathwick 2009). Several statistical approaches are available to produce SDMs (Guisan et al. 2017, Valavi et al. 2022), with regression methods (Guisan et al. 2002) in particular applied in the earlier stages of the SDMs' development, and machine learning methods receiving more attention in recent years (e.g., Evans and Cushman (2009); Cushman et al. (2017); Stupariu et al. (2021)).

Among regression models, generalised linear models (GLM) are extensions of linear model allowing non-linearity and non-constant variance of the data (Hastie and Tibshirani 1990, Guisan et al. 2002). Consequently, GLM can deal with different families of data, making them an appropriate tool for analysing species-habitat relationships, which rarely are represented by normal distributions (Austin 1987). GLM have long been the predominant algorithm used in SDMs (Guisan and Zimmermann 2000, Guisan and Thuiller 2005, Elith and Leathwick 2009, McGarigal et al. 2016).

Among machine learning techniques, random forest (RF; Breiman (2001)) is the most commonly used algorithm (Stupariu et al. 2021). RF are Classification and Regression Tree (CART) based ensemble approaches, which overcome limitations of CARTs such as model over-fitting by creating multiple subsets of the training data by random sampling with replacement, a process known as bagging (bootstrap aggregating), and by training separate decision trees with each subset. Then, the predictions of all the decision trees are combined by averaging (for regressions) or voting (for classification), helping the model to stabilise, reducing in this way the impact of any individual decision tree that could have over-fitted the

data. Additionally, RF can readily accommodate nonlinear relationships between predictor and response variables, and can handle large numbers of predictor variables, although when RF are trained with correlated variables, these tend to have lower variables importance compared to uncorrelated variables, leading to potential underestimate of the importance of correlated variables in the model, an effect known as variable inflation (Breiman 2001, Evans and Cushman 2009, Kumar et al. 2021). Furthermore, by permutation of variables, RF provides accessible measures of variable importance (Cutler et al. 2007, Evans and Cushman 2009).

The main difference between machine learning techniques and likelihood based methods, such as GLM, is that the former aims to optimise the predictive accuracy of a model, while the latter focuses on estimating the parameters of a statistical model that best fit the observed data. The applications of GLM and other parametric models, however, are constrained by often neglected or violated assumptions, such as data independence. Therefore, non-parametric algorithms, like RF, in many cases show better performance in comparison to parametric methods (Cushman et al. 2017, Cushman and Wasserman 2018, Kumar et al. 2021, Valavi et al. 2022). However, there are applications where RF perform poorly, often because non-parametric algorithms have little control over the shape of the fitted function (Roberts et al. 2017). Additionally, RF's performances are sometimes weak when using presence-absence data with an imbalance in favour of absences (e.g., Evans and Cushman (2009); Freeman et al. (2012)). Therefore, comparing the performances of GLM and RF is an important and contemporary issue to assess for developing reliable SDMs.

To produce accurate SDMs, not just the effects of different algorithms should be investigated, but also the effects of resampling techniques on model's training data (Freedman 1981, Efron 1982). Specifically, bootstrapping (i.e., random subsampling with replacement) the training data has been shown to increase models' precision by providing a

combination of models, which reduces stochastic errors in estimation (e.g., Vaughan and Ormerod (2005); Hefley et al. (2014); Xu and Goodacre (2018)).

Additionally, species-habitat relationships occur at multiple spatial scales (Wiens 1989, Levin 1992). Therefore, scale optimisation has recently been an area of rapid development (Vergara et al. 2016, Wan et al. 2019, Chiaverini et al. 2021). Given that scale optimization is a critical step in the framework to produce reliable SDMs (McGarigal et al. 2016), it is important to understand how modelling algorithm and bootstrapping affect multiscale model optimisation.

Here, using presence-absence data of felid species sampled in three study regions in Southeast Asia (mainland, Borneo and Sumatra), we investigated nine hypotheses related to the performances and predictions of SDMs. Specifically, our hypotheses were distinguished in three categories, respectively related to the predictive performances of the models, to the correlation and the similarities between the habitat suitability models, and to the spatial scale selection. For the first category (i.e., predictive performances), we hypothesised that: (1) RF would outperform GLM in terms of their predictive performances; (2) predictive performances of the models trained with bootstrapped data would outperform non-bootstrapped models. For the second category (i.e., correlation between the habitat suitability models), we hypothesises that (3) models trained with the same algorithm would be more correlated; (4) bootstrapping would have relatively little effect on correlation of the SDMs. Finally, for the third category (i.e., spatial scale selection), we hypothesised that (5) species that are more similar in size would have more similarities in the representative spatial scales at which predictor variables were selected; (6) models for the same species would have more similar scales than models for different species; (7) models for the same genus would have more similar scales than models for different genera; (8) models for species from the same study area would have more similar scales than species from different study areas; (9) models

trained with the same algorithm would have more similar scales than models trained with a different algorithm.

## 2. Materials and methods

### 2.1.Data collection

Camera trap surveys were carried out in Southeast Asia between 2007 and 2016, covering the range of the two clouded leopard species (i.e., mainland clouded leopard (*Neofelis nebulosa*) in mainland Southeast Asia and Sunda clouded leopard (*Neofelis diardi*) in Borneo and Sumatra), which represented the original focus of our surveys. Sampling grids occurred mainly in protected areas, covering a wide altitudinal gradient and different environments. Grids consisted of approximately 80 camera stations placed 1.0-2.0km from each other, with two cameras per station at ~40cm above the ground, and deployed along forest trails and disused logging roads to maximise felids' detection success (Macdonald et al. 2018, Macdonald et al. 2019). Additionally, we incorporated opportunist camera trap data from studies conducted by Ash et al. (2021) to monitor tiger populations in five protected areas in Dong Phayayen-Khao Yai Forest Complex, in eastern Thailand.

Even though the surveys were designed for sampling clouded leopards, we also collected data for Asiatic golden cat (*Catopuma temminckii*), marbled cat (*Pardofelis marmorata*) and leopard cat (*Prionailurus bengalensis*) in mainland, Borneo bay cat (*Catopuma badia*), marbled cat and Sunda leopard cat (*Prionailurus javanensis*) in Borneo, and Asiatic golden cat, marbled cat and Sunda leopard cat in Sumatra. To minimise overestimation bias, we discounted records of the same species at the same camera trap stations within 1 hour, except when animals were individually recognisable and when sex and/or age classes were unambiguous.

### 2.2.Habitat covariates

We chose two preliminary sets of covariates to model habitat suitability for felids in the mainland and in the offshore extents of the study area, reflecting different habitat composition of the regions. Covariates were selected based on previous knowledge of felids' ecological requirements in Southeast Asia (Hearn et al. 2018, Macdonald et al. 2018, Macdonald et al. 2019), and included environmental, anthropogenic, topographic and climatic covariates (Tables S1-S3).

To investigate the scales at which sampled felids select habitat features, we assessed each covariate at eight spatial scales, by using circular buffers of varying radii, centred on each camera trap location. For mainland, we investigated representative spatial scales by using circular buffers of 250m, 500m, 1,000m, 2,000m, 4,000m, 8,000m, 16,000m and 32,000m. For Borneo and Sumatra we used buffers of 250m, 500m, 1,000m, 2,000m, 4,000m, 6,000m, 8,000m and 10,000m (Macdonald et al. 2020, Chiaverini et al. 2022).

At each of the aforementioned scales, we calculated biologically informative metrics by using FRAGSTATS (McGarigal et al. 2012) for discrete raster layers and ArcMap v10.6.1 (ESRI 2018) for continuous raster layers: in FRAGSTATS, we implemented percentage of landscape (PLAND) metric, while in ArcMap we calculated focal mean (FM). Both the functions average the pixel values of the covariates' raw raster layers to obtain a smoothened version of the original layer. Additionally, PLAND function calculates the percentage of the focal covariate within the circular buffer, and this is why it has been preferred for discrete layers. Finally, we removed poorly sampled covariates occurring at <10% of camera trap stations.

## 2.3.Modelling framework

### 2.3.1. Generalised Linear Model – Scale optimisation and covariates selection

To investigate the most representative spatial scales, we performed univariate generalised linear model (GLM), independently at each scale for each covariate, using the presence-absence data independently for each felid as the response variable. We subset the dataset by randomly selecting 80% of the camera traps to train the univariate GLM, leaving the remaining 20% for model validation. For each covariate, we selected the scale whose univariate GLM showed the lowest Akaike Information Criterion (AIC). We then checked the best-scaled covariates for multicollinearity by calculating Pearson's correlation index between all covariates pairs and, when two covariates were highly correlated (i.e., $|r| \geq 0.7$), by removing the one whose univariate GLM showed the lower adjusted-$R^2$ (Guisan and Zimmermann 2000).

### 2.3.2. Generalised Linear Model – Model training

We trained GLM independently for each felid species, by applying two different approaches. For the first approach (hereafter *non-bootstrapped GLM*), we trained the models by using the same dataset used to evaluate the best scales, composed of the 80% of the camera traps, and we produced species distribution probability surfaces for each sampled felid.

For the second approach (hereafter *bootstrapped GLM*) we subsampled the entire camera trap dataset by randomly selecting 20% of the presence and of the absence locations, trained a GLM and produced a species distribution probability surface. We reiterated this process 100 times with replacement, and produced the final species distribution probability surfaces by averaging the 100 probability surfaces, independently for each felid.

### 2.3.3. Generalised Linear Model – Model validation

We evaluated the performances of the non-bootstrapped GLM by calculating the Area Under the ROC Curve (AUC), Percent Correctly Classified (PCC), sensitivity, specificity and True Skill Statistic (TSS) using the 20% camera traps retained from the training dataset.

To evaluate the bootstrapped GLM, for each of the 100 iterations, we calculated AUC, PCC, sensitivity, specificity and TSS using the data retained from the model training, composed of 80% of the presence and of the absence locations. We obtained the final metrics by averaging the results of 100 validation iterations. We implemented this framework independently for each felid. Validation metrics were calculated in R v3.5.1 (R Core Team 2018) using the package *PresenceAbsence* (Freeman and Moisen 2007).

### 2.3.4. Random Forest – Scale optimisation and covariates selection

Independently for each felid, we investigated the most representative scale for each covariate by performing univariate Random Forest (RF), using the full camera traps dataset and the presence-absence data as response variable. We trained RF by using 9,999 bootstrap replicates and, for each covariate, we selected the scale whose univariate RF showed the lowest Out-Of-Bag (OOB) error. Other techniques have been developed to identify the most representative spatial scales for habitat suitability models trained with RF; specifically, Bradter et al. (2013) developed a technique that allows the simultaneous evaluation of several spatial scales of multiple predictors, which strongly diverges from our univariate approach. While acknowledging that other approaches might have led to different optimal scales, we decided to implement a univariate scale selection approach to allow the framework to be comparable to the one used for GLM.

### 2.3.5. Random Forest – Model training

Similar to the GLM approaches, for RF we used two different training data selection frameworks. Both the modelling frameworks were composed of two steps: 1) variables selection, implemented on the scale-optimised covariates by using 2,000 bootstrap replicates, and 2) RF training, implemented on the selected covariates by using 500 bootstrap replicates and by setting the number of randomly drawn candidate variables (*mtry*) to 10. For the first modelling framework (hereafter *non-bootstrapped RF*), we trained the RF by using the same training dataset used for the non-bootstrapped GLM, composed of randomly selected 80% of camera trap locations, and produced species distribution probability surfaces for each felid.

For the second approach (hereafter *bootstrapped RF*), we subsampled the entire camera taps dataset by randomly selecting 20% of the presence and of the absence locations. We then trained RF with the subsampled dataset and produced species distribution probability surface. We reiterated this framework 100 times, replacing each time the data used to train the models, and produced 100 different probability surfaces. We then obtained the final species distribution probability surfaces by averaging the 100 probability surfaces, independently for each felid.

### 2.3.6. Random Forest – Model validation

We evaluated the non-bootstrapped RF by calculating AUC, PCC, sensitivity, specificity and TSS using the 20% camera traps retained from the RF training.

To evaluate the bootstrapped RF, we used the retained dataset from the RF training, composed of the 80% of presence and absence locations, to calculate AUC, PCC, sensitivity, specificity and TSS for each iterations, independently for each species, using the package

*PresenceAbsence* in R v3.5.1. To obtain the final metrics, we averaged the metric results of the 100 iterations.

### 2.4.Effects of algorithm and bootstrapping on predictive performances

To evaluate whether the algorithm and the bootstrapping affected the predictive performances of the models, we performed ANOVA tests. We performed the tests independently for each performance metric we calculated, and tested the effect of algorithm, bootstrapping and combination of algorithm and bootstrapping.

### 2.5.Effects of algorithm and bootstrapping on models similarity

To investigate the effects of algorithm and bootstrapping on the similarities between species distribution surfaces, we performed Mantel tests in R v3.5.1 using the package *ecodist* (Goslee and Urban 2007). We first calculated Pearson's correlation indexes between the four models, for each species, for each study area, resulting in a total of twelve independent matrices (Tables S4-S15). Then, we calculated Mantel tests independently between these matrices and two binary matrices representing whether the models have been trained with GLM or RF (Table S16), and whether they have been bootstrapped or non-bootstrapped (Table S17). Finally, we averaged the results of the Mantel tests for each species, independently for algorithm and bootstrapping, to investigate their overarching effect on the species distribution probability surfaces.

### 2.6.Factors determining spatial scales selection

To evaluate the effects of different factors on the selection of the representative spatial scales, we performed Mantel tests in R v3.5.1 using the package *ecodist*. Specifically, we ran five simple Mantel tests between the difference in the ordinal scales among pairs of covariates across models and the difference in each of the factors we hypothesised to affect scale selection: body size (indexed to reflect the relative differences between species), species, genus, study area and algorithm.

Additionally, we ran eight partial Mantel tests separating the effects of the methodological factor (i.e., algorithm) and of the four ecological factors (i.e., body size, species, genus and study area). We ran four partial Mantel tests between the difference in ordinal scales and difference in algorithm, partialling out the difference in each of the four ecological factors, sequentially. Then, we also ran four partial Mantel tests between the difference in ordinal scales and difference in each of the four ecological factors, partialling out the difference in algorithm.

## 3. Results

### 3.1.Data collection

Camera trap grids were deployed in 15 sampling locations in seven countries in the mainland study area (Table S18) and in 22 sampling locations in the Sunda Islands (15 in Borneo and 7 in Sumatra; Table S19). This sampling effort yielded a total of 1,384 camera trap stations in mainland and 1,544 in the Sunda Islands (801 in Borneo and 743 in Sumatra). We achieved a combined sampling effort of 115,389 trap nights in the mainland (mean= 83.4±2.6SE) and of 138,515 trap nights in the Sunda Islands (77,687 in Borneo (mean= 97.0±2.1SE) and 60,828 in Sumatra (mean= 81.9±0.7SE)).

In the mainland we sampled clouded leopard at 235 camera trap stations, Asiatic golden cat at 148 stations, marbled cat at 116 stations and leopard cat at 345 stations (Table S20). In Borneo, we sampled Sunda clouded leopard at 154 camera trap stations, Borneo bay cat at 25 stations, marbled cat at 85 stations and Sunda leopard cat at 202 stations (Table S21). In Sumatra, we sampled Sunda clouded leopard at 101 camera trap stations, Asiatic golden cat at 102 stations, marbled cat at 39 stations and Sunda leopard cat at 23 stations (Table S22).

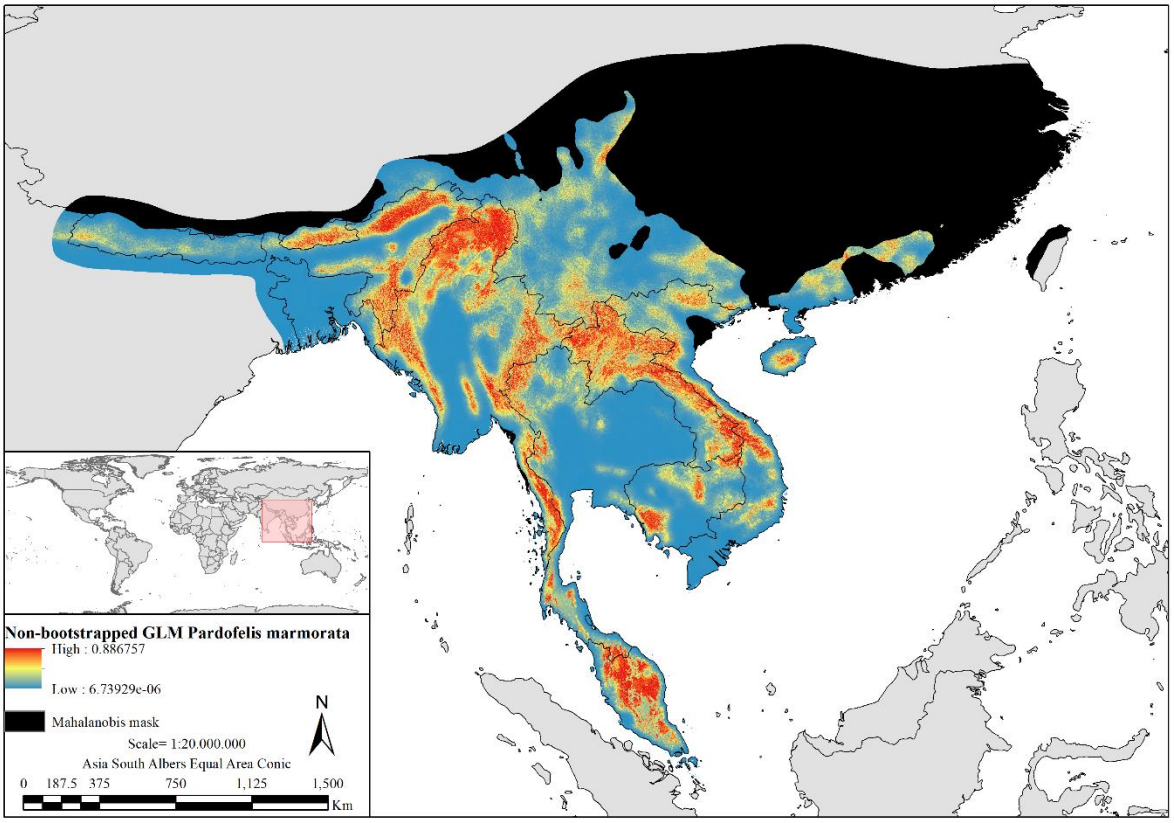
### 3.2.Effects of algorithm and bootstrapping on predictive performances

Predictive performances were substantially different comparing the species distribution probability surfaces produced with the four modelling frameworks (see *Model validation* in Supplementary Information; Tables S23-S27).

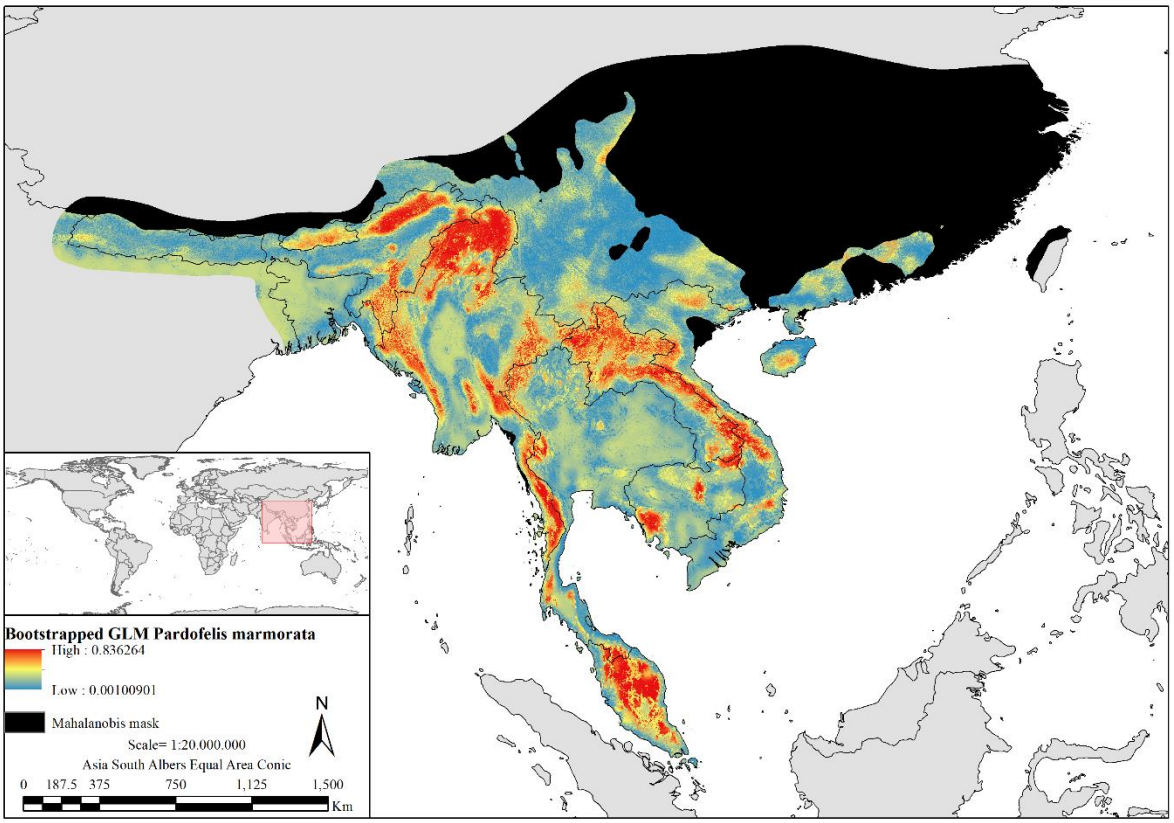
The ANOVA tests performed to assess whether the algorithm, the bootstrapping or their combination had significant effects on the models' predictive performances, revealed that there was a statistically significant difference in the PCC scores based on the bootstrapping ($F_{(1, 44)}$= 9.72, $p$< 0.01) and a less strong difference based on the interaction of algorithm and bootstrapping ($F_{(1, 44)}$= 3.56, $p$< 0.1; Figure S1). We also found sensitivity to be significantly related to algorithm ($F_{(1, 44)}$= 3.72, $p$< 0.1; Figure S2), and specificity to be related to bootstrapping ($F_{(1, 44)}$= 8.53, $p$< 0.01) and, more weakly, to algorithm ($F_{(1, 44)}$= 3.30, $p$< 0.1; Figure S3). We also found TSS to be related to bootstrapping ($F_{(1, 44)}$= 3.30, $p$< 0.1; Figure S4). We did not find any statistically significant difference for AUC (Figure S5).

### 3.3.Effects of algorithm and bootstrapping on models similarity

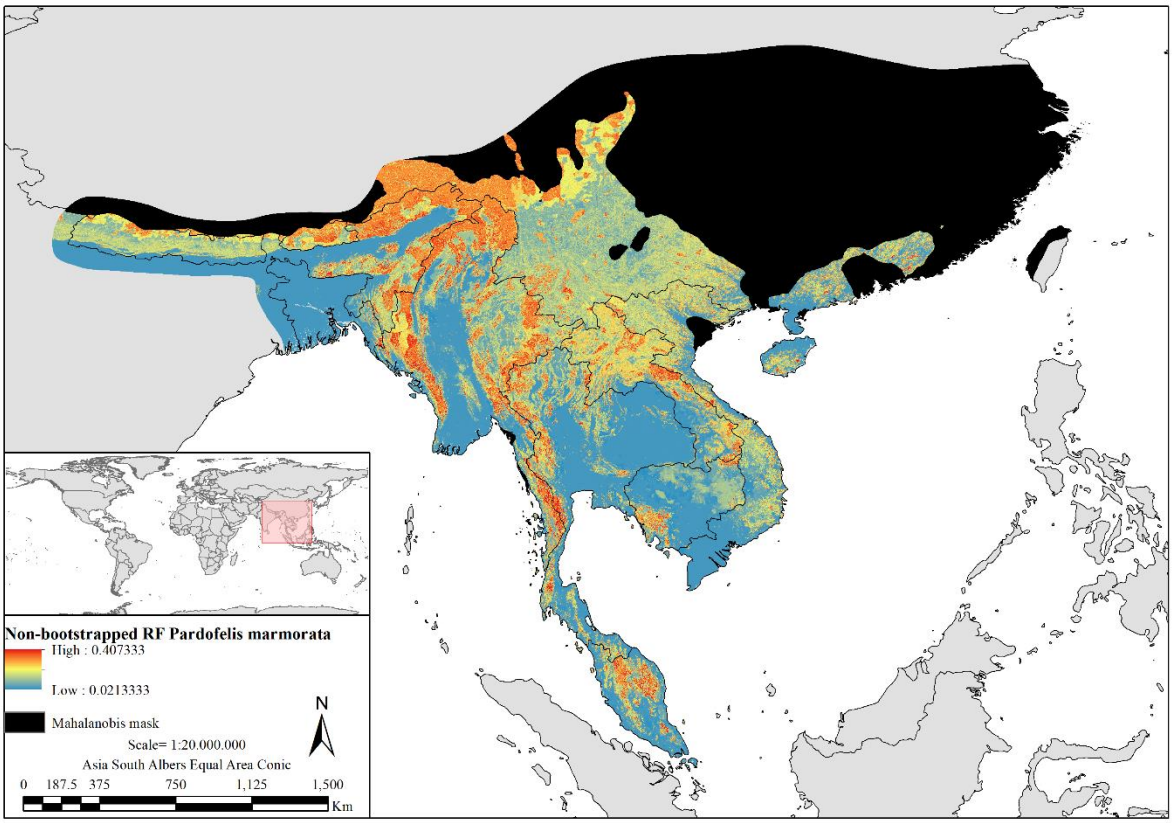
Using the final sets of scale-optimised, uncorrelated covariates, we produced species distribution probability surfaces for the sampled felids by using the four modelling approaches (Figure 1 and Figures S6-S53).



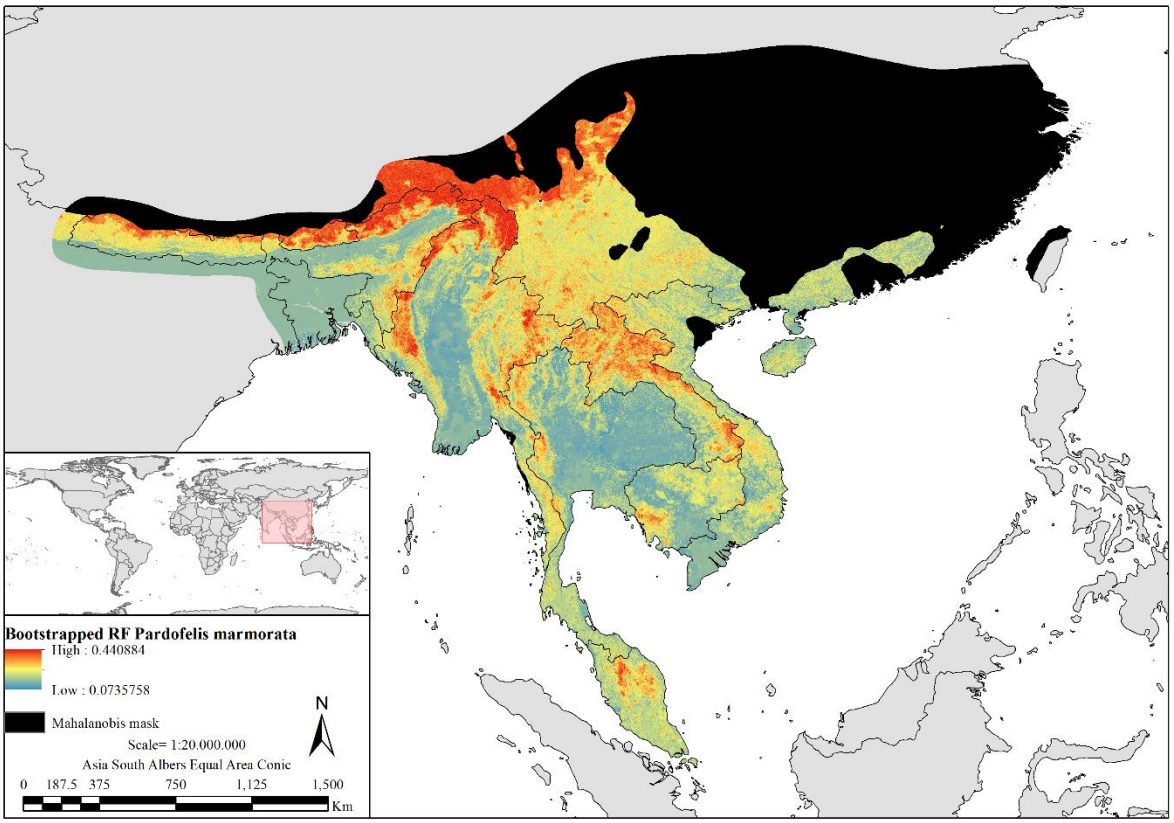

(a)


Bootstrapped GLM Pardofelis marmorata
High : 0.836264
Low : 0.00100901
Mahalanobis mask
Scale= 1:20.000.000
Asia South Albers Equal Area Conic
0 187.5 375 750 1,125 1,500
Km
N

(b)


Non-bootstrapped RF Pardofelis marmorata
High : 0.407333
Low : 0.0213333
Mahalanobis mask
Scale= 1:20.000.000
Asia South Albers Equal Area Conic
0 187.5 375 750 1,125 1,500
Km
N

(c)




(d)

**Figure 1.** Examples of different probability surfaces for marbled cat (*Pardofelis marmorata*) in mainland Southeast Asia, produced with (a) non-bootstrapped GLM, (b) bootstrapped GLM, (c) non-bootstrapped RF and (d) bootstrapped RF.

The four models produced substantially different probability surfaces, showing sometimes very low correlation between models for the same species (see *Model predictions* in Supplementary Information; Tables S4-S15).

The Mantel tests, performed to evaluate whether there was a significant effect of algorithm and bootstrapping on the similarities observed between species distribution probability surfaces, showed a clear pattern in which there is strong effect of algorithm on the correlations in the output (mean Mantel= 0.84; Table 1). However, the Mantel tests showed

also that there was no effect of bootstrapping on the correlations between probability surfaces (mean Mantel= -0.40; Table 1).

**Table 1.** Results of the Mantel tests to evaluate the effects of algorithm and bootstrapping on the correlations between species distribution probability surfaces.

| Study areas | Species | Algorithm Mantel tests | Bootstrapping Mantel tests |
|---|---|---|---|
| Mainland | *Neofelis nebulosa* | 0.98 | -0.49 |
| | *Catopuma temminckii* | 0.24 | -0.16 |
| | *Pardofelis marmorata* | 0.93 | -0.47 |
| | *Prionailurus bengalensis* | 0.94 | -0.46 |
| Borneo | *Neofelis diardi* | 0.99 | -0.46 |
| | *Catopuma badia* | 0.93 | -0.42 |
| | *Pardofelis marmorata* | 0.71 | -0.23 |
| | *Prionailurus javanensis* | 0.96 | -0.48 |
| Sumatra | *Neofelis diardi* | 0.64 | -0.21 |
| | *Catopuma temminckii* | 0.90 | -0.57 |
| | *Pardofelis marmorata* | 0.95 | -0.34 |
| | *Prionailurus javanensis* | 0.96 | -0.48 |
| Mean | | 0.84 | -0.40 |

### 3.4.Factors determining spatial scales selection

The scale optimisation revealed that the most representative scales for each covariate, for each felid, varied substantially based on the algorithm applied, with GLM selecting overall broader scales than RF (see *Scale optimisation and covariates selection* in Supplementary Information; Tables S1-S3). Mantel *r* value is a good indicator of relative effect size (Cushman et al. 2013b) and, based on this criterion, difference in algorithm had the strongest relationship with difference in ordinal scales among covariates across species models. The second highest relationship was for study area, which had approximately 70% of the effect size of difference in algorithm, based on magnitude of Mantel *r*. The third strongest relationship between ordinal scales and predictive factors was with species, which showed approximately 20% of the effect size of difference in algorithm (Table 2).

**Table 2.** Simple Mantel tests results for scales relationships.

| Covariate | Mantel r | *p* |
|---|---|---|
| Algorithm | 0.14 | 0.001 |
| Study area | 0.10 | 0.001 |
| Species | 0.03 | 0.001 |
| Genus | 0.003 | 0.15 |
| Size | -0.0003 | 0.49 |

The partial Mantel tests were carried out to determine whether the methodological differences (i.e., algorithm) were independent of the ecological differences (i.e., species' size, species, genus and study area) in relation to correlation between difference in ordinal scales and difference in predictive variables. The partial Mantel tests showed that all relationships identified in the single Mantel tests were independent of relationships with other factors. Indeed, in all cases the partial Mantel tests slightly increased the Mantel *r* values, showing that removing the co-varying effect of other factors strengthened the correlation between difference in ordinal scales and difference in predictive variables. The correlation between ordinal scales difference and algorithm difference was strongest, followed also in this case by study area, species, genus and species' size (Table 3).

**Table 3.** Partial Mantel tests results for scales relationships.

| Covariate | Partial | Mantel r | *p* |
|---|---|---|---|
| Algorithm | Study area | 0.13590 | 0.001 |
| Algorithm | Species | 0.13521 | 0.001 |
| Algorithm | Genus | 0.13503 | 0.001 |
| Algorithm | Size | 0.13501 | 0.001 |
| Study area | Algorithm | 0.09804 | 0.001 |
| Species | Algorithm | 0.02928 | 0.001 |
| Genus | Algorithm | 0.00329 | 0.11 |
| Size | Algorithm | 0.00001 | 0.49 |

## 4. Discussion

### 4.1.Similarity of prediction and model performance hypotheses

Of the two hypotheses related to model similarities and correlations, our results supported one of them. Specifically, the Mantel tests supported the hypothesis that there are strong differences between spatial predictions produced with GLM and RF, across levels of bootstrapping. Species distribution surfaces resulting from the application of these algorithms are used to prioritise areas for protection or to identify the areas richest in biodiversity (e.g., Kaszta et al. (2020); Penjor et al. (2021)), and differences in the model predictions based on algorithm applied, and not on the underlying ecological relationships, would add a high level of uncertainty to interpreting the results of species distribution models.

Given this large systematic difference between model prediction by GLM and RF, it is important to identify *a priori* which method was superior to avoid the uncertainty and ambiguity of conflicting predictions. A number of past studies have suggested that RF is generally more robust, accurate and reliable than GLM for species distribution modelling (e.g., Cushman et al. (2017); Cushman and Wasserman (2018); Kumar et al. (2021)). Therefore, we hypothesised that RF would consistently out-perform GLM in terms of model performance and accuracy of predictions. However, our hypothesis was not supported. Nominally, GLM outperformed RF based on AUC in all three study areas for the majority of species. The only marginally significant differences due to the algorithm were observed in the scores of the sensitivity and specificity. However, the combined effect of algorithm and bootstrapping was also statistically significant in exampling variance in PCC scores.

We found our results surprising given the rich literature supporting RF as a much stronger modelling algorithm than GLM. Cushman and Wasserman (2018) compared the species distribution models for American marten (*Martes americana*) in northern USA, by using

GLM and RF, concluding that RF produced the most supported model. Similarly, Kumar et al. (2021), in their effort to develop of a new technique to model species-habitat relationships (i.e., the *smoothing function*), also compared GLM and RF, demonstrating that the latter outperformed the former. However, there are cases in which RF has been reported to perform poorly when trained with presence-absence data characterised by several absences (Evans and Cushman 2009, Freeman et al. 2012), likely due to the sensitivity of this algorithm to imbalanced classes. This is indeed the case of our data characterised, especially for some species, by several absences and, even though we corrected for imbalance data by implementing RF class-balance modelling approach (Evans and Cushman 2009), we believe that the effect of zero-inflation on RF has to be evaluated.

Additionally, recent highlighted showed that more complex models, such as RF and other machine learning techniques, showed only a minor improvement in predictive performance compared to simpler models, such as GLM. Ramampiandra et al. (2023), for example, highlighted that while machine learning techniques would only slightly improve predictive performances compared to simpler algorithm, the main impact of the increasing complexity of machine learning techniques was observed on overfitting. Stock et al. (2020) highlighted that, while RF usually performed better than other algorithms they tested in their effort to model fisheries bycatch, the predictive performances of RF were more sensitive when the models were spatially extrapolated beyond the geographic range of the data, confirming that more complex modelling techniques can have better interpolation performances than simpler methods, but are worse at spatial extrapolation (Heikkinen et al. 2012).

Finally, even though we demonstrated that differences in the algorithms led to large and systematic differences in SDM predictions, the Mantel tests did not support our hypothesis that models trained with bootstrapped data were more similar to other bootstrapped models, across algorithms. Therefore, bootstrapping the training data of the models, regardless of the

algorithm, did not lead to significant differences from models trained with non-bootstrapped data. These results are corroborated by the weak support for our hypothesis that bootstrapping would produce SDMs with higher predicting power. Bootstrapping had only a slightly stronger effect on explaining the predictive performances, given that it had strong statistical power in explaining PCC and specificity scores and, to a lesser extent, TSS. This conflicts with our expectation that bootstrapping would improve the model's ability to correctly classify presence-absence data by reducing single models' stochastic errors. However, an interesting behaviour of bootstrapping that we partially observed here, and which has already been demonstrated by, for example, Osawa et al. (2011), is that it improved SDMs' ability to discern true positive rates (i.e., sensitivity) and true negative rates (i.e., specificity). Consequently, TSS, calculated as "sensitivity + specificity – 1", benefitted from bootstrapping, showing significant difference from non-bootstrapped SDMs. Our results generally found that non-bootstrapped models had higher performance based on the majority of our performance measures than bootstrapped models for most species in all three study areas. This suggests that the large effort to conduct bootstrapped sampling to train SDMs may be misplaced and that generally bootstrapping does not improve model predictions.

### 4.2.Scaling hypotheses

The Mantel test analyses supported three of the five hypotheses related to spatial scales. First, contrary to predictions, our results suggested that there is no relationship between similarities in scales of relationship and body size based on species' weight. One would expect for a coherent guild, such as felids, that there would be strong and monotonic association between scales of habitat relationships and body size, with larger-bodied species generally related to habitat factors at broader scales. Large felids are expected to select

habitat factors at broad scales, reflecting their mobility (Elliot et al. 2014) and home-range selection (Khosravi et al. 2019). It is not uncommon, however, to observe species selecting habitat factors at scales that are uncorrelated with their body sizes. Hearn et al. (2018), for example, highlighted that in Borneo, Sunda leopard cat, marbled cat and Sunda clouded leopard selected tree cover at increasing spatial scales, as one would expect based on their sizes, but that they selected elevation at spatial scales that would be unlikely based on their sizes, with Sunda clouded leopard and marbled cat selecting elevation at slightly more than 100m, and the small Sunda leopard cat at almost 2km. These examples suggest that simple ideas about the association between body sizes and scales of species-habitat relationships might be misplaced, and species-specific formal optimal scales optimisation has to be carried out to assess the representative spatial scales (McGarigal et al. 2016).

Second, we found significant relationships between species and scales, such that covariates for the same species across the three study areas had more similarities in scales than covariates for different species. This is expected if there are species-level associations with scales of relationships with habitat factors which are general and consistent (e.g., Cushman and McGarigal (2002)). However, our results suggested that there is no significant association between the scales of covariates selected and taxonomic genera, rejecting our third scaling hypothesis. This suggests that the species-level scales associations with habitat factors do not extend to the next higher level of taxonomic association. However, differences in the spatial scales between members of the same genus have been highlighted, for example, between mainland clouded leopard and Sunda clouded leopard (Macdonald et al. 2018, Macdonald et al. 2019). Of the covariates used to model suitable habitat for the two species, compound topographic index was selected at 500m by mainland clouded leopard and 1km by Sunda clouded leopard, and the difference became stronger looking at other shared covariates, with focal mean of slope position selected at 8km by mainland clouded leopard

and at 250m by Sunda clouded leopard. Similarly, Vergara et al. (2016) found strong differences in the selection of spatial scales between pine marten (*Martes martes*) and beech marten (*Martes foina*) in the northern Iberian Peninsula, with some of the covariates assessed to model habitat suitability showing strong displacement between the two species (e.g., % crops= 16km for pine marten and 4km for beech marten, and % forests= 2km for pine marten and 32km for beech marten).

Our results indicated that there is strong relationship between scales of covariates and study area, suggests that there are regional-scale ecological or geographical factors that affect scales of habitat relationships across multiple species. This may reflect differences in regional limiting factors (e.g., Cushman et al. (2013a); Vergara et al. (2016)) that result in different factors at different spatial scales being the best predictors in different study areas. Multivariate assessment of the suitable habitat for terrestrial vertebrate fauna in Southeast Asia (Macdonald et al. 2020, Chiaverini et al. 2022) revealed, for example, that altitudinal gradients affected mainland biodiversity at much broader scale (32km) than insular species, which are in turn affected by elevation only at a very fine scale (250m). Differences between study areas have been highlighted also for felids: Atzeni et al. (2020) found that snow leopard (*Panthera uncia*) differently selected the response scale for limiting habitat factors across two landscapes in China. This result also directly reflects on the differential selection of habitat features and spatial scales implemented by animal species at different moments of their life (Johnson 1980, Maiorano et al. 2017). Although we cannot know whether the individuals sampled by camera traps were resident animal or dispersals, our results suggest that there is an underlying and general difference of optimal spatial scales which is dependent solely on the study area, regardless of their behaviour.

Finally, we found an extremely strong correlation between difference in scales and difference in algorithm, supporting our last hypothesis. This relationship was almost twice as

strong as between any other factor and difference in scales, suggesting that selection of spatial scales is differently affected by GLM and RF. GLM generally selected broader scales than RF, and the strength of the Mantel correlations suggests that these methods strongly differ in the scales they selected across models for the 12 species (4 species for 3 study areas) in Southeast Asia. Our result that GLM systematically predicts broader spatial scales than RF is consistent with several other studies that have compared scale selection in the two studies (e.g., Cushman et al. (2017); Cushman and Wasserman (2018)).

This finding has major implications for species distribution modelling using scale optimisation. Specifically, if distinct statistical methods differently select spatial scales, then the results of the scale optimisation process are unreliable and not consistently related to underlying ecological relationships. However, given this is an empirical analysis in which the true relationships between species and habitat factors across scales are not known, these results strongly suggest the need to use a controlled simulation study (e.g., Atzeni et al. (2020); Chiaverini et al. (2021)) to systematically evaluate the performance of the GLM and RF in terms of correctly identifying the underlying scales of relationships. Chiaverini et al. (2021), for example, demonstrated that inconsistencies in the scale optimisation process were due to sampling bias of the training data. Therefore, since also in our study areas camera trap samplings were biased towards forested areas, protected areas and easily accessible areas, the discrepancies observed between different algorithms might be partly due to sampling bias, and further simulations are recommended.

## 5. Conclusions

Our study used a vast empirical camera-trap data set for multiple felid species across three large geographical areas of Southeast Asia to test nine hypotheses about the effects of

species distribution modelling algorithm and bootstrapping on model prediction and performance. The most important results of our study were that (1) there was not a clear pattern of RF being superior to GLM in model performance, (2) their predictions were spatially highly divergent and (3) there were large and systematic differences between GLM and RF in the scales of environmental variables they identified. These results are a cause of concern and caution in species distribution modelling. One of the premises of SDMs is that algorithms used provide objective and informative predictions of the factors driving species-habitat relationships, and ability to accurately predict species distribution. If there are large and systematic differences in the scales of factors identified and the predicted probability surfaces produced between algorithms, irrespective of the underlying ecological relationships, then there is likely a much higher degree of uncertainty in SDM predictions than was previously appreciated. This suggests that much more care should be taken by analysts conducting SDM analyses to compare multiple methods to predict their particular data sets and select those that are shown to be most effective in that system and for those species. Altogether, these results are worrying since SDMs are frequently used to design conservation actions. We recommend further investigations of the effects of different factors on the outcomes of SDMs, preferably by applying simulation experiments in which the scales of the species-habitat relationships and the suitable habitats distribution are known.

## 6. References

Ash, E., Z. Kaszta, A. Noochdumrong, T. Redford, P. Chanteap, C. Hallam, B. Jaroensuk, S. Raksat, K. Srinoppawan, and D. W. Macdonald. 2021. Opportunity for Thailand's forgotten tigers: assessment of the Indochinese tiger Panthera tigris corbetti and its prey with camera-trap surveys. Oryx **55**:204-211.

Atzeni, L., S. A. Cushman, D. F. Bai, J. Wang, P. J. Chen, K. Shi, and P. Riordan. 2020. Meta-replication, sampling bias, and multi-scale model selection: A case study on snow leopard (Panthera uncia) in western China. Ecology and Evolution **10**:7686-7712.

Austin, M. P. 1987. Models for the Analysis of Species Response to Environmental Gradients. Vegetatio **69**:35-45.

Bradter, U., W. E. Kunin, J. D. Altringham, T. J. Thom, and T. G. Benton. 2013. Identifying appropriate spatial scales of predictors in species distribution models with the random forest algorithm. Methods in Ecology and Evolution **4**:167-174.

Breiman, L. 2001. Random forests. Machine Learning **45**:5-32.

Chiaverini, L., D. W. Macdonald, H. M. Bothwell, A. J. Hearn, S. M. Cheyne, I. Haidir, L. T. B. Hunter, Ż. Kaszta, E. A. Macdonald, J. Ross, and S. A. Cushman. 2022. Multi-scale, multivariate community models improve designation of biodiversity hotspots in the Sunda Islands. Animal Conservation.

Chiaverini, L., H. Y. Wan, B. Hahn, A. Cilimburg, T. N. Wasserman, and S. A. Cushman. 2021. Effects of non-representative sampling design on multi-scale habitat models: flammulated owls in the Rocky Mountains. Ecological Modelling **450**.

Cushman, S. A., E. Macdonald, E. Landguth, Y. Malhi, and D. Macdonald. 2017. Multiple-scale prediction of forest loss risk across Borneo. Landscape Ecology **32**:1581-1598.

Cushman, S. A., and K. McGarigal. 2002. Hierarchical, multi-scale decomposition of species-environment relationships. Landscape Ecology **17**:637-646.

Cushman, S. A., A. J. Shirk, and E. L. Landguth. 2013a. Landscape genetics and limiting factors. Conservation Genetics **14**:263-274.

Cushman, S. A., and T. N. Wasserman. 2018. Landscape applications of machine learning: comparing random forests and logistic regression in multi-scale optimized predictive

modeling of American Marten Occurrence in Northern Idaho, USA. Pages 185-203 Machine Learning for Ecology and Sustainable Natural Resource Management. Springer.

Cushman, S. A., T. N. Wasserman, E. L. Landguth, and A. J. Shirk. 2013b. Re-Evaluating Causal Modeling with Mantel Tests in Landscape Genetics. Diversity **5**:51-72.

Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, and K. T. Hess. 2007. Random forests for classification in ecology. Ecology **88**:2783-2792.

Efron, B. 1982. The jackknife, the bootstrap, and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia, Pa.

Elith, J., and J. R. Leathwick. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annual Review of Ecology Evolution and Systematics **40**:677-697.

Elliot, N. B., S. A. Cushman, D. W. Macdonald, and A. J. Loveridge. 2014. The devil is in the dispersers: predictions of landscape connectivity change with demography. Journal of Applied Ecology **51**:1169-1178.

ESRI. 2018. ArcGIS Desktop: Release 10.6.1. Environmental System Research Institute. Redlands, CA, USA.

Evans, J. S., and S. A. Cushman. 2009. Gradient modeling of conifer species using random forests. Landscape Ecology **24**:673-683.

Freedman, D. A. 1981. Bootstrapping Regression-Models. Annals of Statistics **9**:1218-1228.

Freeman, E., and G. Moisen. 2007. PresenceAbsence: An R Package for Presence Absence Analysis. Journal of Statistical Software **23**.

Freeman, E. A., G. G. Moisen, and T. S. Frescino. 2012. Evaluating effectiveness of down-sampling for stratified designs and unbalanced prevalence in Random Forest models of tree species distributions in Nevada. Ecological Modelling **233**:1-10.

Goslee, S. C., and D. L. Urban. 2007. The ecodist package for dissimilarity-based analysis of ecological data. Journal of Statistical Software **22**:1-19.

Guisan, A., T. C. Edwards, and T. Hastie. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling **157**:89-100.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters **8**:993-1009.

Guisan, A., W. Thuiller, and N. E. Zimmermann. 2017. Habitat suitability and distribution models : with applications in R. Cambridge University Press, Cambridge.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling **135**:147-186.

Hastie, T., and R. Tibshirani. 1990. Generalized additive models. Chapman and Hall, London.

Hearn, A. J., S. A. Cushman, J. Ross, B. Goossens, L. T. B. Hunter, and D. W. Macdonald. 2018. Spatio-temporal ecology of sympatric felids on Borneo. Evidence for resource partitioning? Plos One **13**.

Hefley, T. J., D. M. Baasch, A. J. Tyre, and E. E. Blankenship. 2014. Correction of location errors for presence-only species distribution models. Methods in Ecology and Evolution **5**:207-214.

Heikkinen, R. K., M. Marmion, and M. Luoto. 2012. Does the interpolation accuracy of species distribution models come at the expense of transferability? Ecography **35**:276-288.

Johnson, D. H. 1980. The Comparison of Usage and Availability Measurements for Evaluating Resource Preference. Ecology **61**:65-71.

Kaszta, Z., S. A. Cushman, and D. W. Macdonald. 2020. Prioritizing habitat core areas and corridors for a large carnivore across its range. Animal Conservation.

Khosravi, R., M. R. Hemami, and S. A. Cushman. 2019. Multi-scale niche modeling of three sympatric felids of conservation importance in central Iran. Landscape Ecology **34**:2451-2467.

Kumar, S. U., P. K. Maini, L. Chiaverini, A. J. Hearn, D. W. Macdonald, Z. Kaszta, and S. A. Cushman. 2021. Smoothing and the environmental manifold. Ecological Informatics **66**.

Levin, S. A. 1992. The Problem of Pattern and Scale in Ecology. Ecology **73**:1943-1967.

Macdonald, D. W., H. M. Bothwell, A. J. Hearn, S. M. Cheyne, I. Haidir, L. T. B. Hunter, A. Kaszta, M. Linkie, E. A. Macdonald, J. Ross, and S. A. Cushman. 2018. Multi-scale habitat selection modeling identifies threats and conservation opportunities for the Sunda clouded leopard (Neofelis diardi). Biological Conservation **227**:92-103.

Macdonald, D. W., H. M. Bothwell, Ż. Kaszta, E. Ash, G. Bolongon, D. Burnham, Ö. E. Can, A. Campos-Arceiz, P. Channa, G. R. Clements, A. J. Hearn, L. Hedges, S. Htun, J. F. Kamler, K. Kawanishi, E. A. Macdonald, S. W. Mohamad, J. Moore, H. Naing, M. Onuma, U. Penjor, A. Rasphone, D. Mark Rayan, J. Ross, P. Singh, C. K. W. Tan, J. Wadey, B. P. Yadav, and S. A. Cushman. 2019. Multi-scale habitat modelling identifies spatial conservation priorities for mainland clouded leopards (Neofelis nebulosa). Diversity and Distributions **0**:1-16.

Macdonald, D. W., L. Chiaverini, H. M. Bothwell, Z. Kaszta, E. Ash, G. Bolongon, O. E. Can, A. Campos-Arceiz, P. Channa, G. R. Clements, A. J. Hearn, L. Hedges, S. Htun, J. F. Kamler, E. A. Macdonald, J. Moore, H. Naing, M. Onuma, A. Rasphone, D. M. Rayan, J. Ross, P. Singh, C. K. W. Tan, J. Wadey, B. P. Yadav, and S. A. Cushman.

2020. Predicting biodiversity richness in rapidly changing landscapes: climate, low human pressure or protection as salvation? Biodiversity and Conservation.

Maiorano, L., L. Boitani, L. Chiaverini, and P. Ciucci. 2017. Uncertainties in the identification of potential dispersal corridors: The importance of behaviour, sex, and algorithm. Basic and Applied Ecology **21**:66-75.

McGarigal, K., S. A. Cushman, and E. Ene. 2012. FRAGSTATS v4: spatial pattern analysis program for categorical and continuous maps. Computer software program produced by the authors at the University of Massachusetts, Amherst, MA. Available: http://www.umass.edu/landeco/research/fragstats/fragstats.html.

McGarigal, K., H. Y. Wan, K. A. Zeller, B. C. Timm, and S. A. Cushman. 2016. Multi-scale habitat selection modeling: a review and outlook. Landscape Ecology **31**:1161-1175.

Osawa, T., H. Mitsuhashi, Y. Uematsu, and A. Ushimaru. 2011. Bagging GLM: Improved generalized linear model for the analysis of zero-inflated data. Ecological Informatics **6**:270-275.

Penjor, U., Z. Kaszta, D. W. Macdonald, and S. A. Cushman. 2021. Prioritizing areas for conservation outside the existing protected area network in Bhutan: the use of multi-species, multi-scale habitat suitability models. Landscape Ecology **36**:1281-1309.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramampiandra, E. C., A. Scheidegger, J. Wydler, and N. Schuwirth. 2023. A comparison of machine learning and statistical species distribution models: Quantifying overfitting supports model interpretation. Ecological Modelling **481**.

Roberts, D. R., V. Bahn, S. Ciuti, M. S. Boyce, J. Elith, G. Guillera-Arroita, S. Hauenstein, J. J. Lahoz-Monfort, B. Schroder, W. Thuiller, D. I. Warton, B. A. Wintle, F. Hartig,

and C. F. Dormann. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography **40**:913-929.

Stock, B. C., E. J. Ward, T. Eguchi, J. Jannot, J. T. Thorson, B. E. Feist, and B. X. Semmens. 2020. Comparing predictions of fisheries bycatch using multiple spatiotemporal species distribution model frameworks. Canadian Journal of Fisheries and Aquatic Sciences **77**:146-163.

Stupariu, M. S., S. A. Cushman, A. I. Plesoianu, I. Patru-Stupariu, and C. Furst. 2021. Machine learning in landscape ecological analysis: a review of recent approaches. Landscape Ecology.

Valavi, R., G. Guillera-Arroita, J. J. Lahoz-Monfort, and J. Elith. 2022. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. Ecological Monographs **92**.

Vaughan, I. P., and S. J. Ormerod. 2005. The continuing challenges of testing species distribution models. Journal of Applied Ecology **42**:720-730.

Vergara, M., S. A. Cushman, F. Urra, and A. Ruiz-Gonzalez. 2016. Shaken but not stirred: multiscale habitat suitability modeling of sympatric marten species (Martes martes and Martes foina) in the northern Iberian Peninsula. Landscape Ecology **31**:1241-1260.

Wan, H. Y., S. A. Cushman, and J. L. Ganey. 2019. Improving habitat and connectivity model predictions with multi-scale resource selection functions from two geographic areas. Landscape Ecology **34**:503-519.

Wiens, J. A. 1989. Spatial Scaling in Ecology. Functional Ecology **3**:385-397.

Xu, Y., and R. Goodacre. 2018. On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the

Generalization Performance of Supervised Learning. Journal of Analysis and Testing **2**:249-262.

# Supplementary material for Chapter 4

## Materials and methods

### Mahalanobis mask

Since the camera trap sampling failed to survey the full gradients of all the habitat factors used to model felids distribution over the entire study area (e.g., camera traps occurred mainly within protected areas and along easily accessible areas such as disused logging roads), we masked out the regions that were ecologically highly divergent from the sampled ones, to avoid projecting unreliable species distribution probability surfaces over unsampled locations. To produce the masks, we first produced Mahalanobis distance surfaces independently for mainland Southeast Asia, Borneo and Sumatra in R v3.5.1 using the package *dismo* (Hijmans et al. 2017) displaying, in a continuous layer, the distance values between points in a multivariate space, that we defined by using the habitat features that most likely determined the geographic bias of the camera trap survey: annual maximum temperature, annual minimum temperature, annual precipitations, elevation, human footprint, protected areas, roughness and tree canopy cover (Chiaverini et al. 2022).

Then, we converted the continuous Mahalanobis distance surfaces to binary form by using the $40^{th}$ and the $20^{th}$ percentiles of the Mahalanobis distance surfaces' pixel values distribution for the mainland study area and for the Sunda Islands, respectively, and excluded all pixels with distance values higher than the thresholds (Macdonald et al. 2020, Chiaverini et al. 2022). We chose two different thresholds to reflect the different extents of the study areas, and the fact that the regions characterised by habitat conditions diverse from the sampled ones were likely more extended in the vaster mainland Southeast Asia than in Borneo and Sumatra.

## Results

### Model validation

In the mainland study area, models trained with GLM outperformed RF models based on AUC scores (Tables S23 and S26). Predictive surfaces produced with non-bootstrapped GLM showed the highest AUC for all the species except clouded leopard, which showed the highest AUC for the bootstrapped GLM. However, based on the TSS scores, only Asiatic golden cat and leopard cat confirmed the results of the AUC, while for clouded leopard the model with the highest TSS was the non-bootstrapped RF, and for marbled cat the bootstrapped GLM (Tables S23 and S27).

In Borneo, generally GLM models outperformed RF models based on AUC scores (Tables S24 and S26). Sunda clouded leopard and Sunda leopard cat's probability distributions showed the highest AUC when modelled with bootstrapped GLM, while Borneo bay cat showed the highest AUC with non-bootstrapped GLM. Bootstrapped RF, however, was the model with the highest AUC for marbled cat, even though both non-bootstrapped GLM and bootstrapped GLM had similar AUC. The most supported models, however, were slightly different when assessing the TSS scores (Tables S24 and S27). Only Borneo bay cat confirmed the result of the AUC, while both Sunda clouded leopard and Sunda leopard cat showed higher TSS when modelled with non-bootstrapped GLM. Marbled cat showed the highest TSS with non-bootstrapped RF.

Overall, in Sumatra, RF models were most frequently the most supported ones based on AUC scores (Tables S25 and S26). Sunda clouded leopard and Asiatic golden cat showed the highest AUC when their predictive surfaces were modelled with non-bootstrapped RF. Marbled cat's most supported model was produced with bootstrapped RF, while Sunda leopard cat's model showed the highest AUC when modelled with non-bootstrapped GLM.

The results were confirmed for all the species when assessing TSS instead of AUC, except for Sunda clouded leopard, whose model showing the highest TSS was the one produced with non-bootstrapped GLM (Tables S25 and S27).

**Model predictions**

In the mainland study area, non-bootstrapped GLM – bootstrapped GLM were the most similar models for clouded leopard, marbled cat and leopard cat, with correlations respectively of 0.97, 0.89 and 0.93, while for Asiatic golden cat non-bootstrapped GLM – bootstrapped RF were the most similar models with correlation of 0.58 (Tables S4-S7). For clouded leopard and Asiatic golden cat, the most different models were bootstrapped GLM – non-bootstrapped RF, with correlations respectively of 0.10 and -0.23, while for marbled cat and leopard cat the most different models were bootstrapped GLM – bootstrapped RF (correlation= 0.31) and non-bootstrapped GLM – non-bootstrapped RF (correlation= 0.30).

In Borneo, non-bootstrapped RF – bootstrapped RF were the most similar models for Sunda clouded leopard, Borneo bay cat and marbled cat, with correlations respectively of 0.90, 0.88 and 0.87, while for Sunda leopard cat the most similar models were non-bootstrapped GLM – bootstrapped GLM, with correlation of 0.97 (Tables S8-S11). The most different models for Sunda clouded leopard were bootstrapped GLM – non-bootstrapped RF, with correlation of 0.61, while the most different models for Borneo bay cat, marbled cat and Sunda leopard cat were non-bootstrapped GLM – non-bootstrapped RF, with correlations respectively of 0.09, 0.46 and 0.60.

In Sumatra, non-bootstrapped GLM – bootstrapped GLM were the most similar models for Sunda clouded leopard, Asiatic golden cat and marbled cat, with correlations of 0.87, 0.90 and 0.87, respectively. The most similar models for Sunda leopard cat were non-

bootstrapped RF – bootstrapped RF, with correlation of 0.89 (Tables S12-S15). The most different models for Sunda clouded leopard and Asiatic golden cat were non-bootstrapped GLM – non-bootstrapped RF, with correlations respectively of 0.53 and 0.61. Bootstrapped GLM – non-bootstrapped RF and bootstrapped GLM – bootstrapped RF were the most different models for marbled cat, both with correlation of 0.64, while the most different models for Sunda leopard cat were bootstrapped GLM – non-bootstrapped RF (correlation= -0.02).

Evaluating together the correlations between all the species distribution predictive surfaces, it becomes clear that the most similar ones are those trained with the same algorithm. Overall, the most similar models are non-bootstrapped GLM – bootstrapped GLM, with a mean correlation of 0.81, followed by non-bootstrapped RF – bootstrapped RF, with a mean correlation of 0.77. Non-bootstrapped GLM – non-bootstrapped RF show a mean correlation of 0.38, while bootstrapped GLM – bootstrapped RF have mean correlation of 0.50. Non-bootstrapped GLM – bootstrapped RF show mean correlation of 0.49, while bootstrapped GLM – non-bootstrapped RF have mean correlation of 0.37.

**Scale optimisation and covariates selection**

In the mainland study area, GLM selected most of the covariates at broad scales for all the felids, with the broadest assessed scales (i.e., 16,000 m and 32,000 m) being selected for more than 60% of the covariates, for all felids. GLM selected the finest scales (i.e., 250 m and 500 m) for 22.2% of the covariates for clouded leopard and Asiatic golden cat, for 27.8% of the covariates for marbled cat and for none of the covariates for leopard cat.

Scale optimisation for mainland using RF algorithm identified very different patterns of scale dependence than did the GLM analysis. For all modelled species, RF selected most

of the covariates at fine scales, with the finest scales being selected for 66.7%, 55.6%, 50.0% and 44.4% of the covariates by marbled cat, clouded leopard, leopard cat and Asiatic golden cat, respectively. Broad scales were selected for 11.1% of the covariates by Asiatic golden cat and marbled cat, for 22.2% of the covariates by clouded leopard and for 27.8% of the covariates by leopard cat.

Also for Borneo, scale optimisation revealed different patterns between the two algorithms, with GLM showing an overall trend toward the selection of variables at broader scales. For all the species, GLM highlighted the broadest scales (i.e., 8,000 m and 10,000 m) to be selected more often than the finest scales (i.e., 250 m and 500 m), with Sunda leopard cat, marbled cat, Sunda clouded leopard and Borneo bay cat selecting the broadest scales for 55.0%, 52.6%, 40.0% and 21.1% of the covariates, respectively, and the finest scales for 15.0%, 15.8%, 20.0% and 15.8% of the covariates, respectively.

Scale optimisation for Bornean species using RF revealed a higher frequency of covariates selected at fine scales. The finest scales were selected for 52.6%, 50.0%, 26.3% and 25.0% of the covariates respectively by marbled cat, Sunda clouded leopard, Borneo bay cat and Sunda leopard cat, while the same species selected the broadest scales for 10.5%, 15.0%, 5.3% and 25.0% of the covariates, respectively. Sunda leopard cat was the only species in Borneo selecting the finest and the broadest scales with the same frequency.

The trend was confirmed also by the scale optimisation for Sumatra, where GLM selected a higher proportion of covariates at broad scales. The broadest scales (i.e., 8,000 m and 10,000 m) were selected for 52.4%, 45.0%, 33.3% and 23.8% of the covariates respectively by Asiatic golden cat, Sunda leopard cat, marbled cat and Sunda clouded leopard. The finest scales (i.e., 250 m and 500 m) were selected for 23.8% of the covariates by Sunda clouded leopard, showing the same proportion of covariate being representative at

the finest and the broadest scales, for 20.0% of the covariates by Sunda leopard cat, and by 9.5% of the covariates by both Asiatic golden cat and marbled cat.

Scale optimisation for Sumatra with RF selected the majority of the covariates at fine scales for all the species. The finest scales were selected for 52.4%, 38.1%, 30.0% and 28.6% of the covariates for Sunda clouded leopard, Asiatic golden cat, Sunda leopard cat and marbled cat, respectively. The broadest scales were selected much less frequently: for 14.3% of the covariates for Asiatic golden cat, for 10.0% of the covariates for Sunda leopard cat and for 9.5% of the covariates for both Sunda clouded leopard and marbled cat.

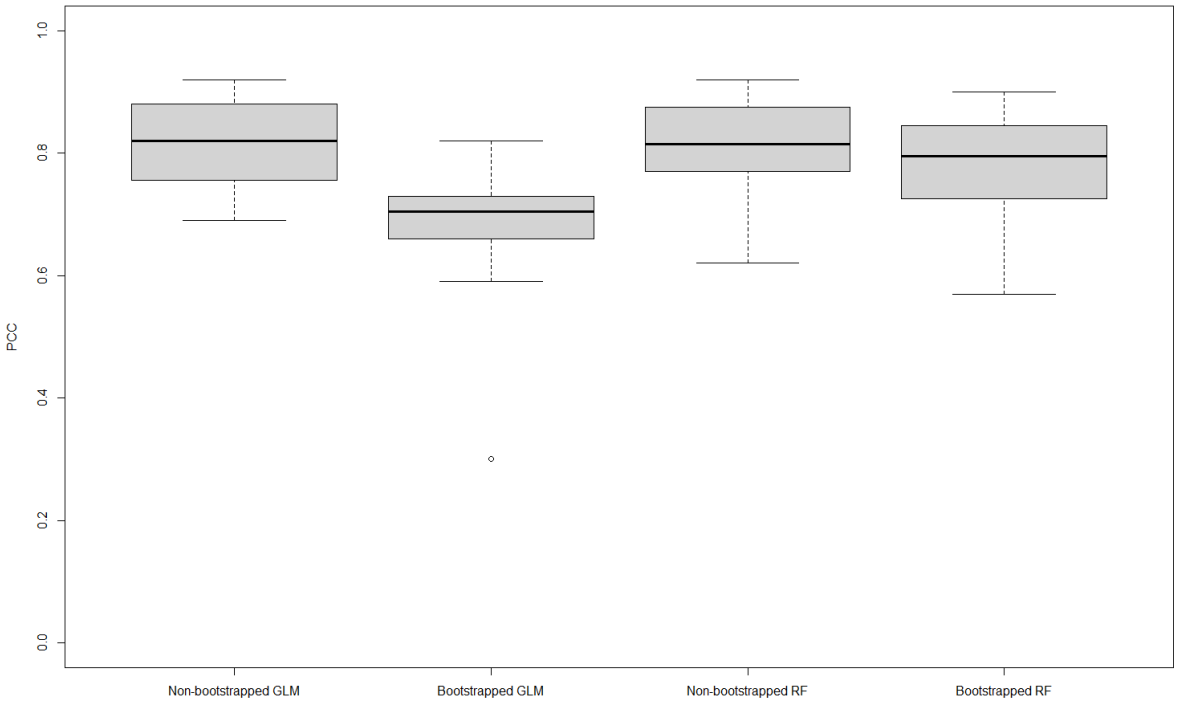



**Figure S1.** Boxplots of the PCC scores.

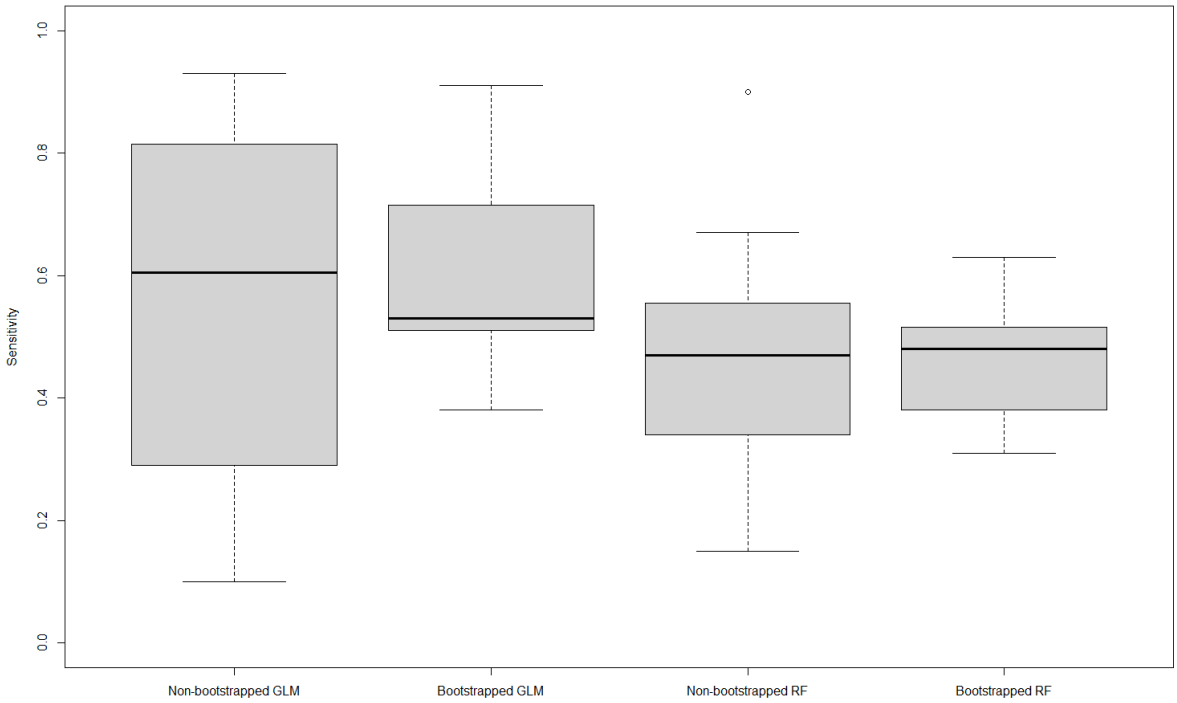



**Figure S2.** Boxplots of the sensitivity scores.

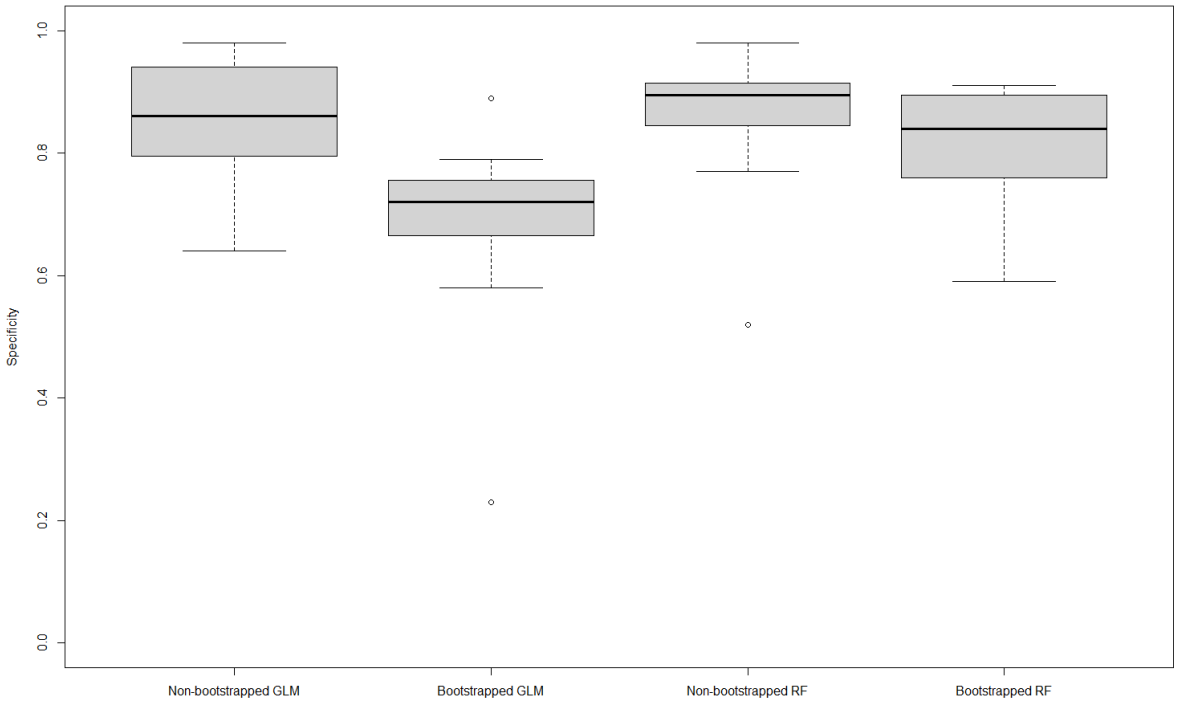



**Figure S3.** Boxplots of the specificity scores.

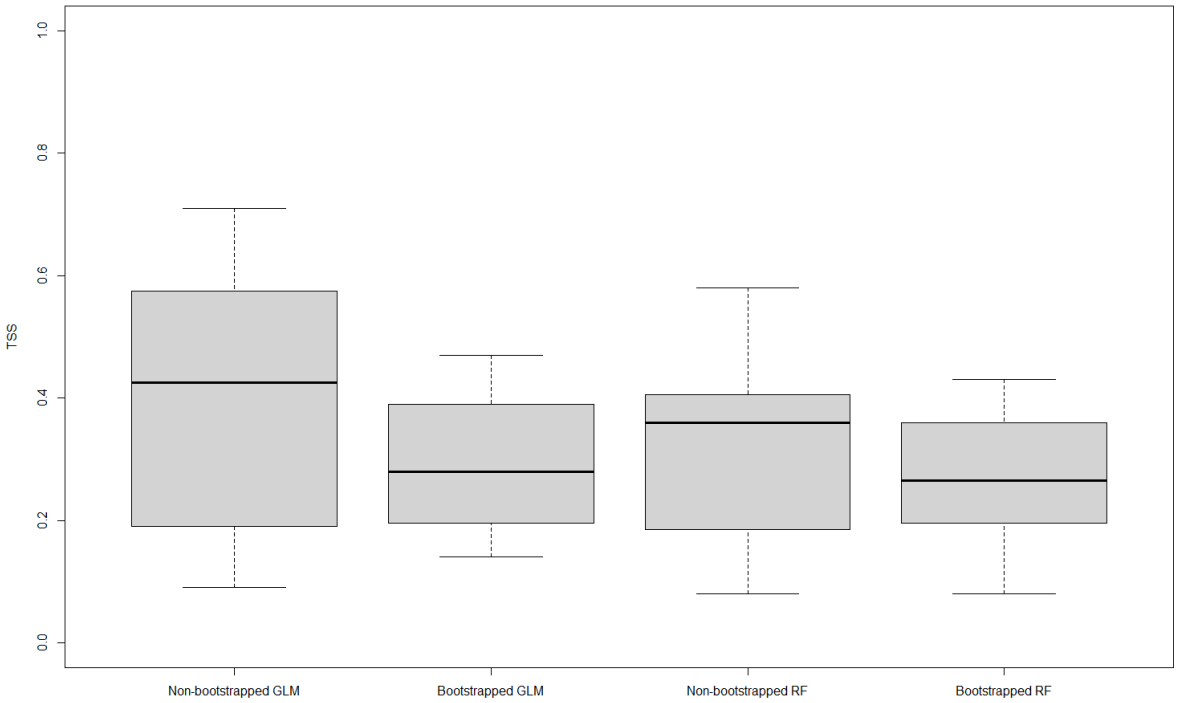



**Figure S4.** Boxplots of the TSS scores.

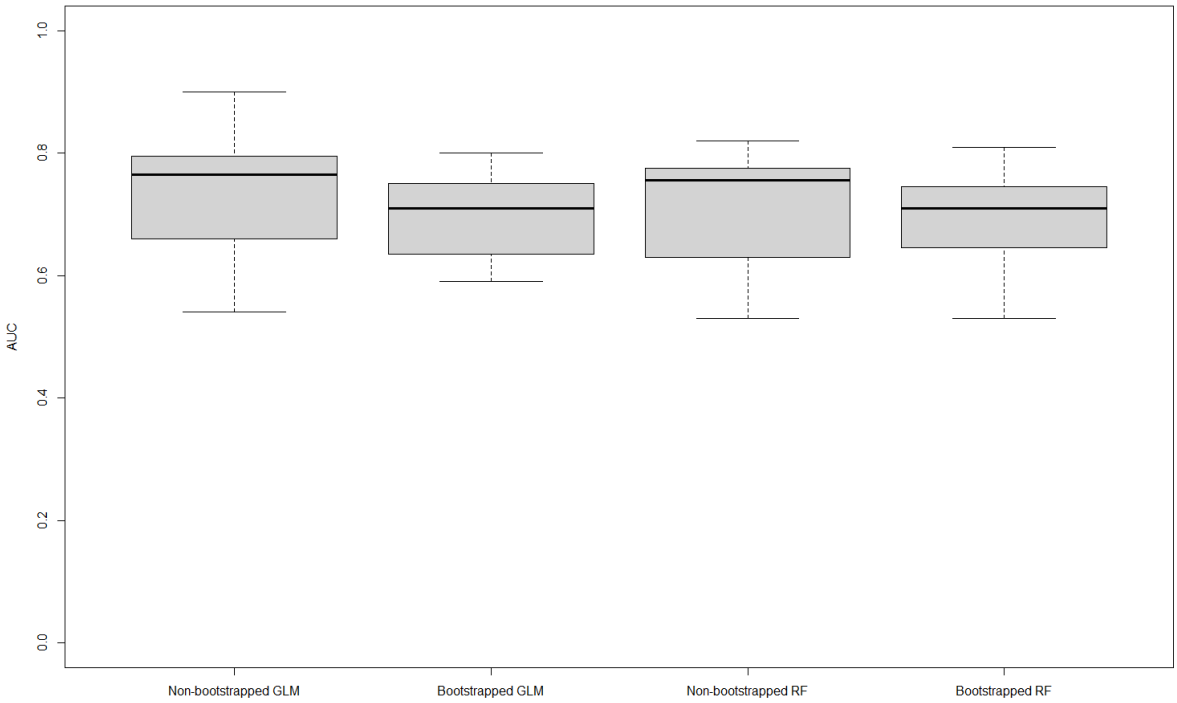



**Figure S5.** Boxplots of the AUC scores.

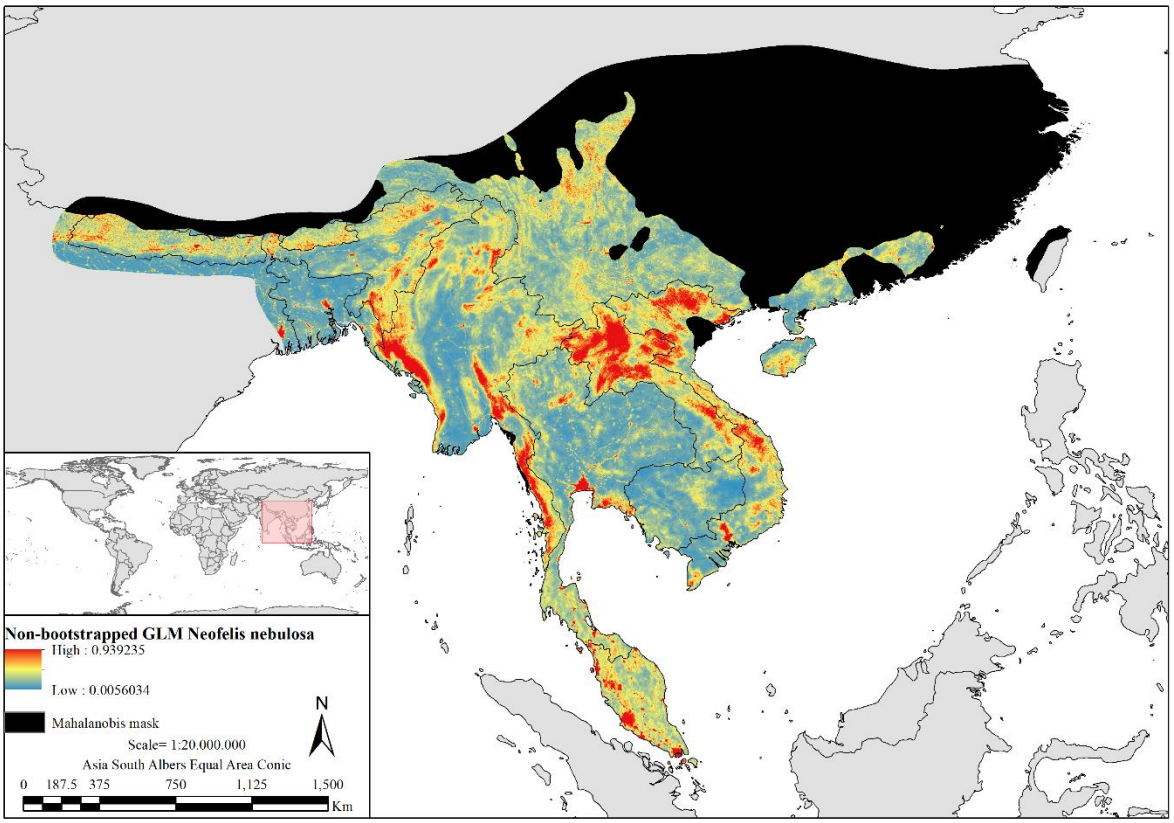



**Figure S6.** Probability surface for clouded leopard (*Neofelis nebulosa*) in mainland Southeast Asia, produced with non-bootstrapped GLM.

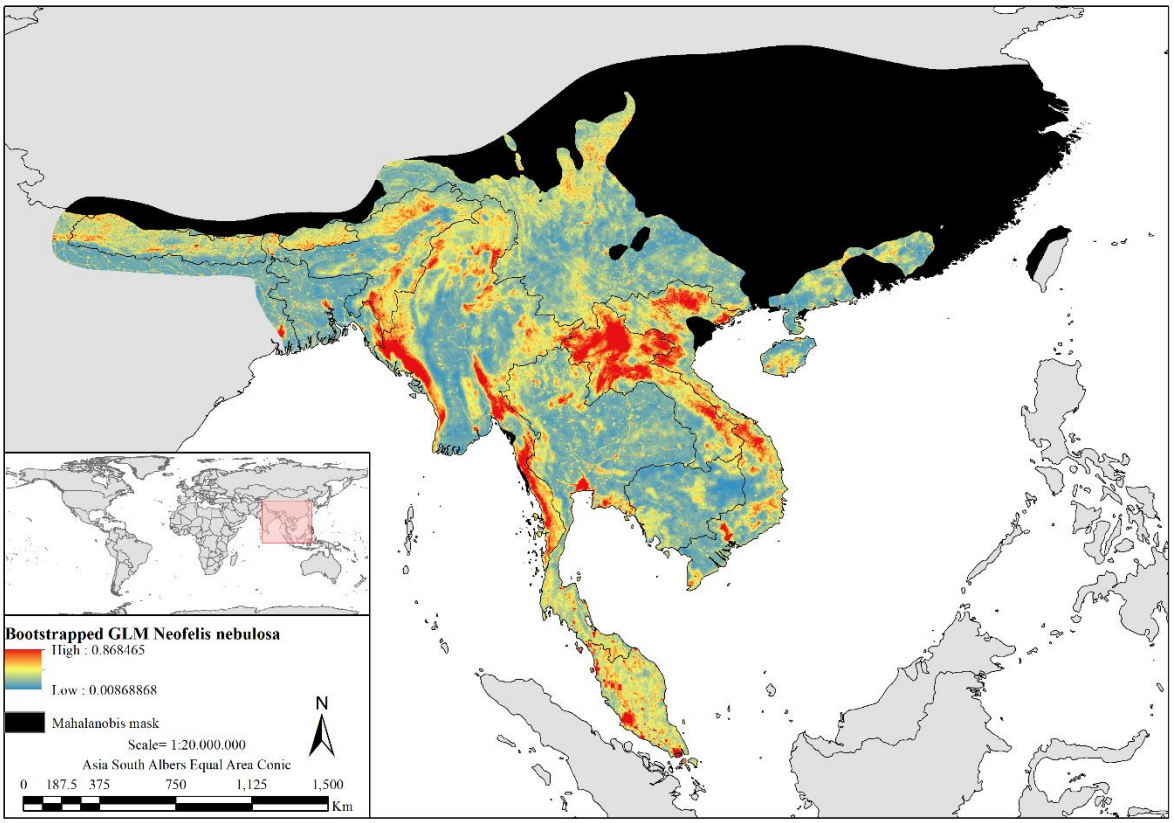



**Figure S7.** Probability surface for clouded leopard (*Neofelis nebulosa*) in mainland Southeast Asia, produced with bootstrapped GLM.

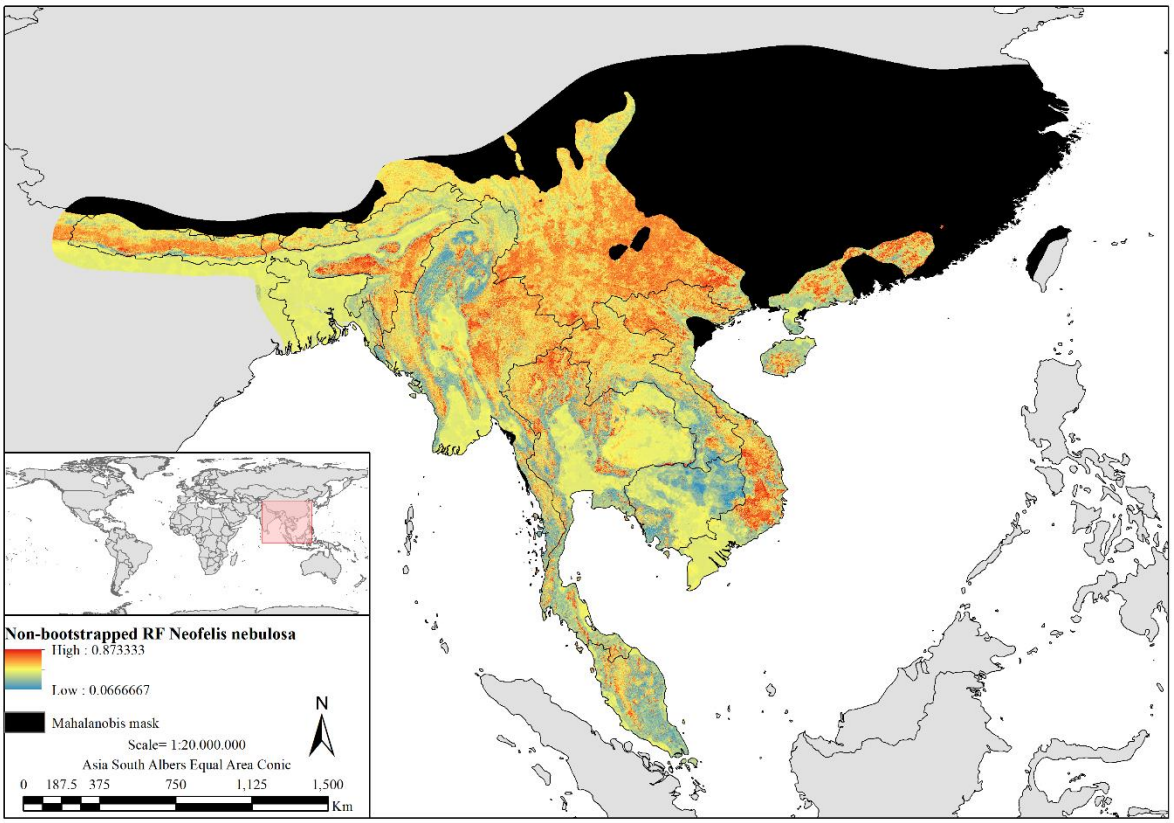



**Figure S8.** Probability surface for clouded leopard (*Neofelis nebulosa*) in mainland Southeast Asia, produced with non-bootstrapped RF.

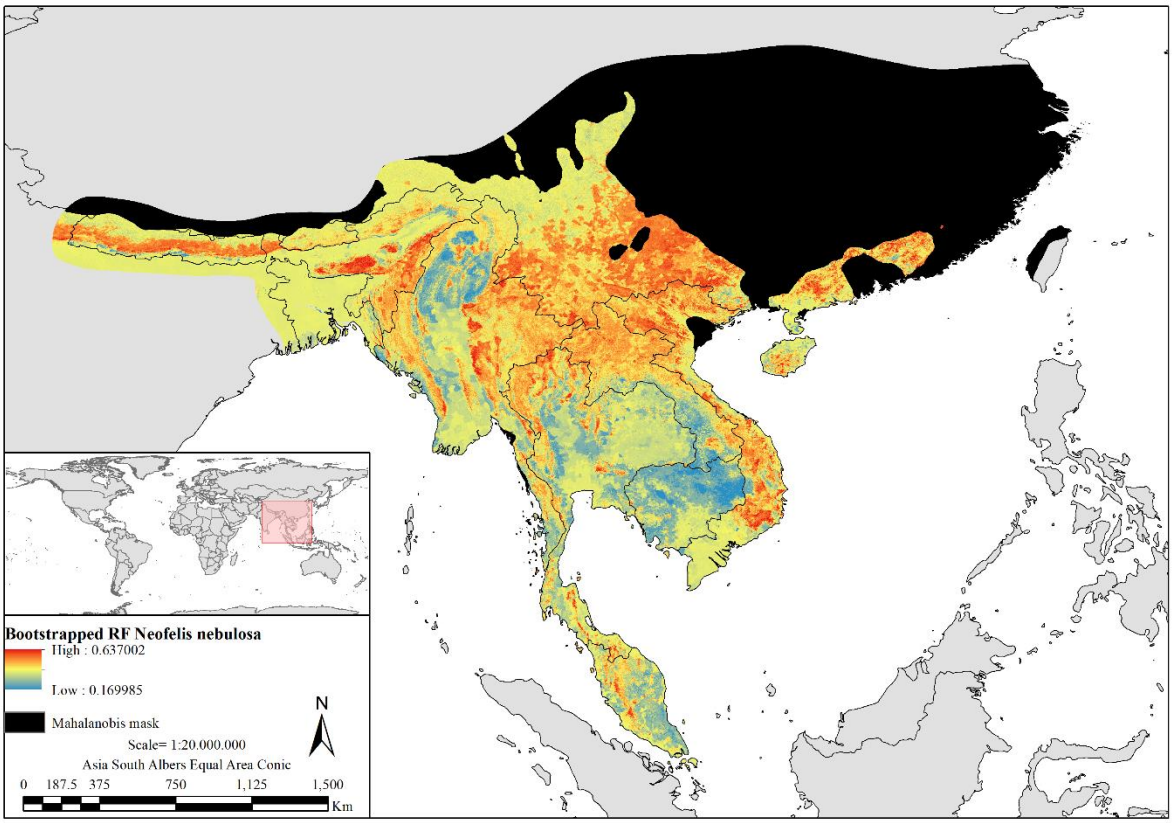



**Figure S9.** Probability surface for clouded leopard (*Neofelis nebulosa*) in mainland Southeast Asia, produced with bootstrapped RF.

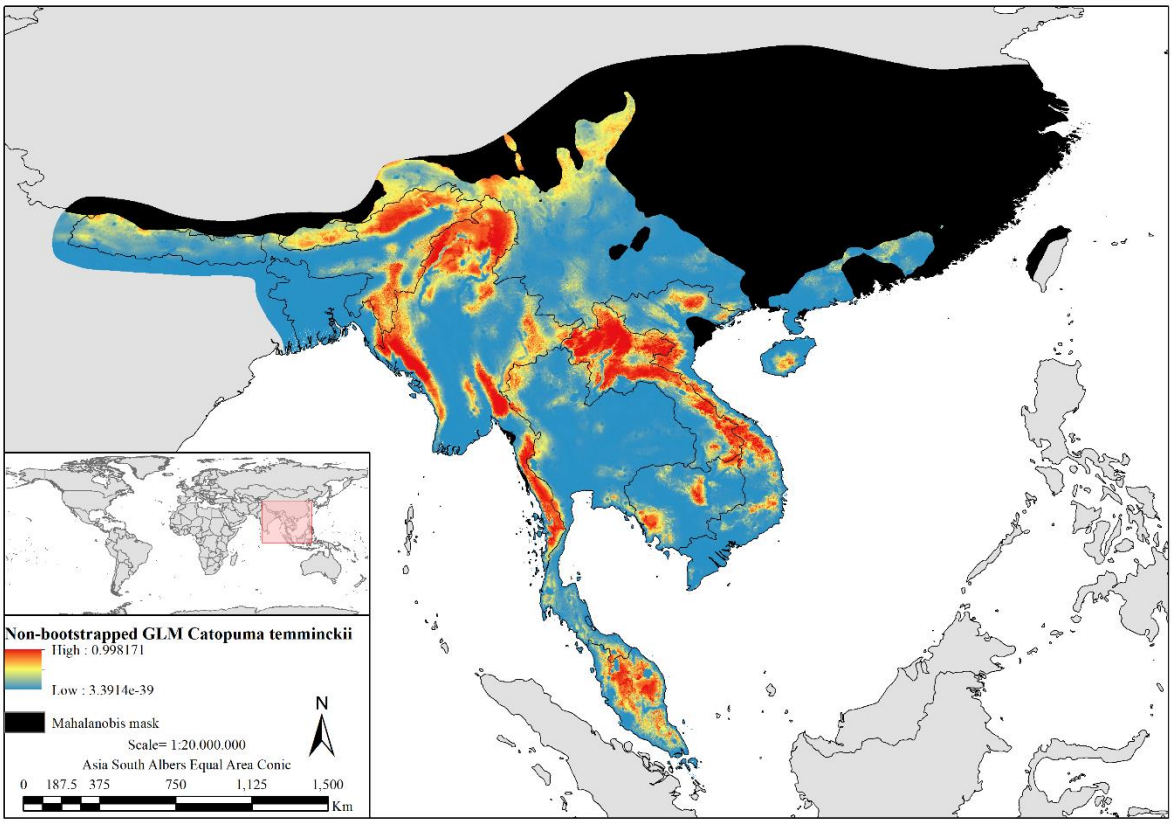



**Figure S10.** Probability surface for Asiatic golden cat (*Catopuma temminckii*) in mainland Southeast Asia, produced with non-bootstrapped GLM.

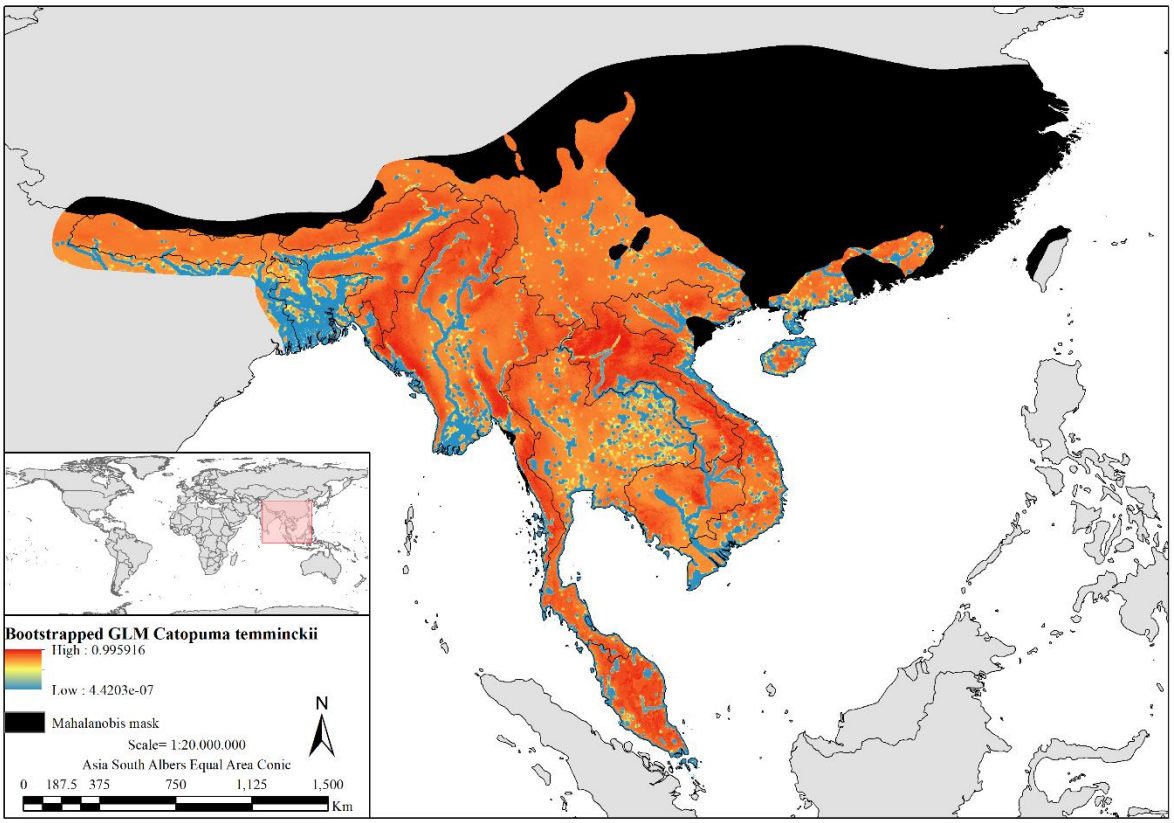



**Figure S11.** Probability surface for Asiatic golden cat (*Catopuma temminckii*) in mainland Southeast Asia, produced with bootstrapped GLM.

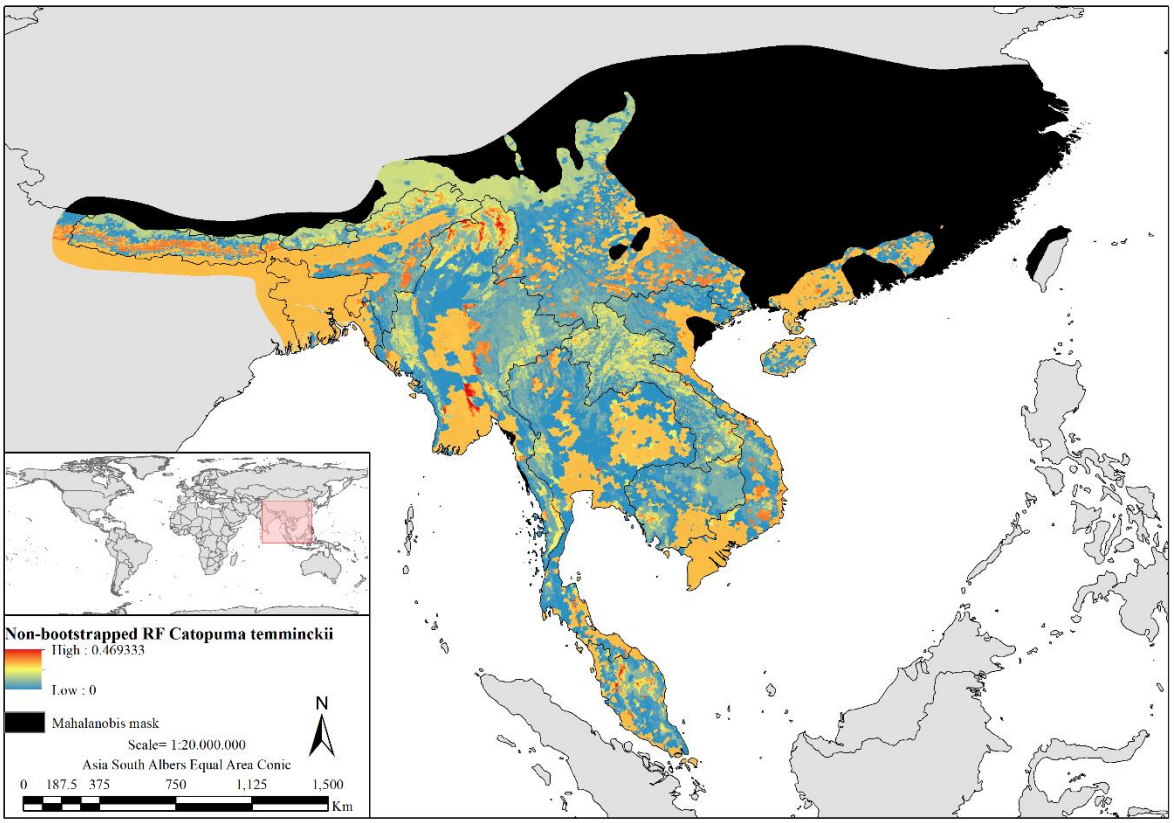



**Figure S12.** Probability surface for Asiatic golden cat (*Catopuma temminckii*) in mainland Southeast Asia, produced with non-bootstrapped RF.

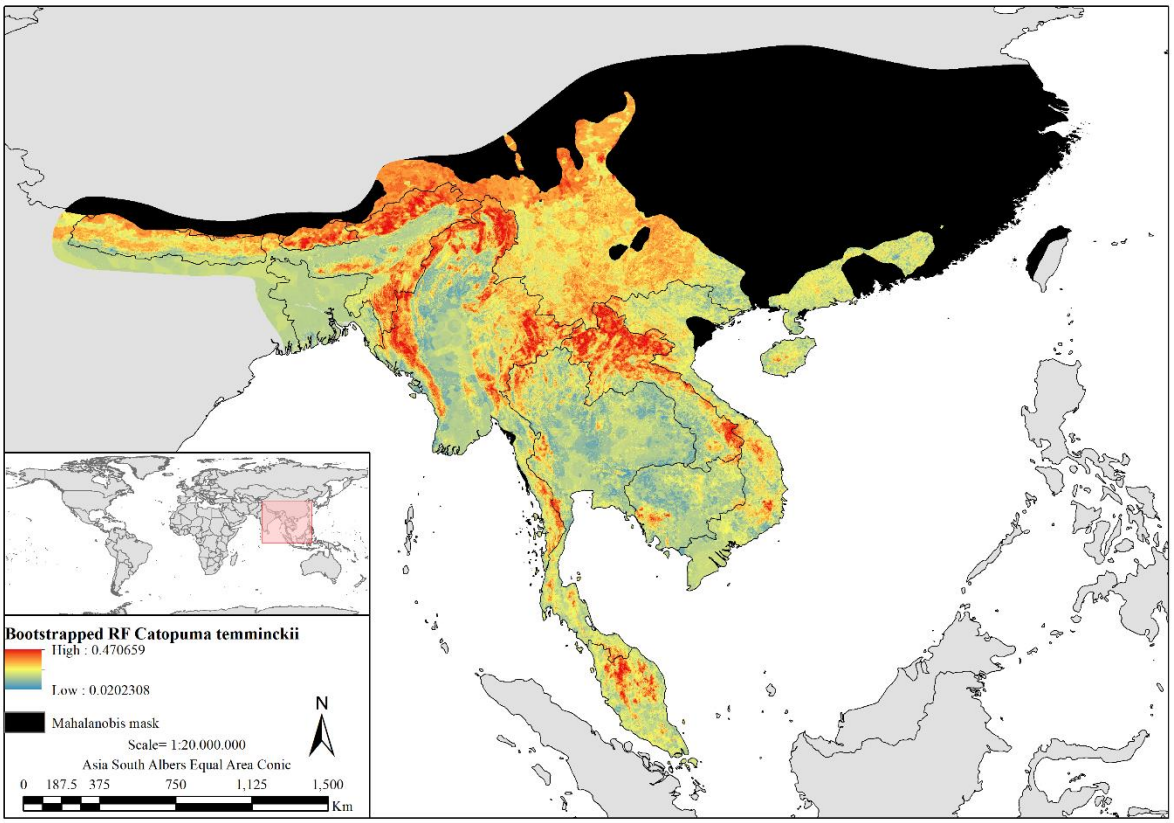



**Figure S13.** Probability surface for Asiatic golden cat (*Catopuma temminckii*) in mainland Southeast Asia, produced with bootstrapped RF.

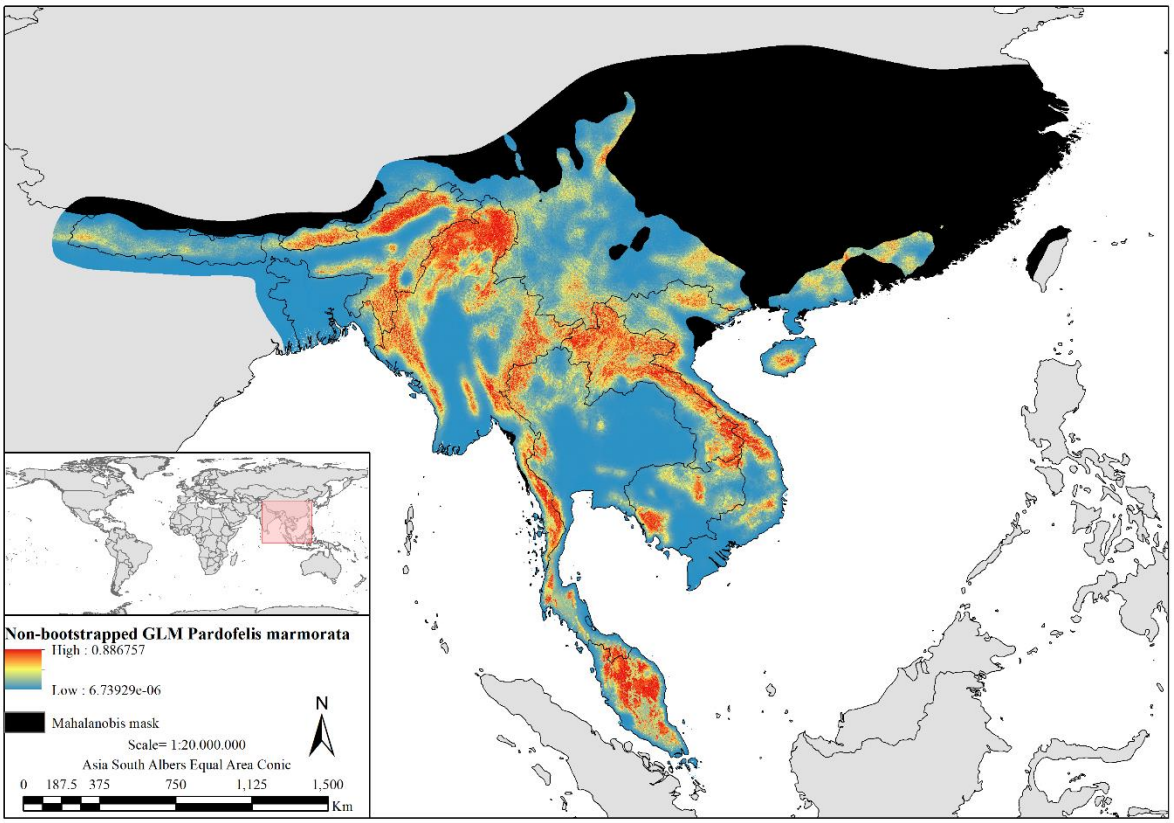



**Figure S14.** Probability surface for marbled cat (*Pardofelis marmorata*) in mainland Southeast Asia, produced with non-bootstrapped GLM.

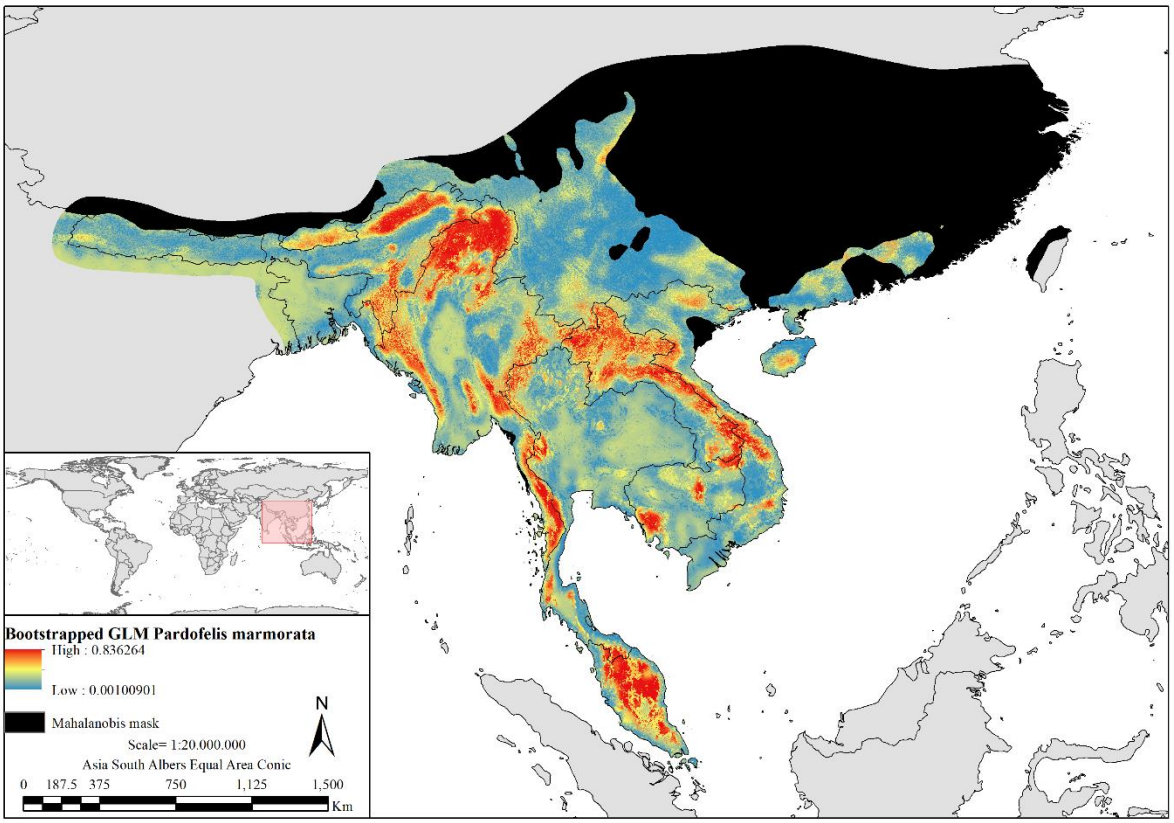



**Figure S15.** Probability surface for marbled cat (*Pardofelis marmorata*) in mainland Southeast Asia, produced with bootstrapped GLM.

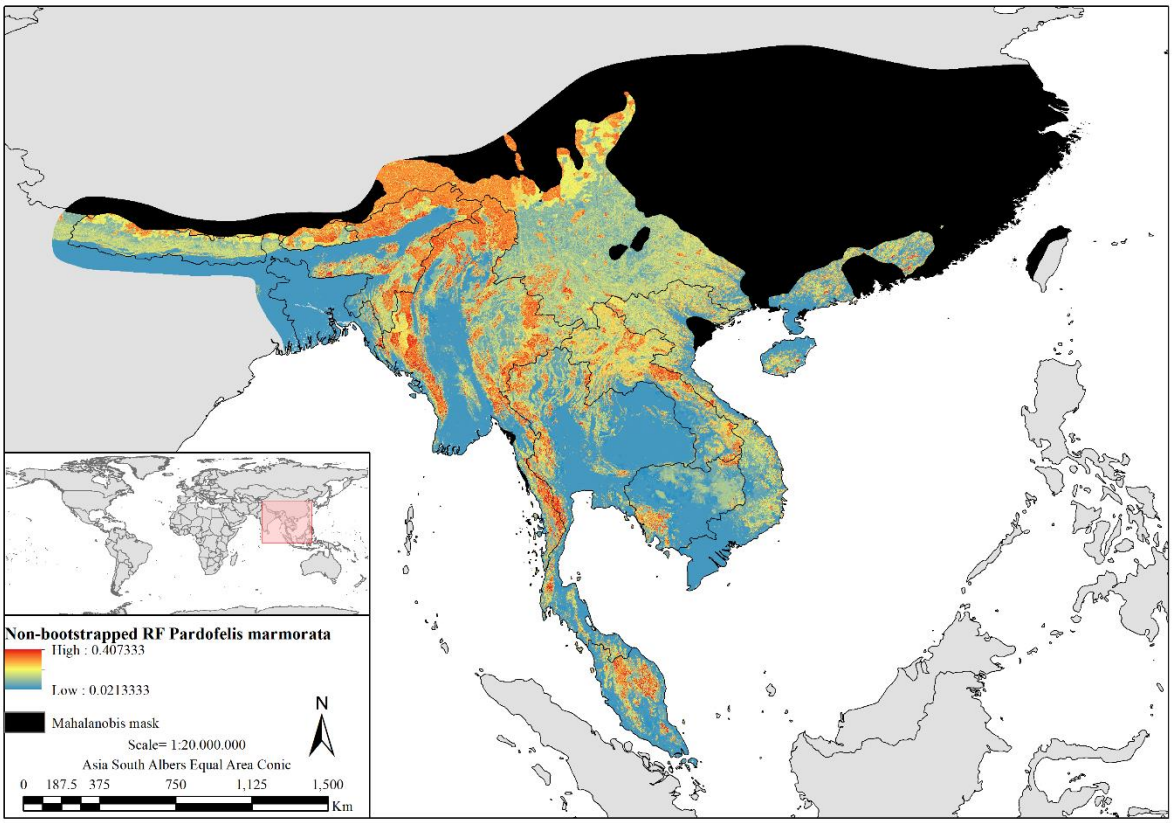



**Figure S16.** Probability surface for marbled cat (*Pardofelis marmorata*) in mainland Southeast Asia, produced with non-bootstrapped RF.

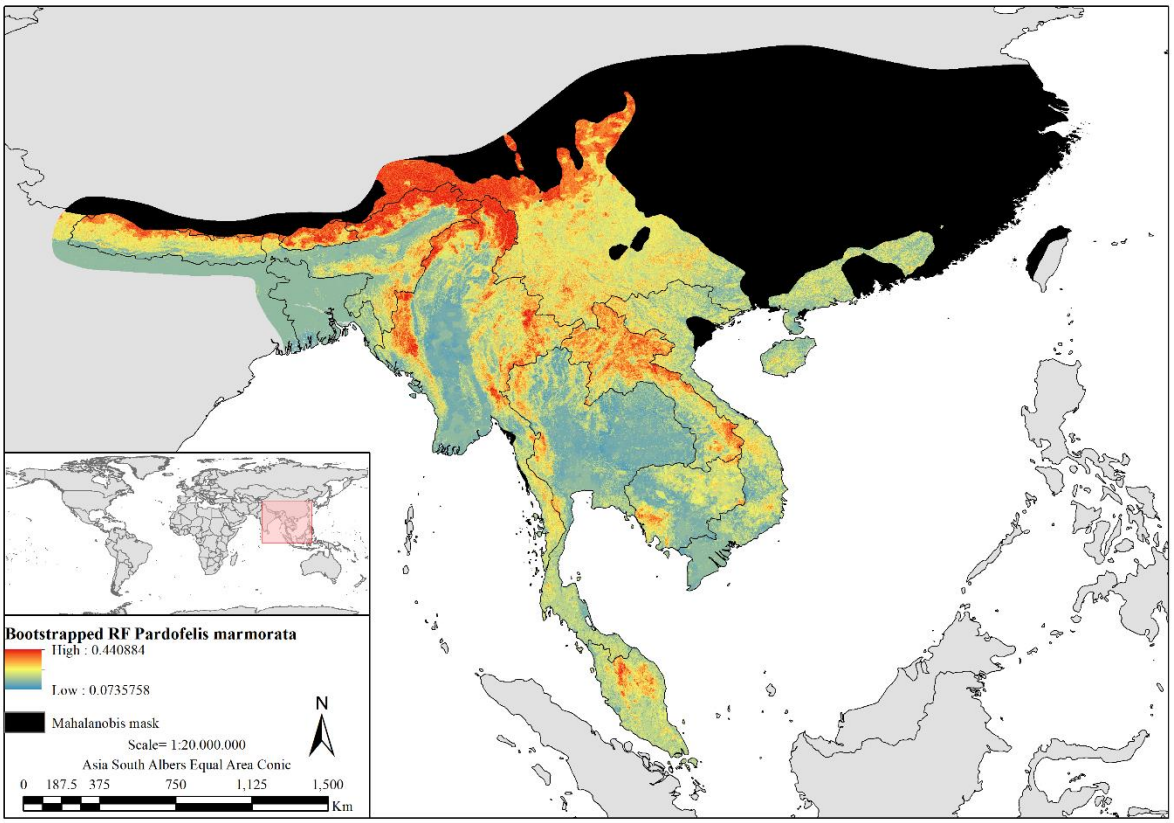



**Figure S17.** Probability surface for marbled cat (*Pardofelis marmorata*) in mainland Southeast Asia, produced with bootstrapped RF.

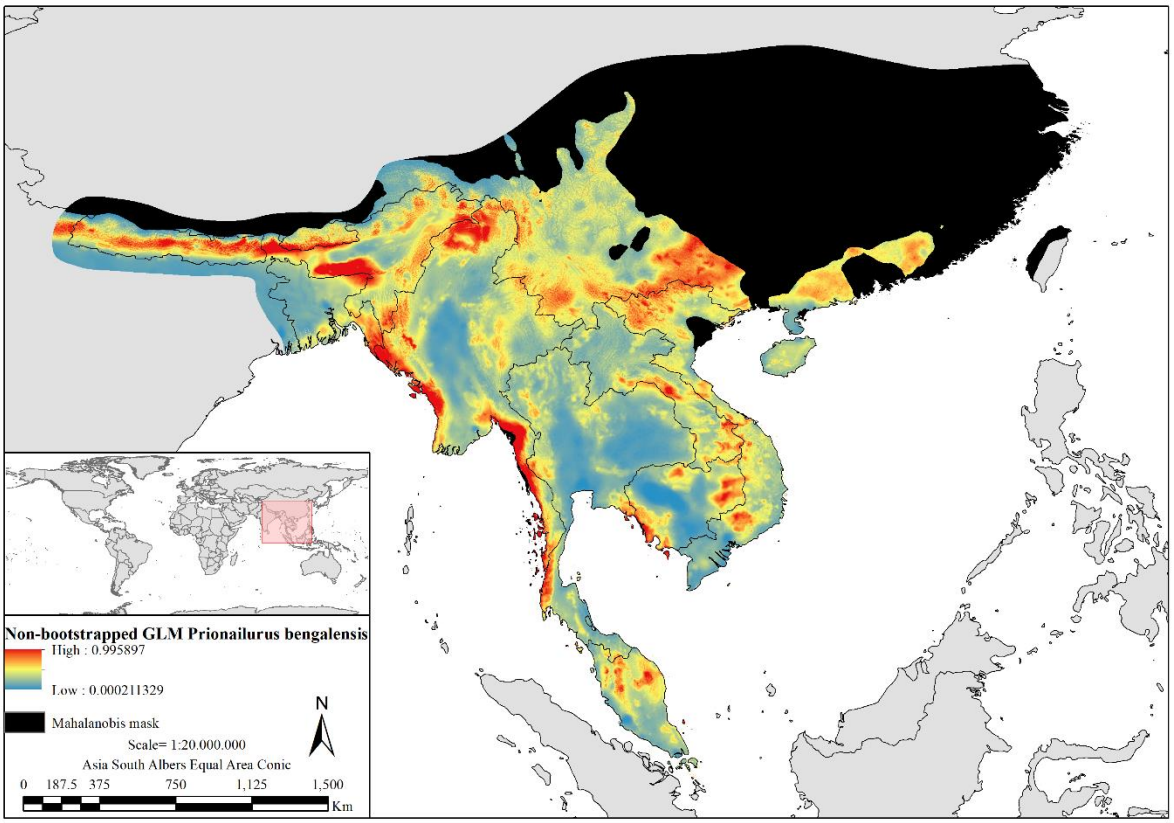



**Figure S18.** Probability surface for leopard cat (*Prionailurus bengalensis*) in mainland Southeast Asia, produced with non-bootstrapped GLM.

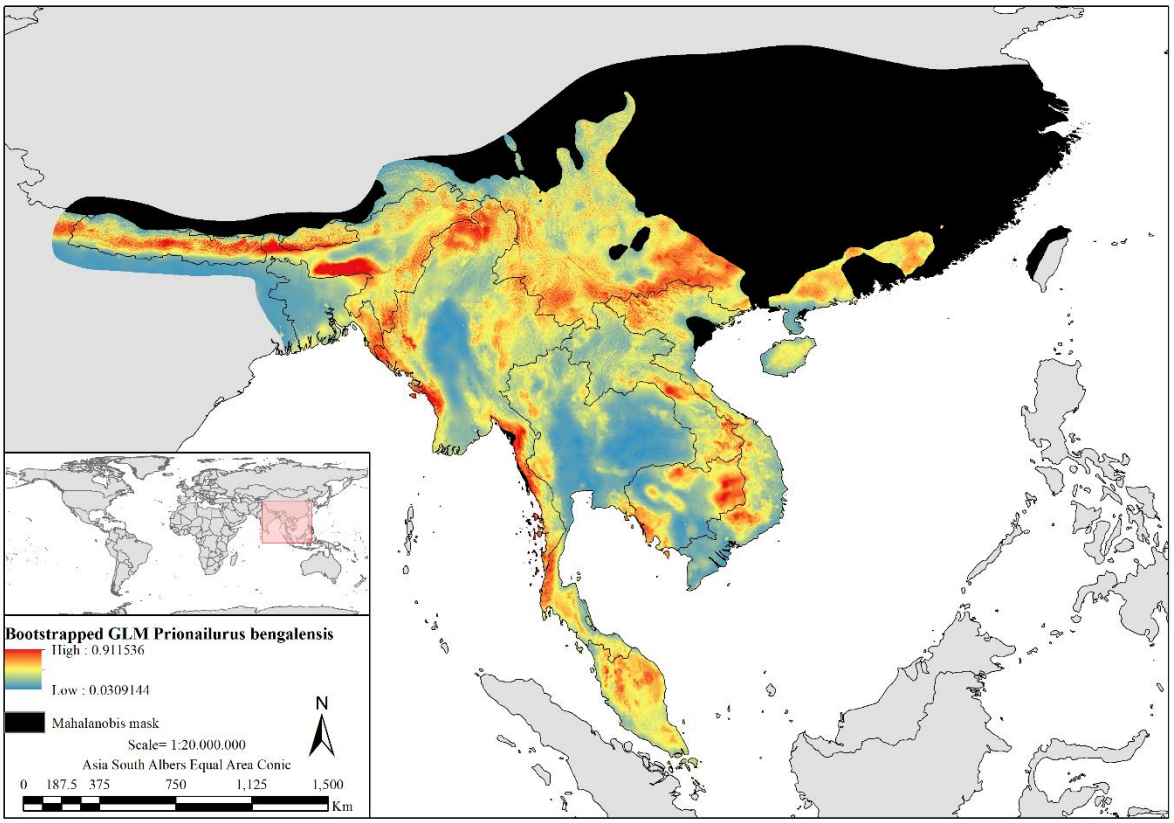



**Figure S19.** Probability surface for leopard cat (*Prionailurus bengalensis*) in mainland Southeast Asia, produced with bootstrapped GLM.

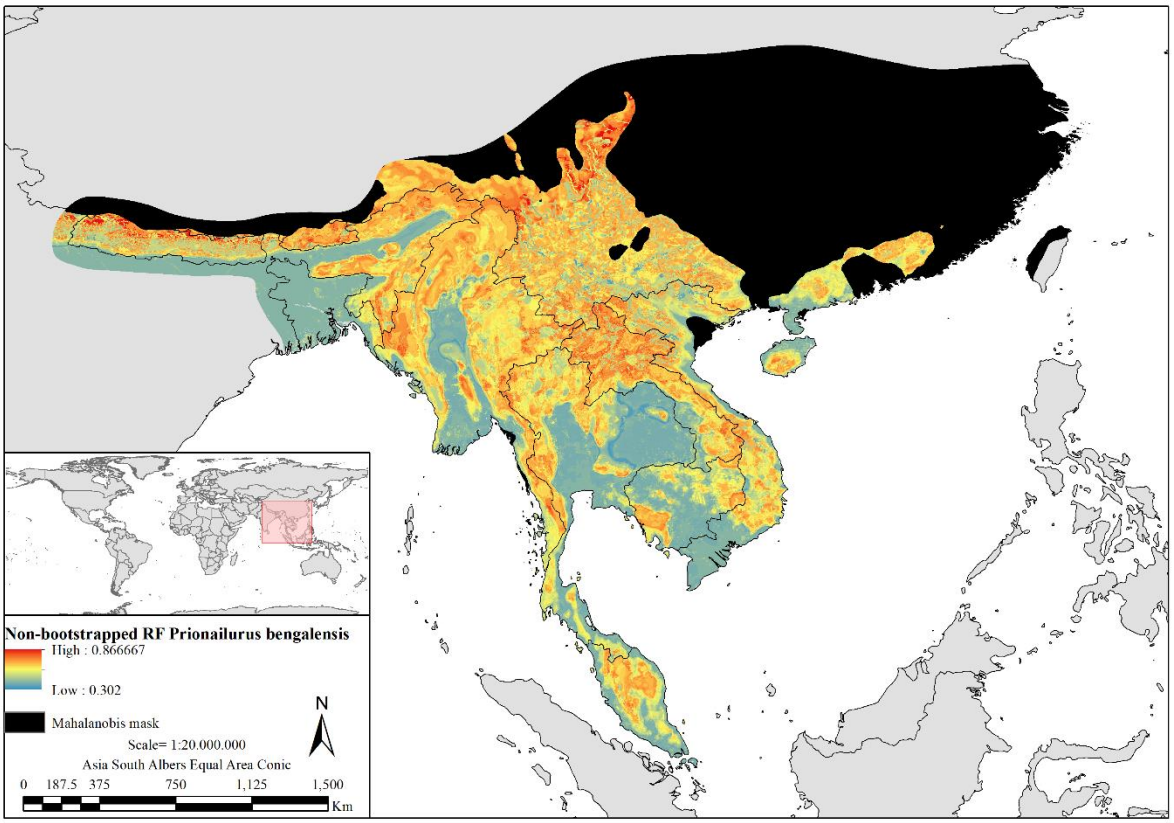



**Figure S20.** Probability surface for leopard cat (*Prionailurus bengalensis*) in mainland Southeast Asia, produced with non-bootstrapped RF.

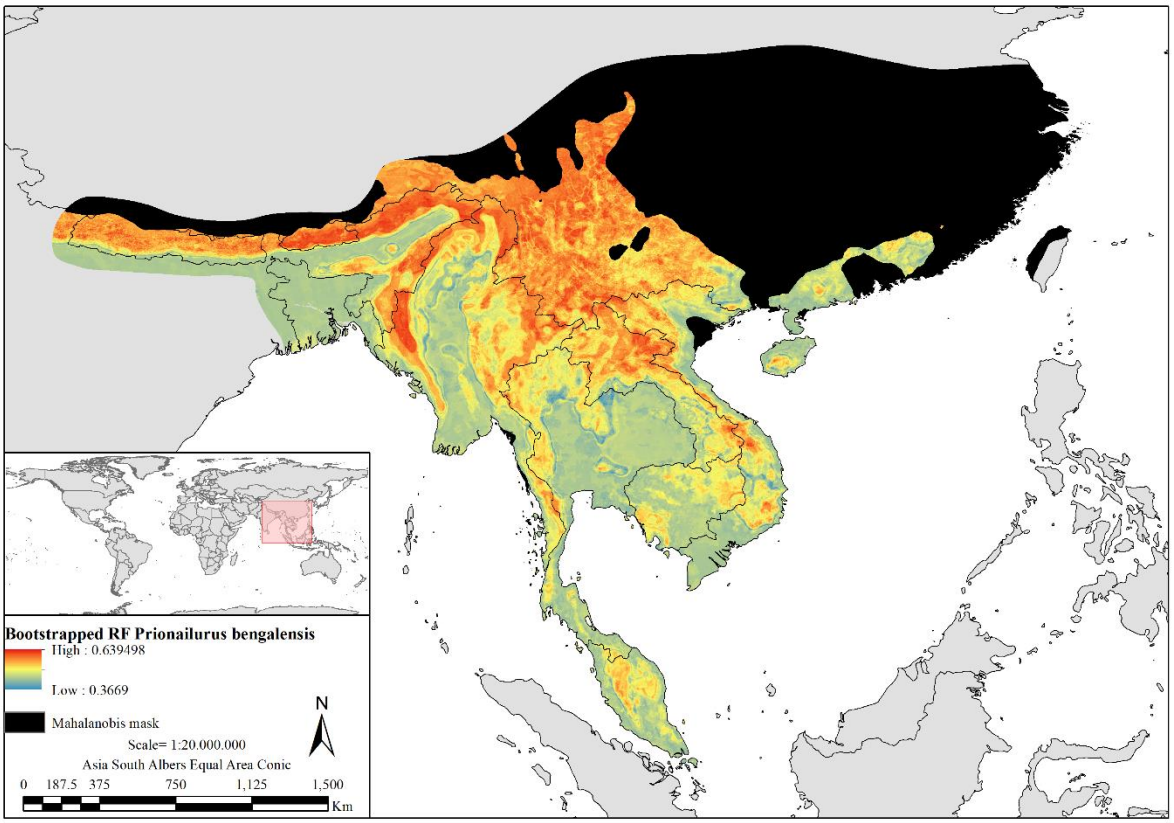



**Figure S21.** Probability surface for leopard cat (*Prionailurus bengalensis*) in mainland Southeast Asia, produced with bootstrapped RF.

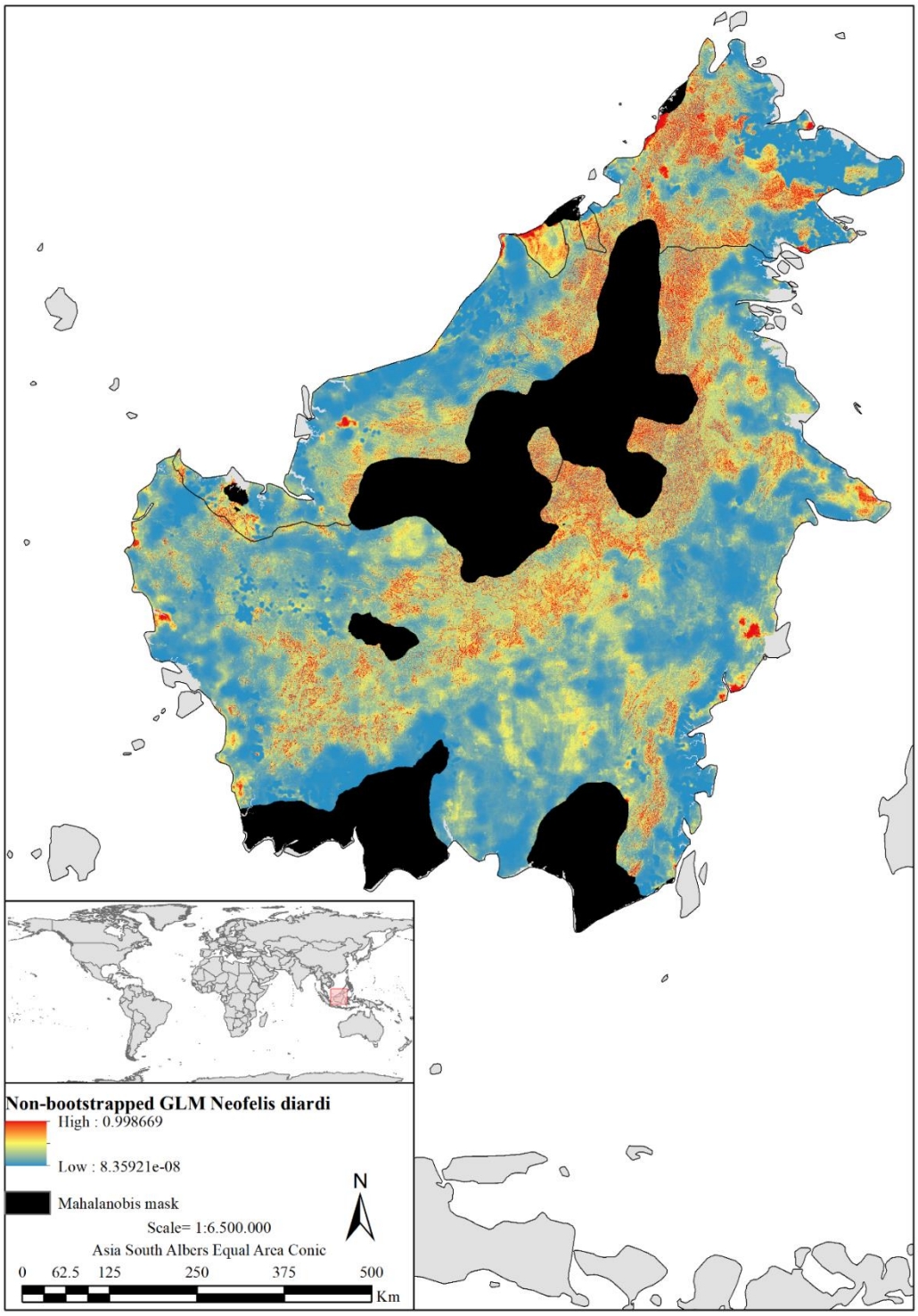



**Figure S22.** Probability surface for Sunda clouded leopard (*Neofelis diardi*) in Borneo, produced with non-bootstrapped GLM.

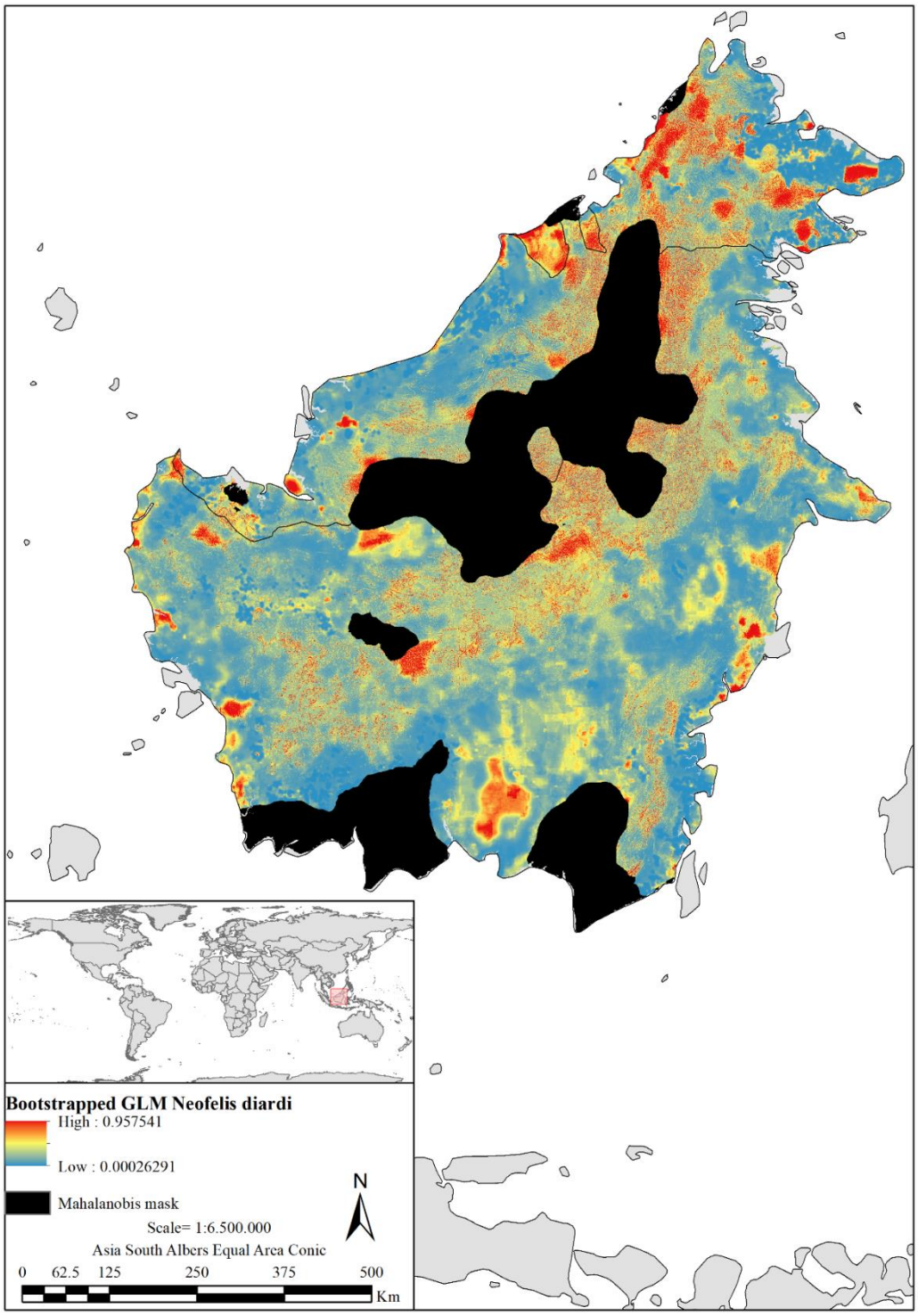



**Figure S23.** Probability surface for Sunda clouded leopard (*Neofelis diardi*) in Borneo, produced with bootstrapped GLM.

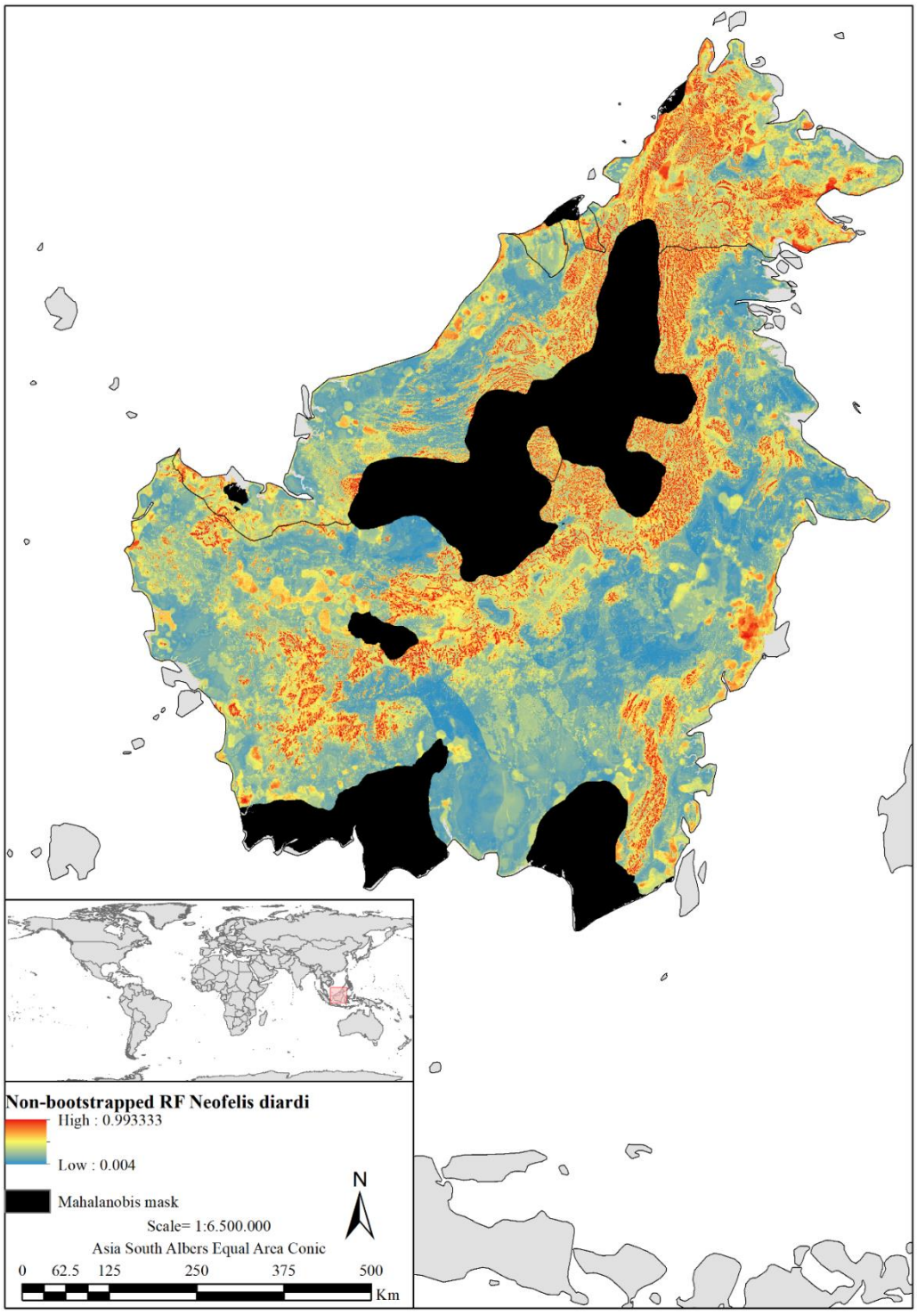



**Figure S24.** Probability surface for Sunda clouded leopard (*Neofelis diardi*) in Borneo, produced with non-bootstrapped RF.

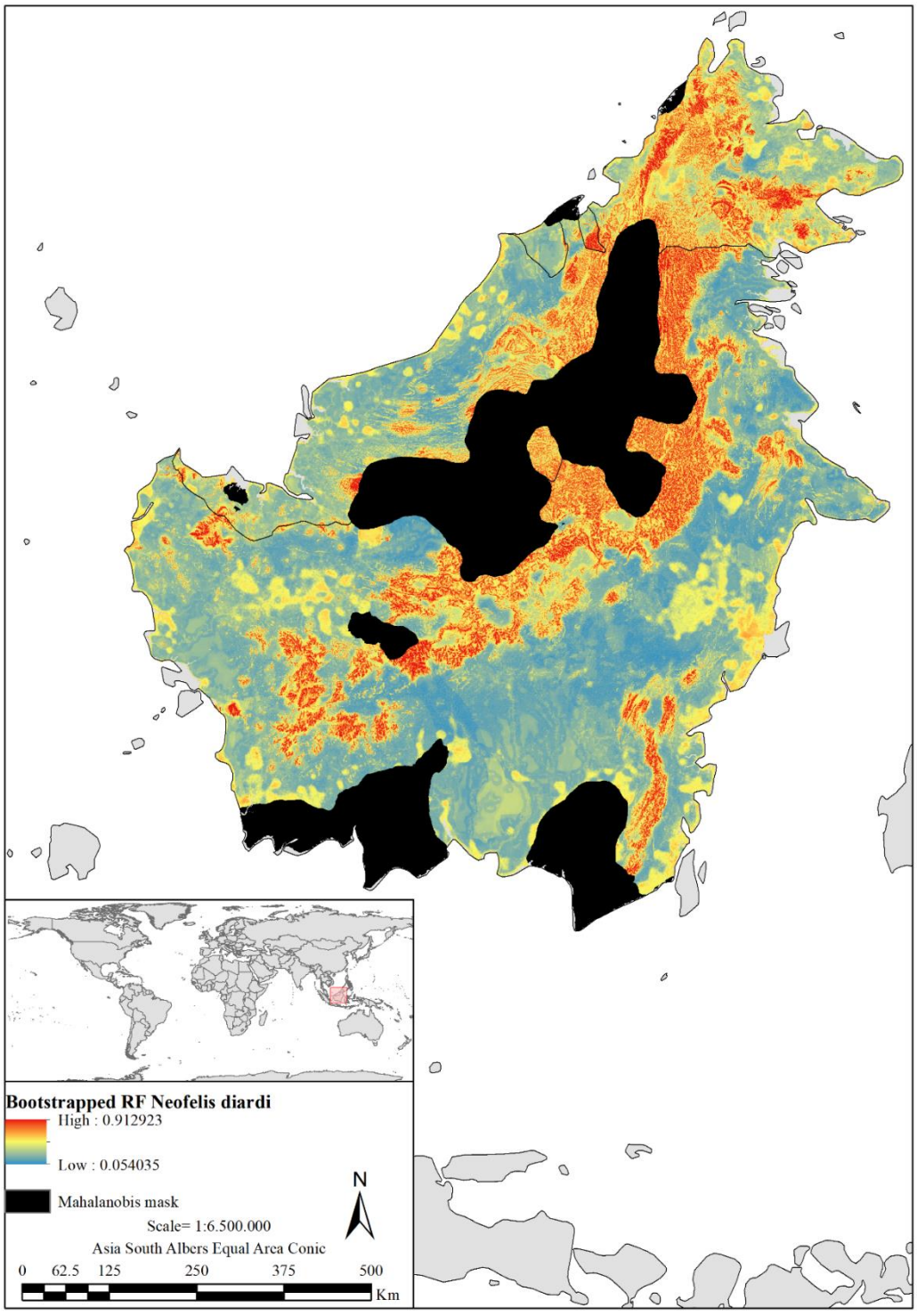



**Figure S25.** Probability surface for Sunda clouded leopard (*Neofelis diardi*) in Borneo, produced with bootstrapped RF.

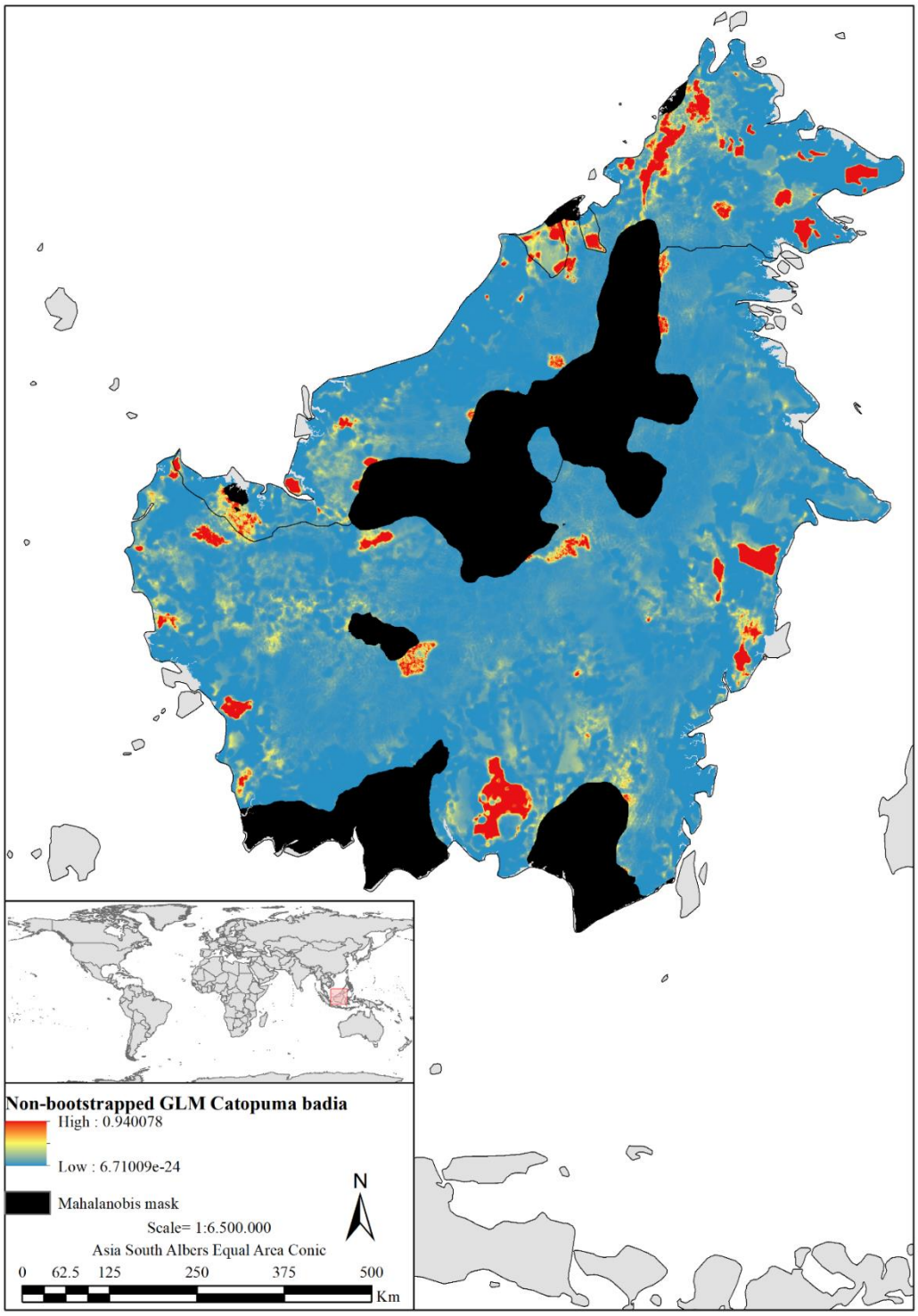



**Figure S26.** Probability surface for Borneo bay cat (*Catopuma badia*) in Borneo, produced with non-bootstrapped GLM.

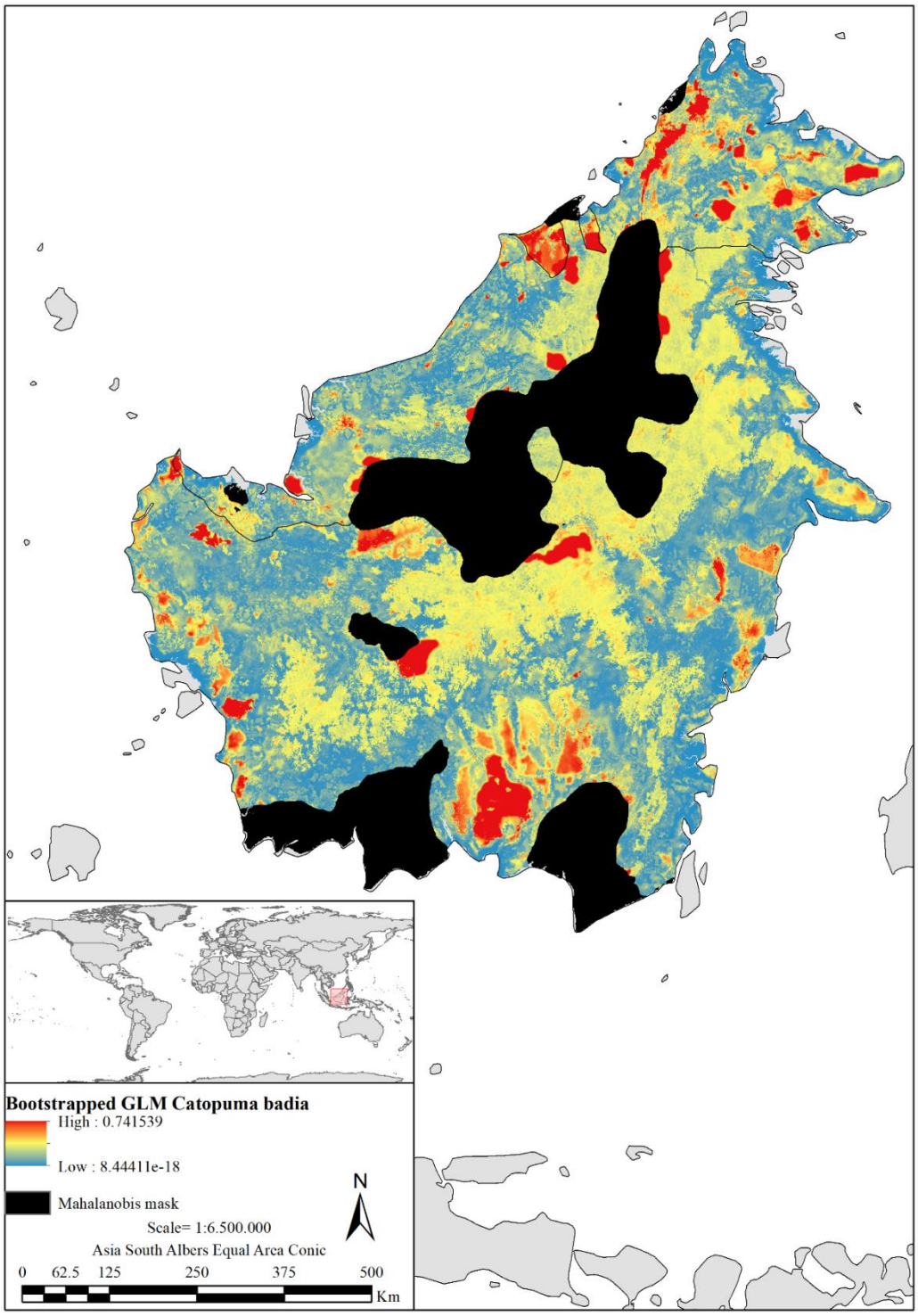



**Figure S27.** Probability surface for Borneo bay cat (*Catopuma badia*) in Borneo, produced with bootstrapped GLM.

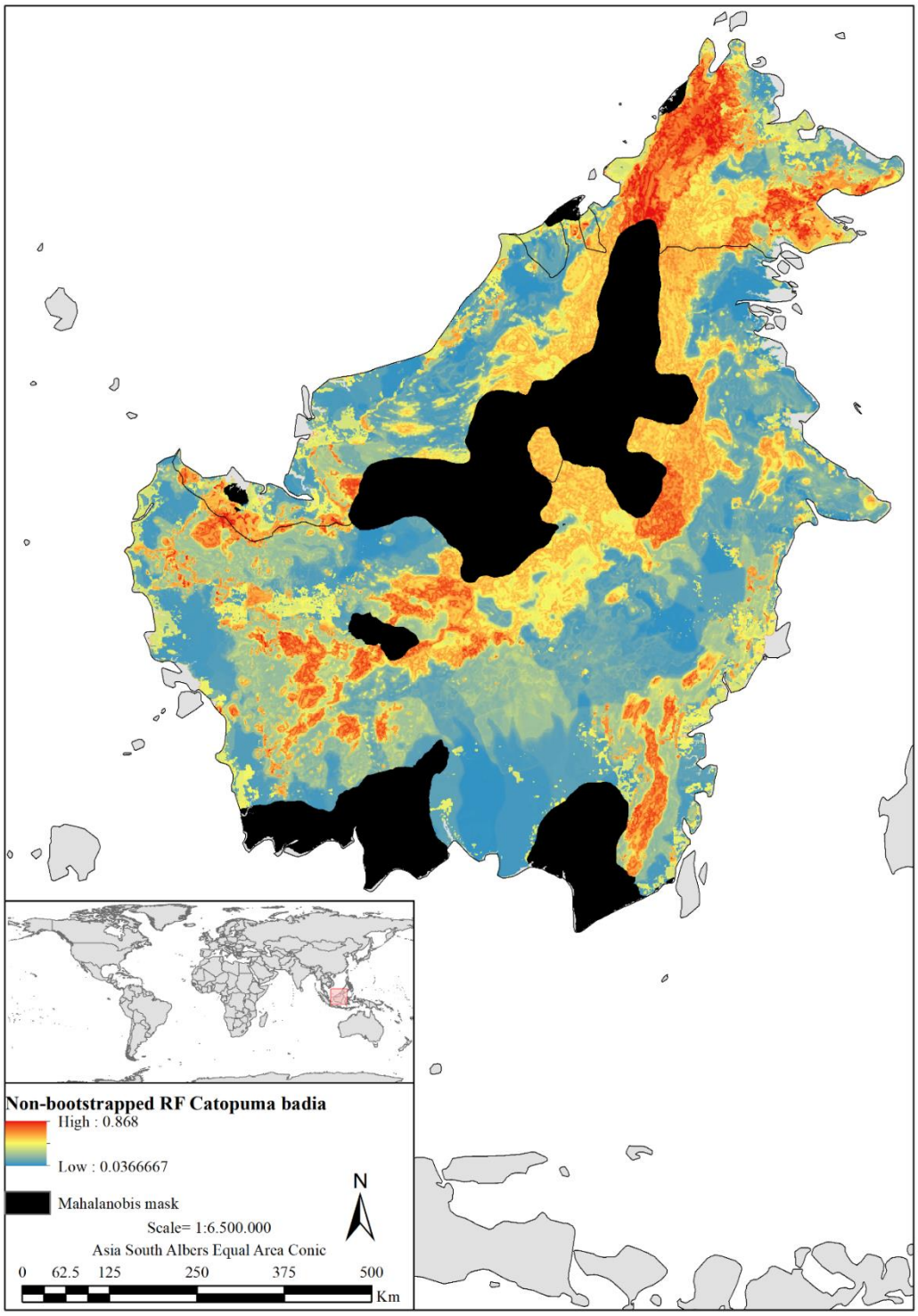



**Figure S28.** Probability surface for Borneo bay cat (*Catopuma badia*) in Borneo, produced with non-bootstrapped RF.

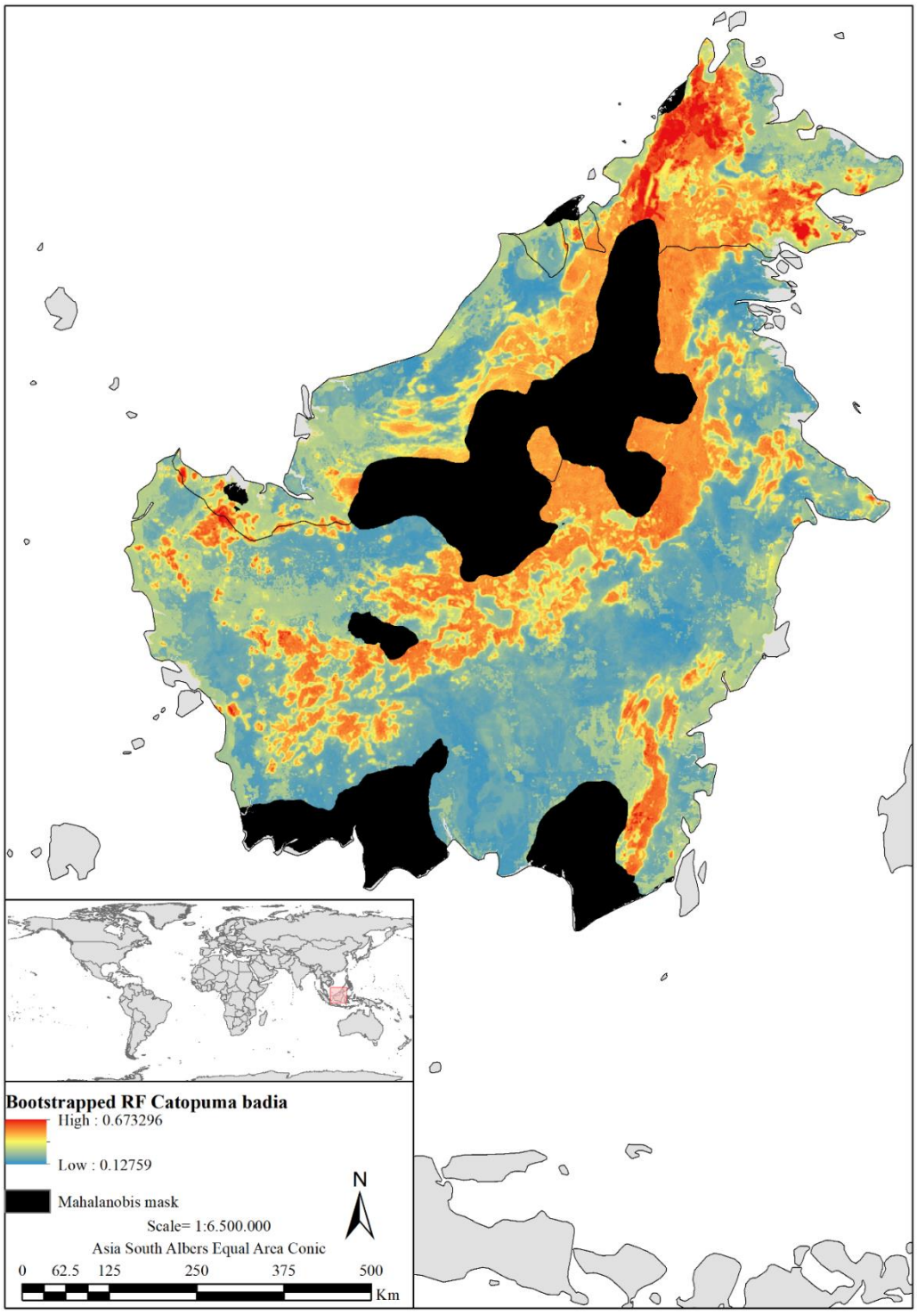



**Figure S29.** Probability surface for Borneo bay cat (*Catopuma badia*) in Borneo, produced with bootstrapped RF.

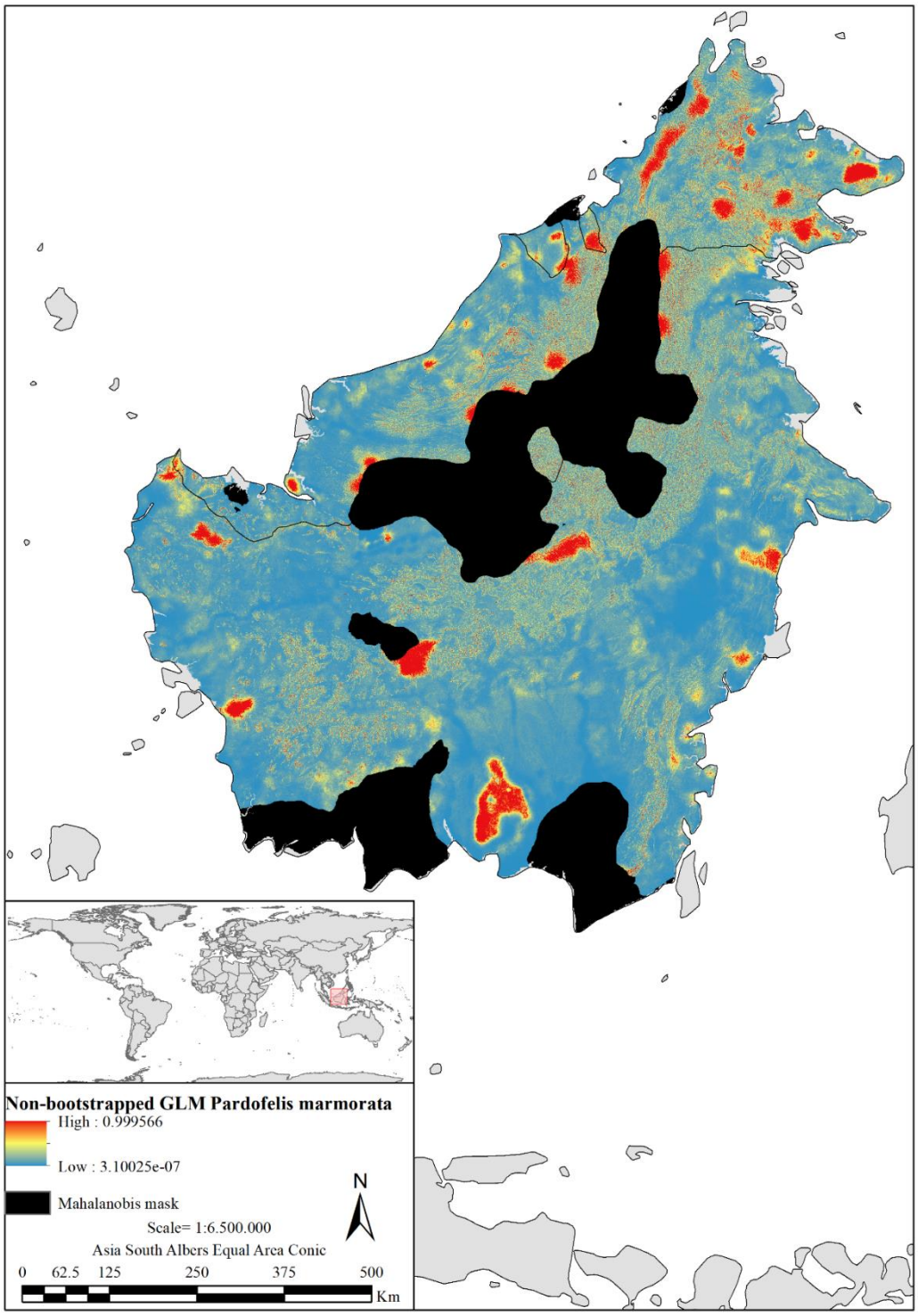



**Figure S30.** Probability surface for marbled cat (*Pardofelis marmorata*) in Borneo, produced with non-bootstrapped GLM.

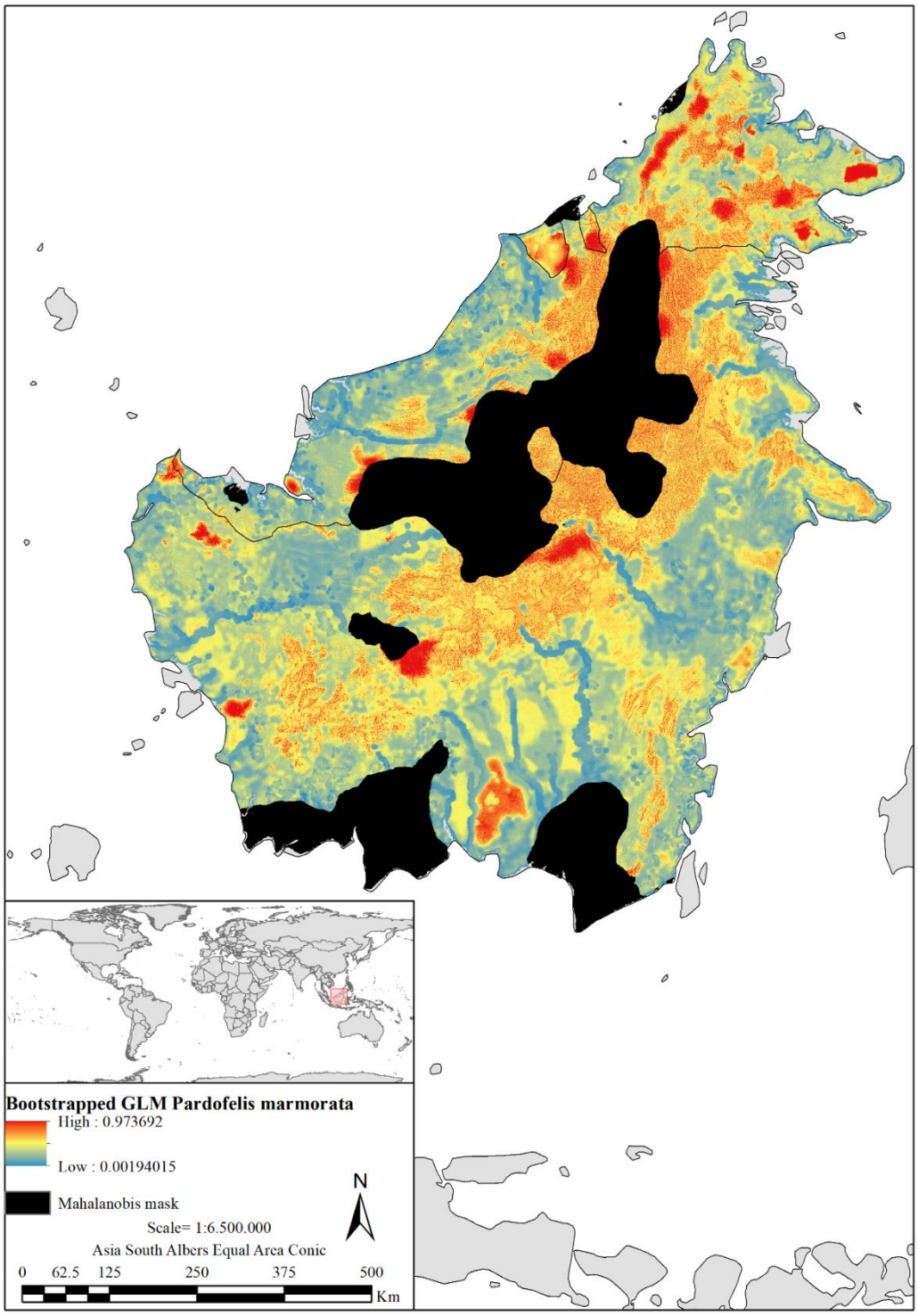



**Figure S31.** Probability surface for marbled cat (*Pardofelis marmorata*) in Borneo, produced with bootstrapped GLM.

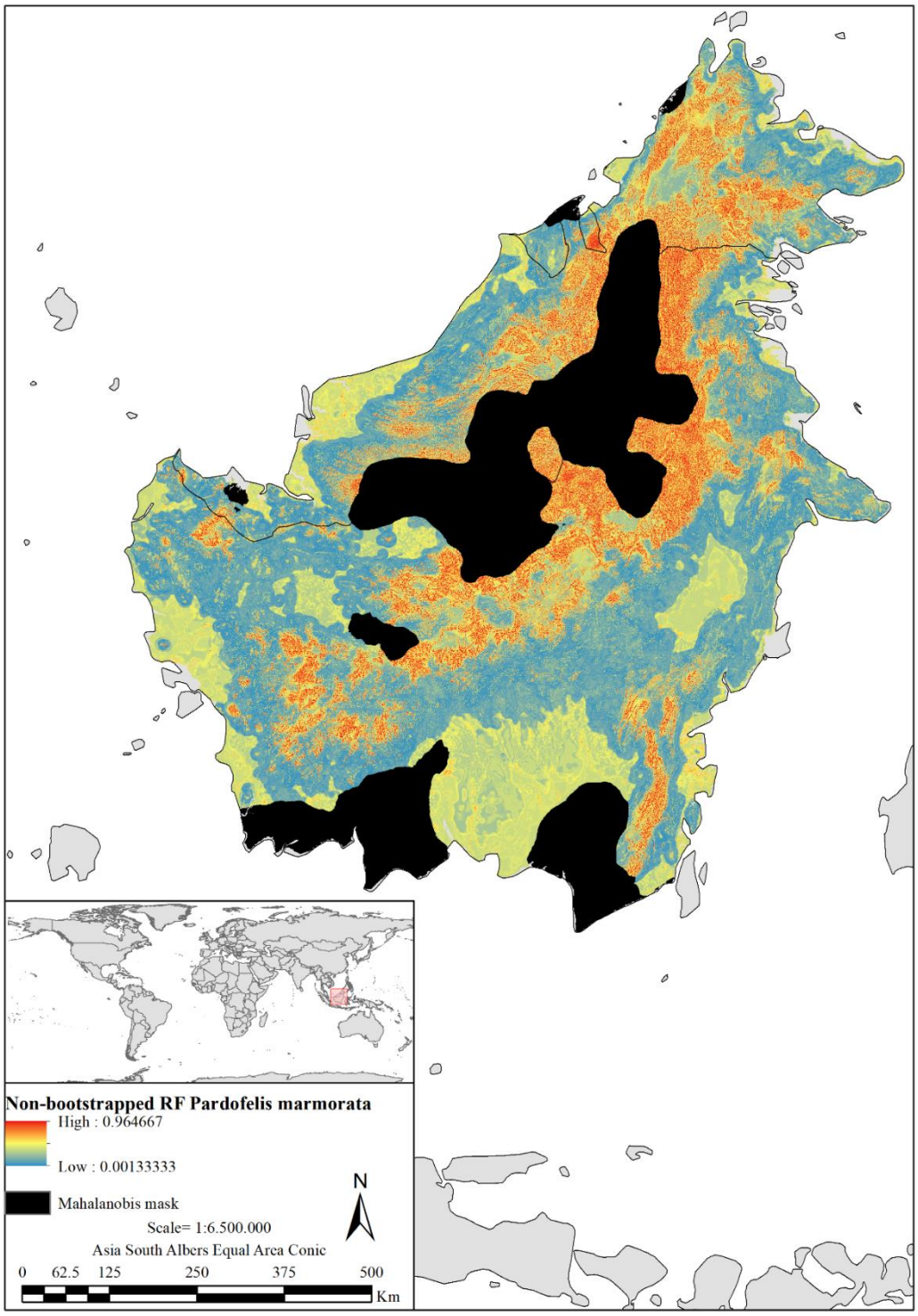



**Figure S32.** Probability surface for marbled cat (*Pardofelis marmorata*) in Borneo, produced with non-bootstrapped RF.

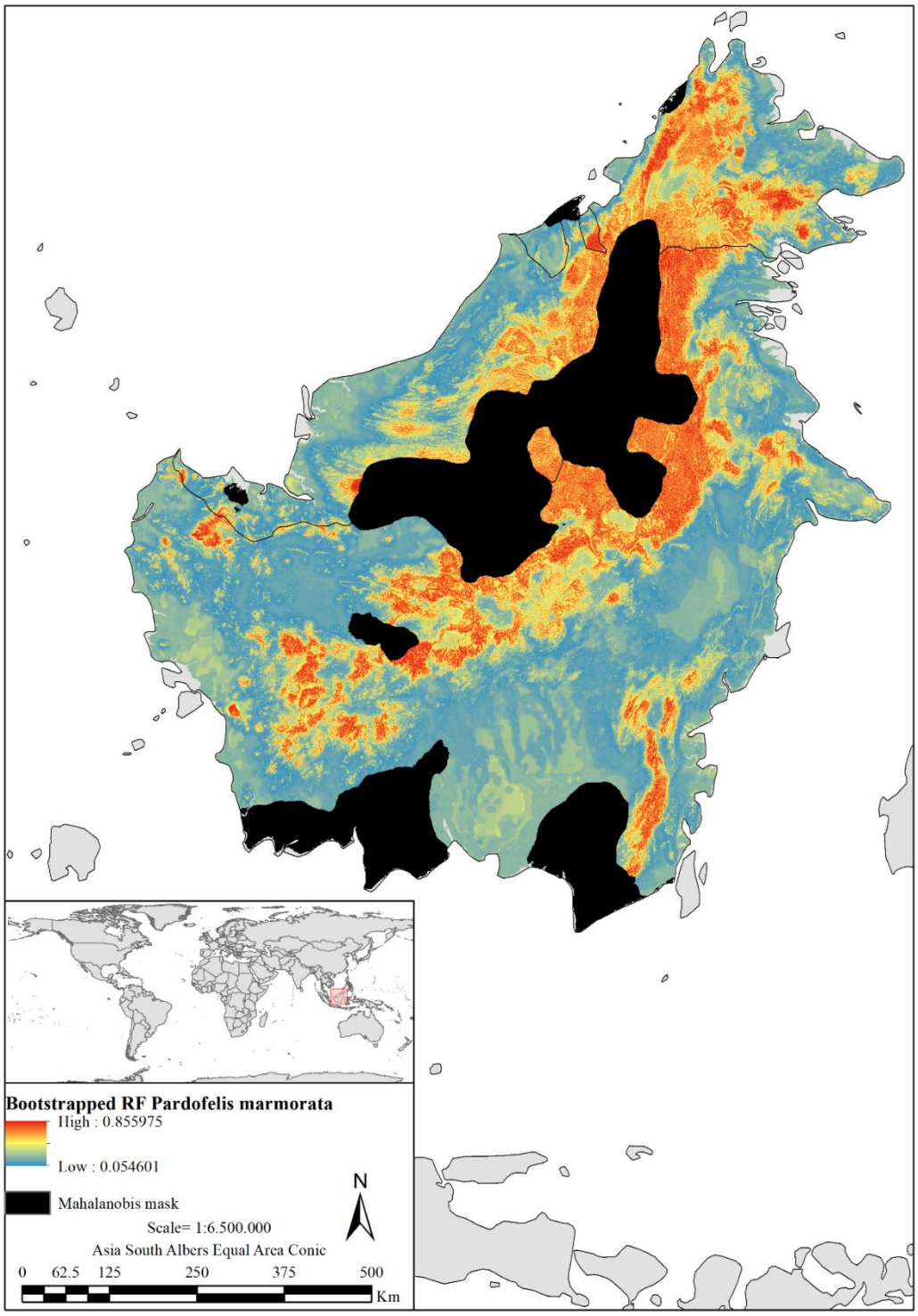



**Figure S33.** Probability surface for marbled cat (*Pardofelis marmorata*) in Borneo, produced with bootstrapped RF.

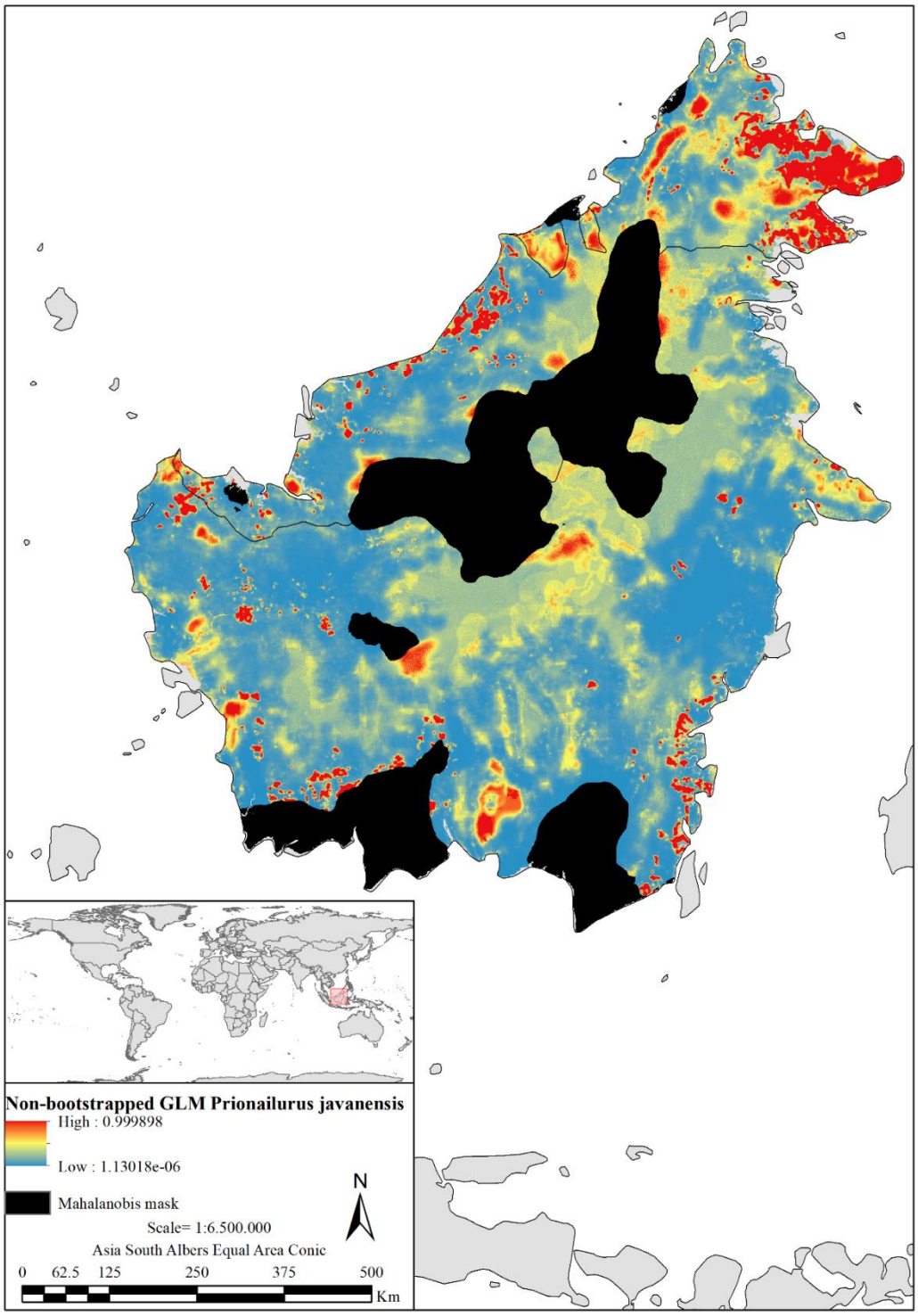



**Figure S34.** Probability surface for Sunda leopard cat (*Prionailurus javanensis*) in Borneo, produced with non-bootstrapped GLM.

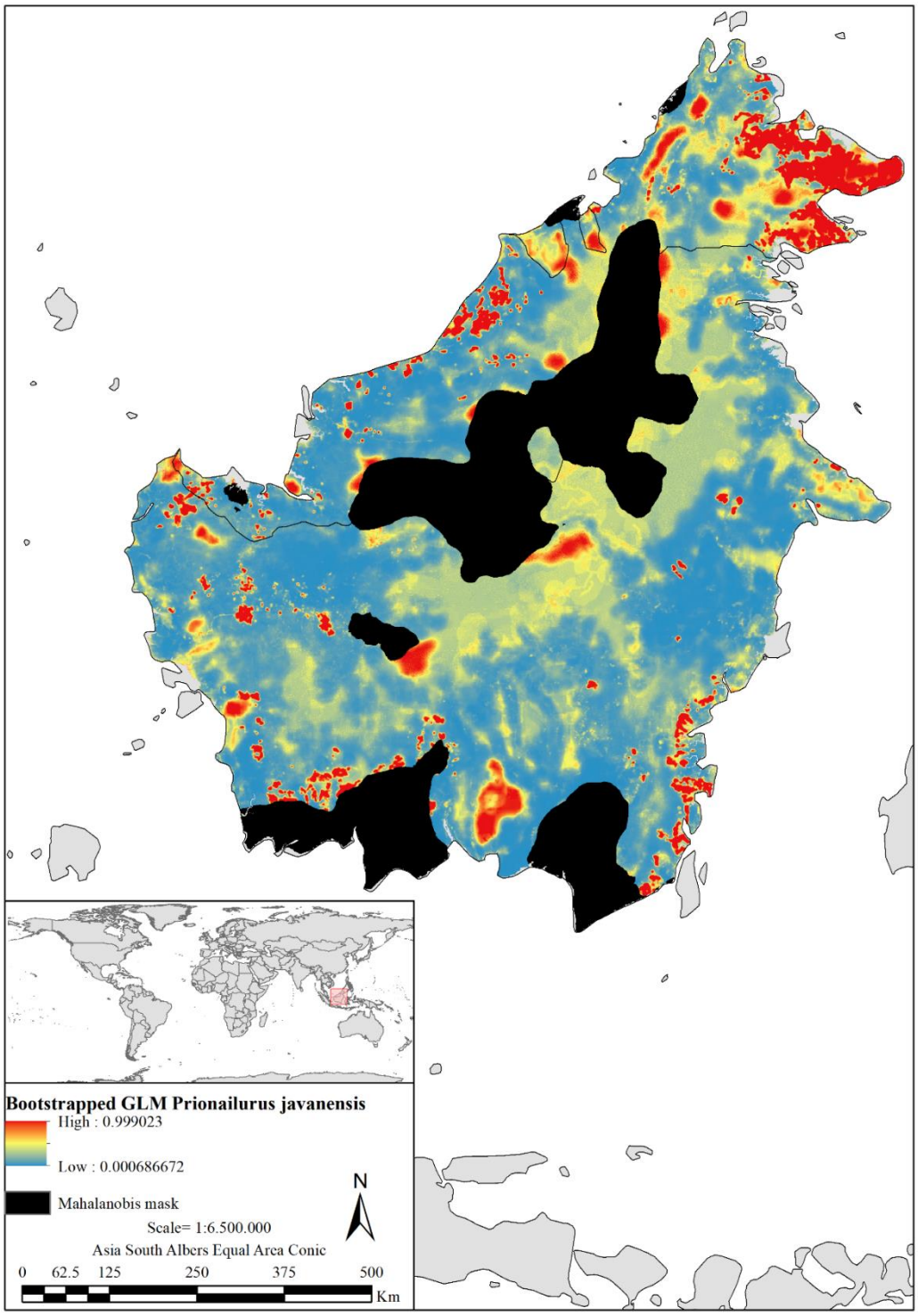



**Figure S35.** Probability surface for Sunda leopard cat (*Prionailurus javanensis*) in Borneo, produced with bootstrapped GLM.

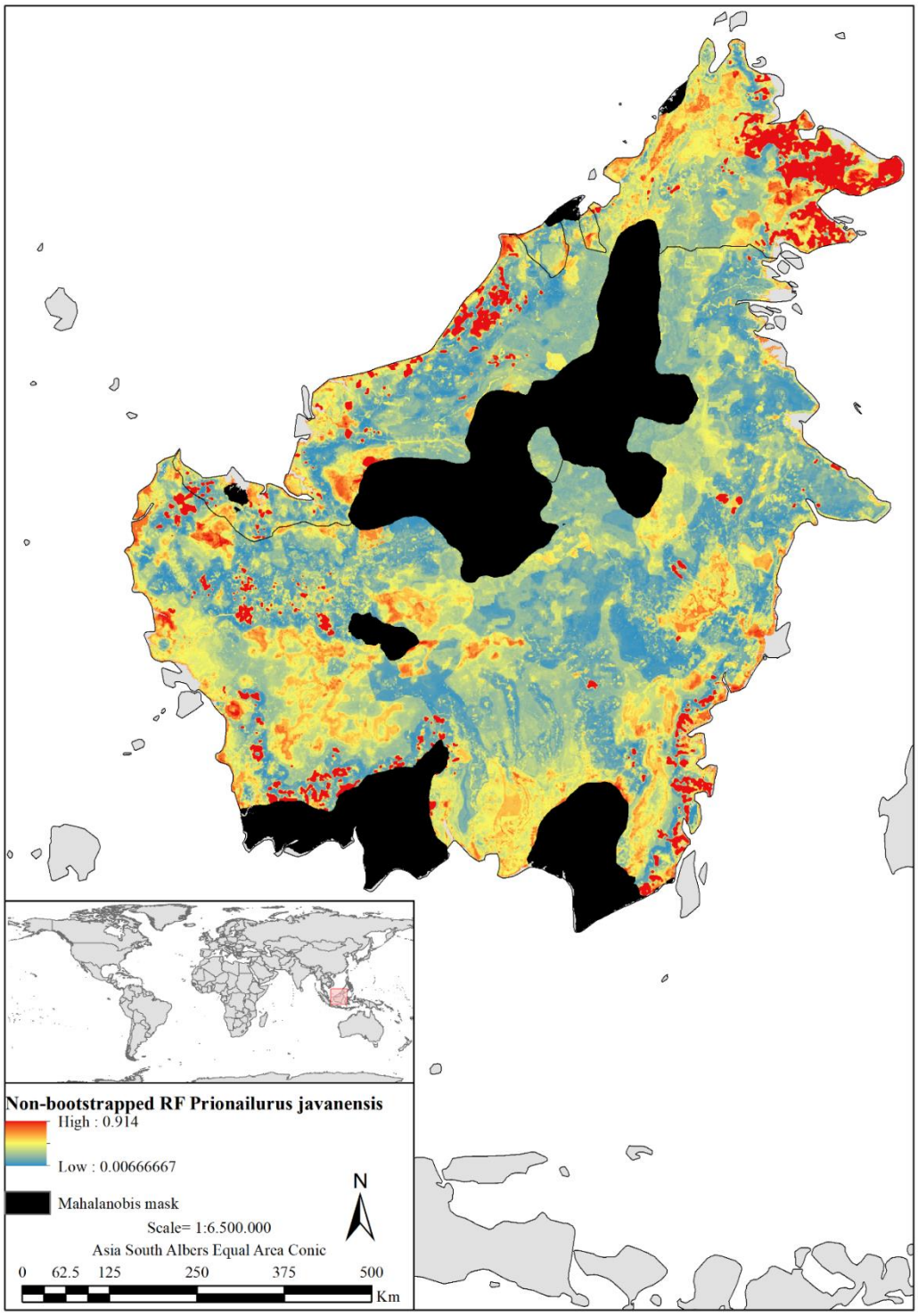



**Figure S36.** Probability surface for Sunda leopard cat (*Prionailurus javanensis*) in Borneo, produced with non-bootstrapped RF.

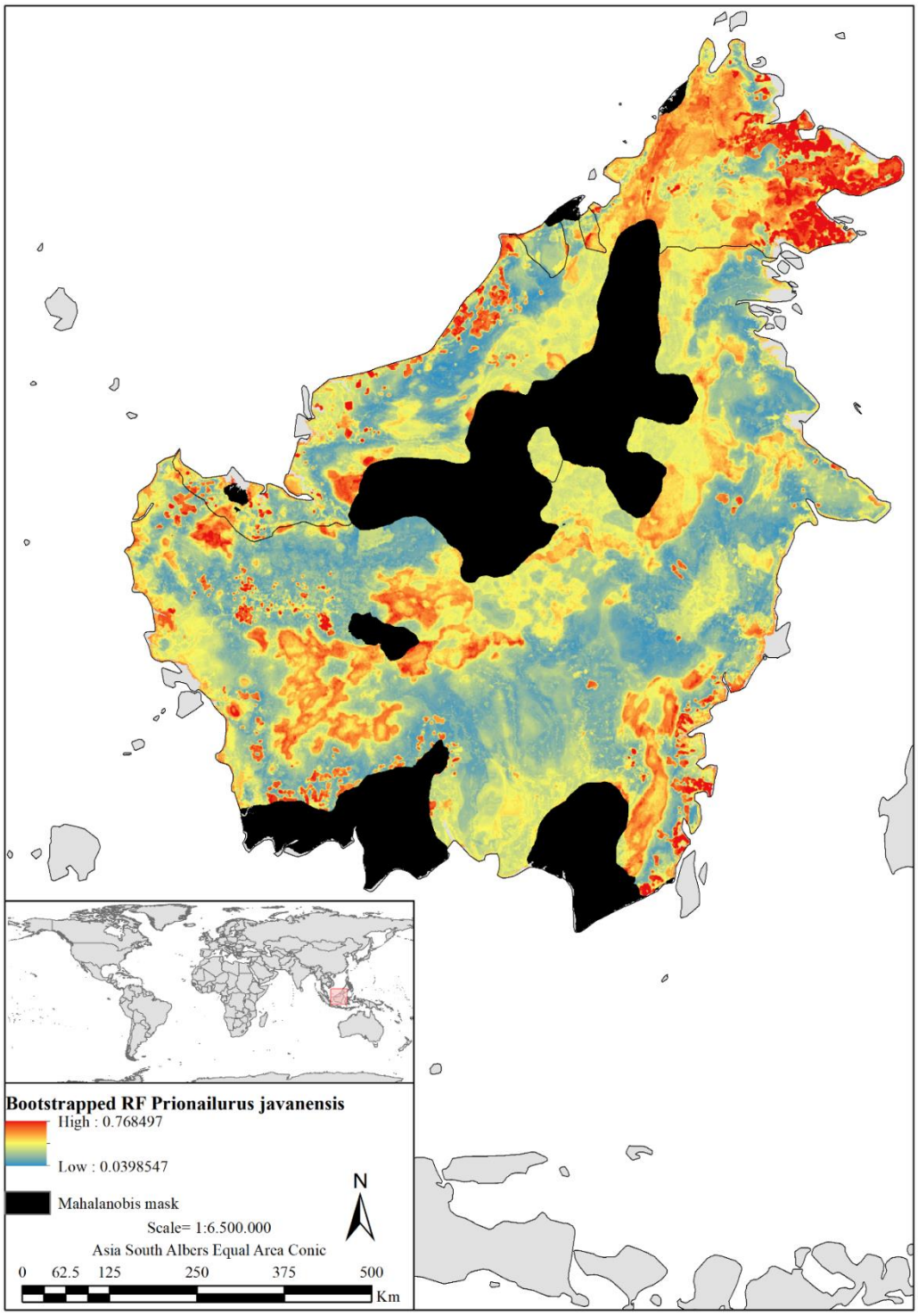



**Figure S37.** Probability surface for Sunda leopard cat (*Prionailurus javanensis*) in Borneo, produced with bootstrapped RF.

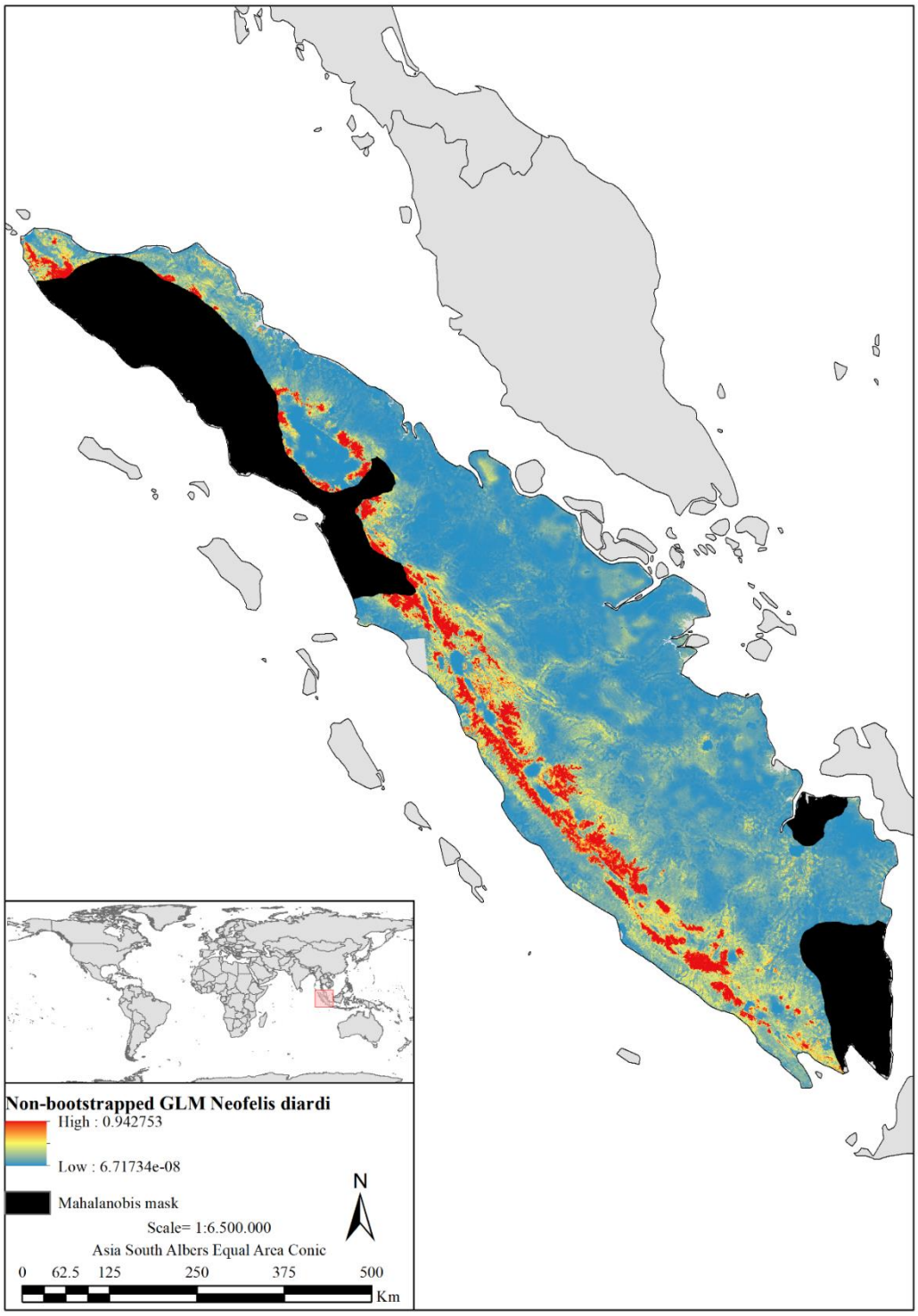



**Figure S38.** Probability surface for Sunda clouded leopard (*Neofelis diardi*) in Sumatra, produced with non-bootstrapped GLM.

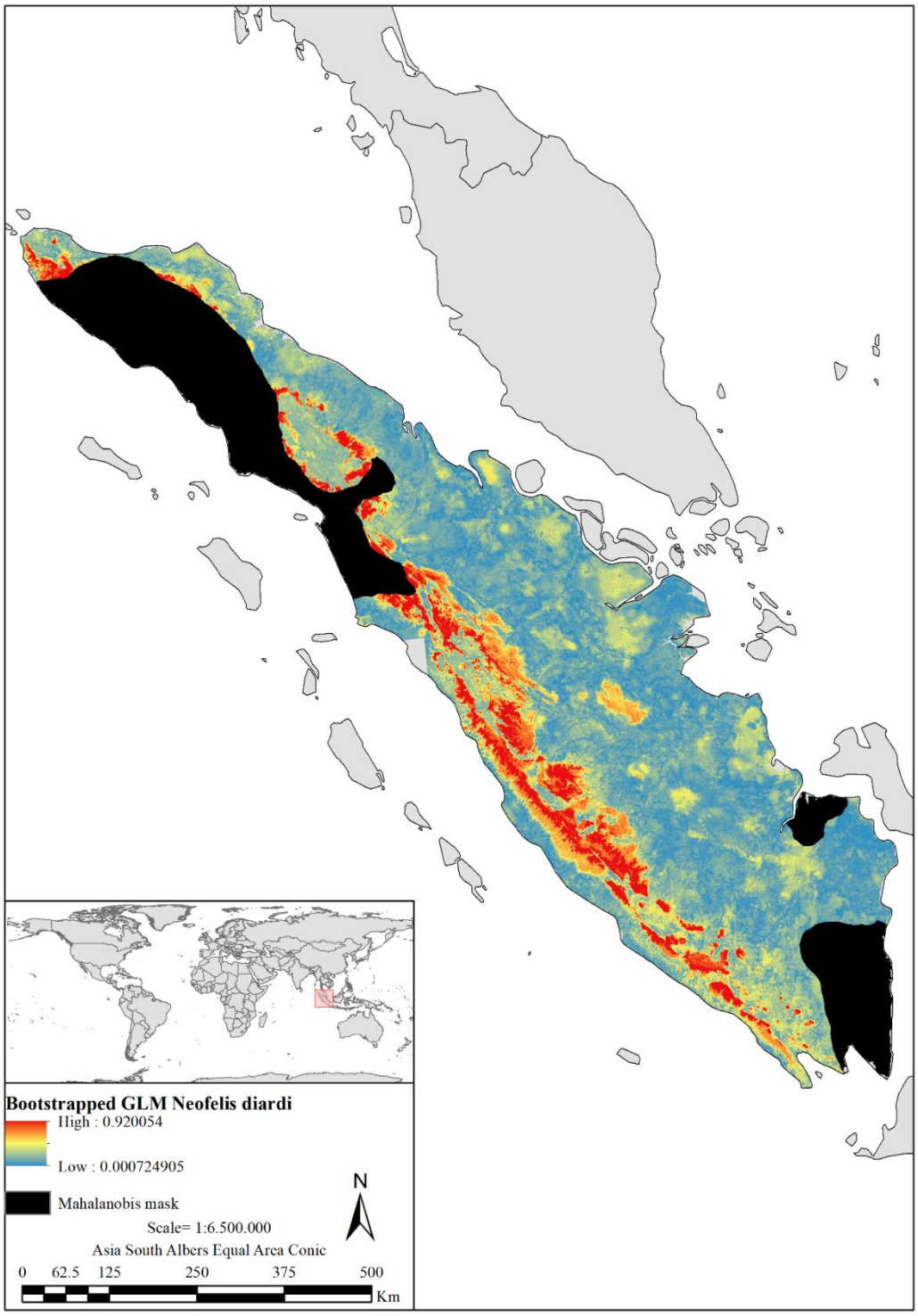



**Figure S39.** Probability surface for Sunda clouded leopard (*Neofelis diardi*) in Sumatra, produced with bootstrapped GLM.

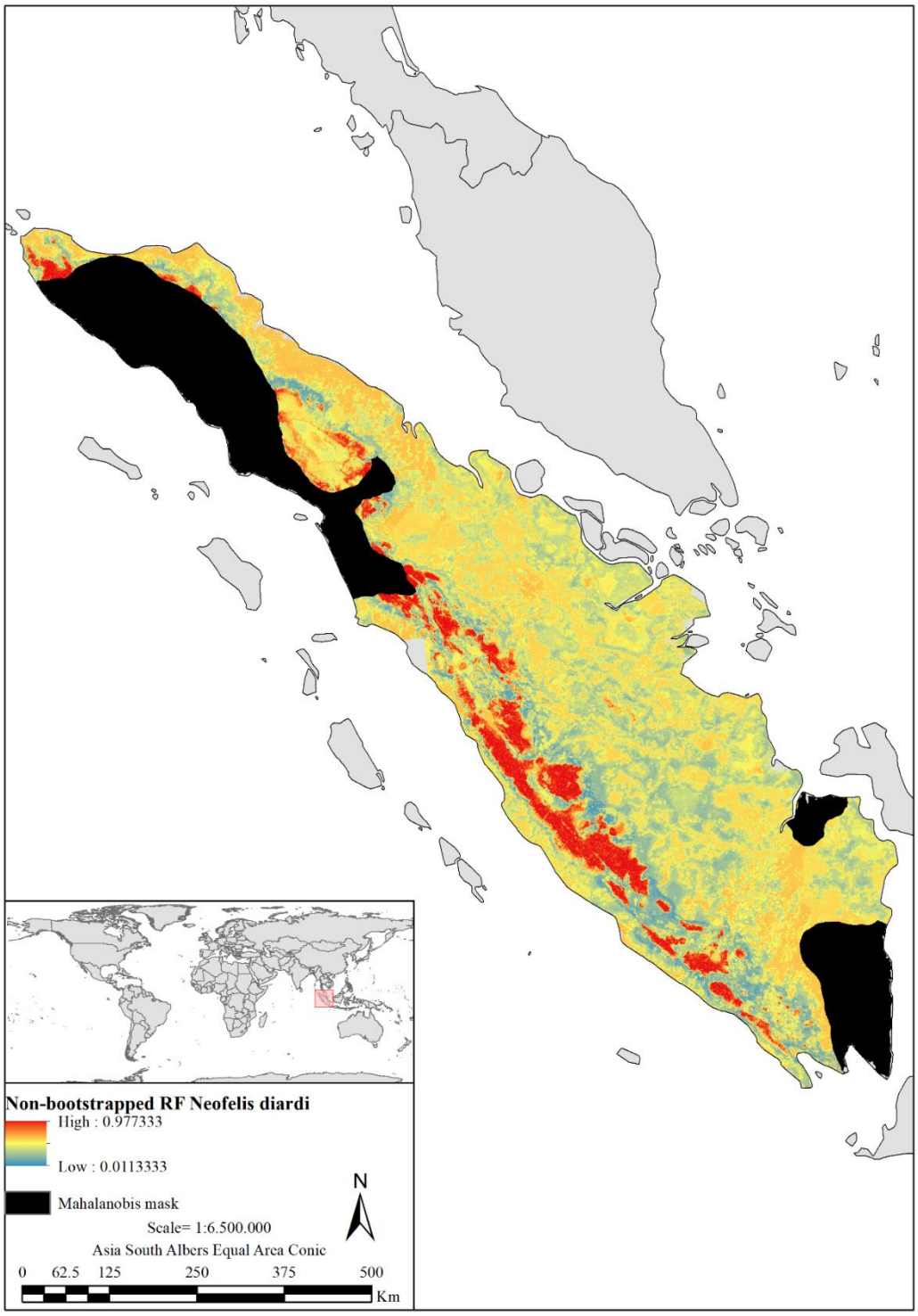



**Figure S40.** Probability surface for Sunda clouded leopard (*Neofelis diardi*) in Sumatra, produced with non-bootstrapped RF.

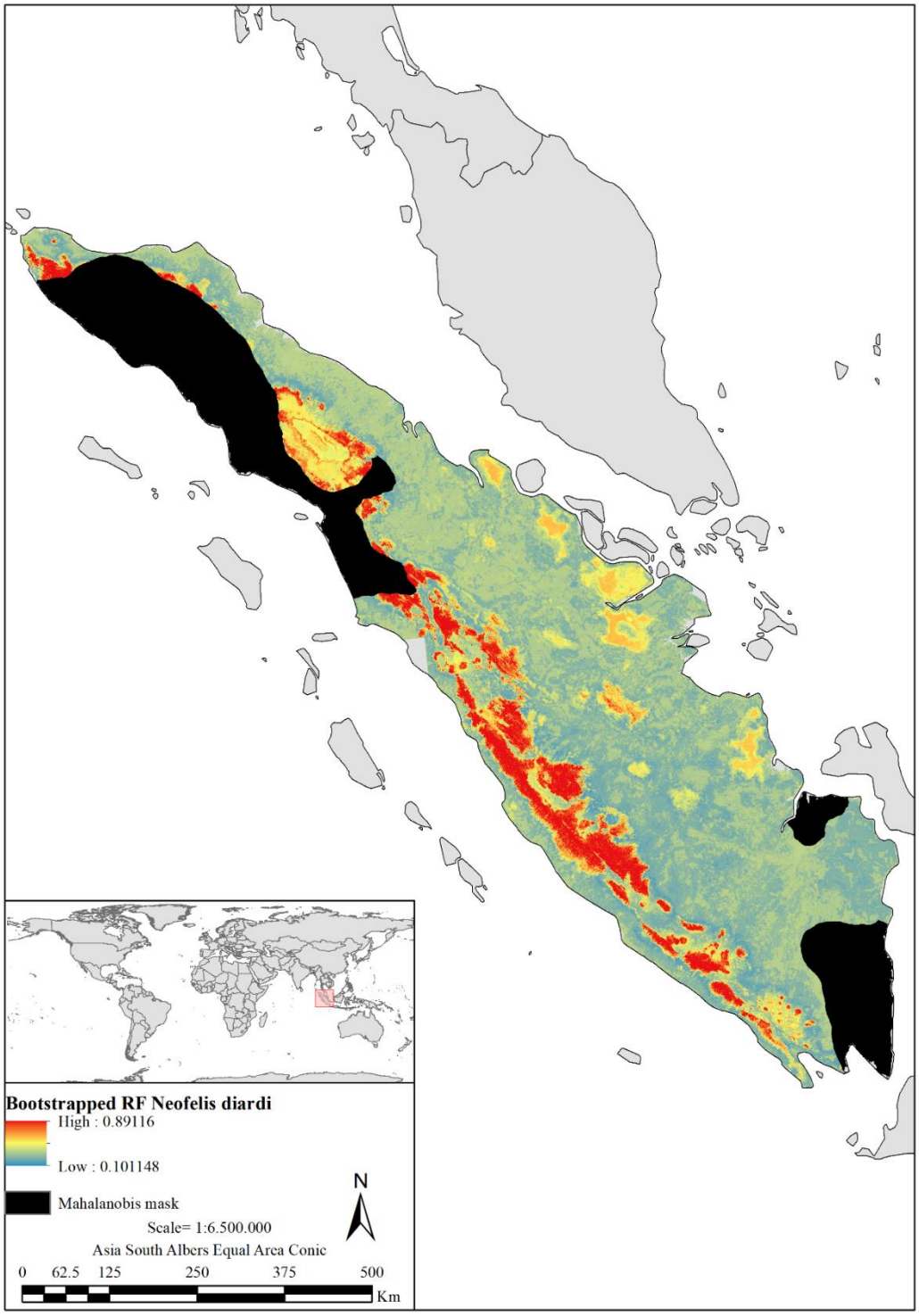



**Figure S41.** Probability surface for Sunda clouded leopard (*Neofelis diardi*) in Sumatra, produced with bootstrapped RF.

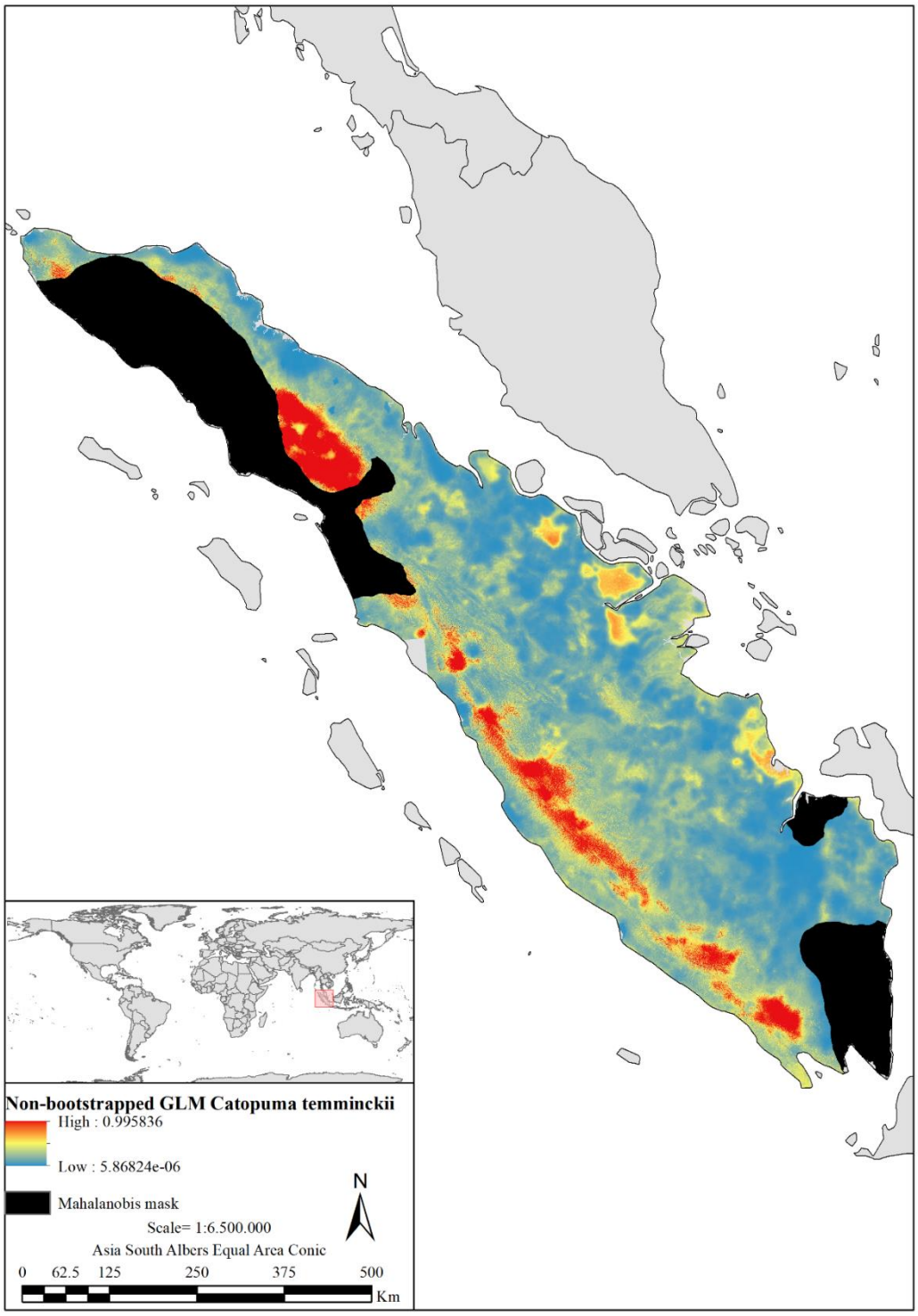



**Figure S42.** Probability surface for Asiatic golden cat (*Catopuma temminckii*) in Sumatra, produced with non-bootstrapped GLM.

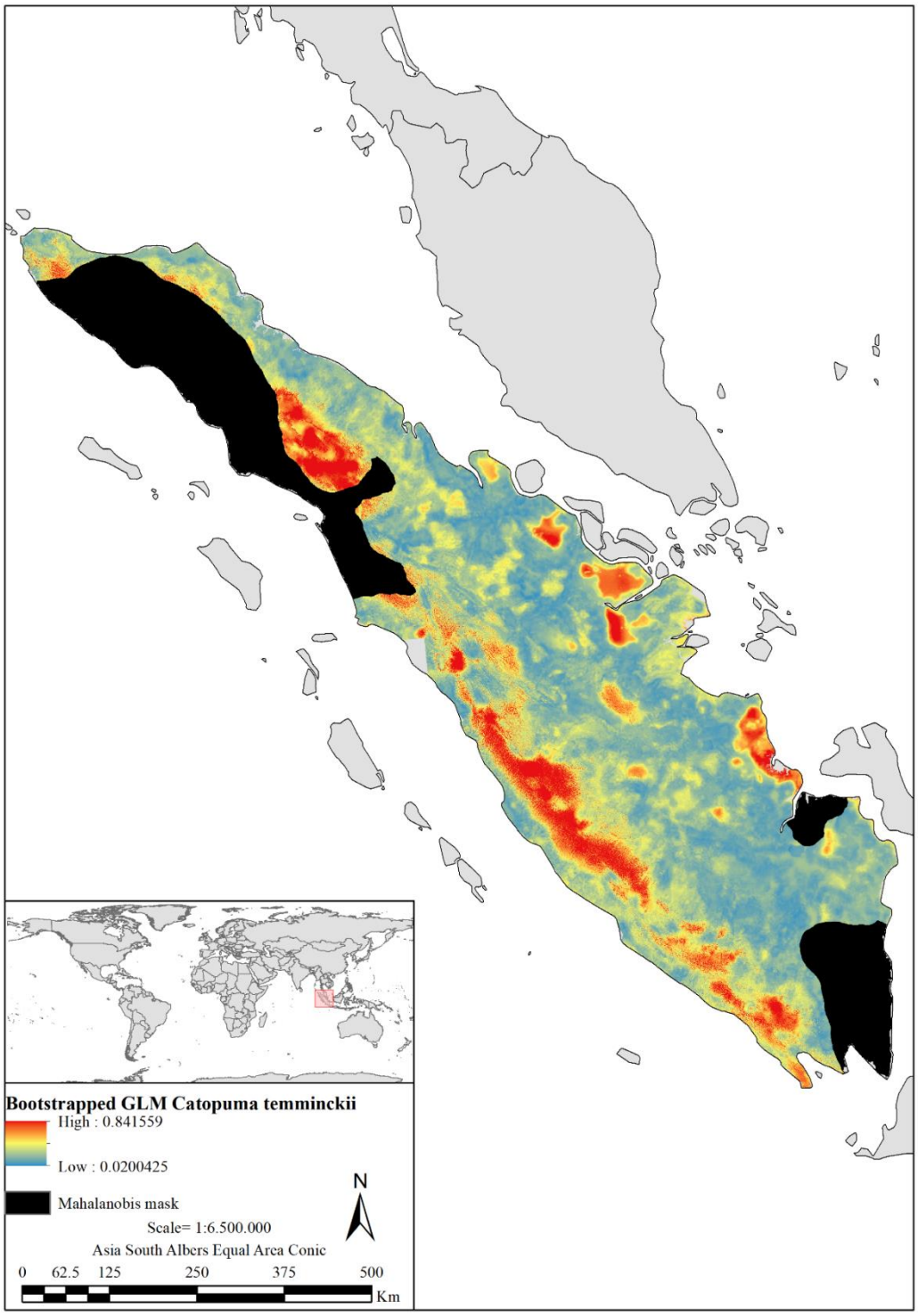



**Figure S43.** Probability surface for Asiatic golden cat (*Catopuma temminckii*) in Sumatra, produced with bootstrapped GLM.

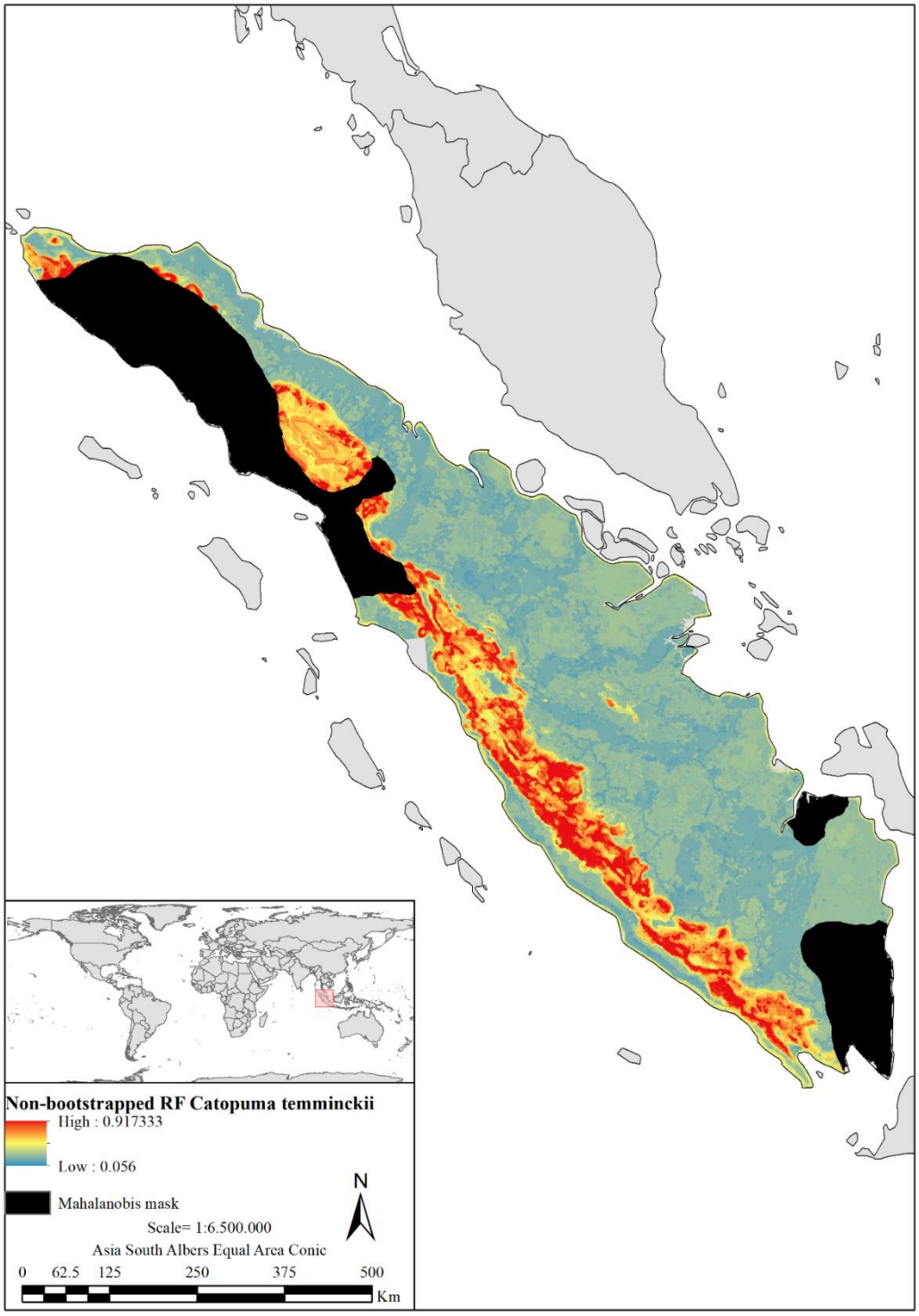



**Figure S44.** Probability surface for Asiatic golden cat (*Catopuma temminckii*) in Sumatra, produced with non-bootstrapped RF.

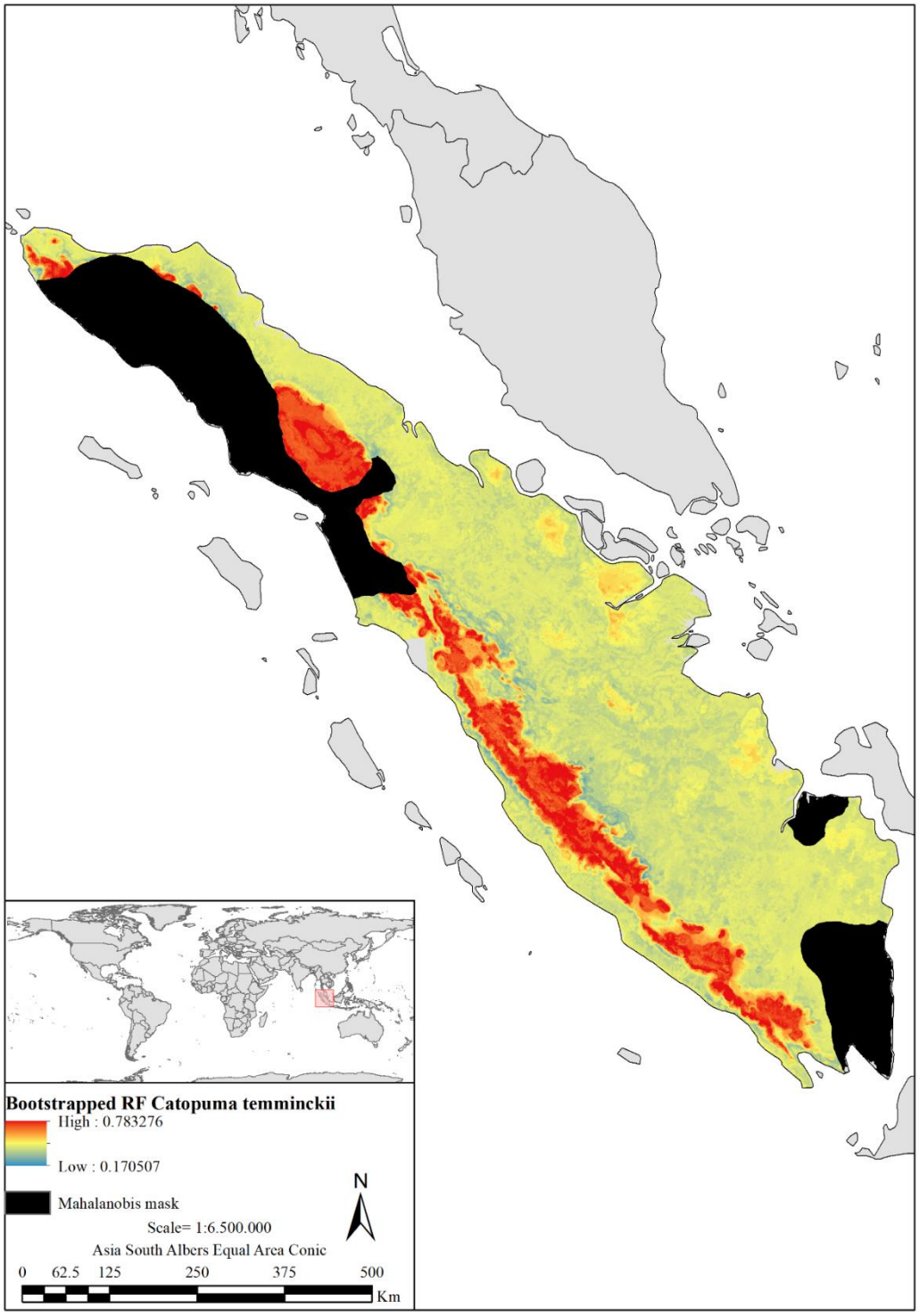



**Figure S45.** Probability surface for Asiatic golden cat (*Catopuma temminckii*) in Sumatra, produced with bootstrapped RF.

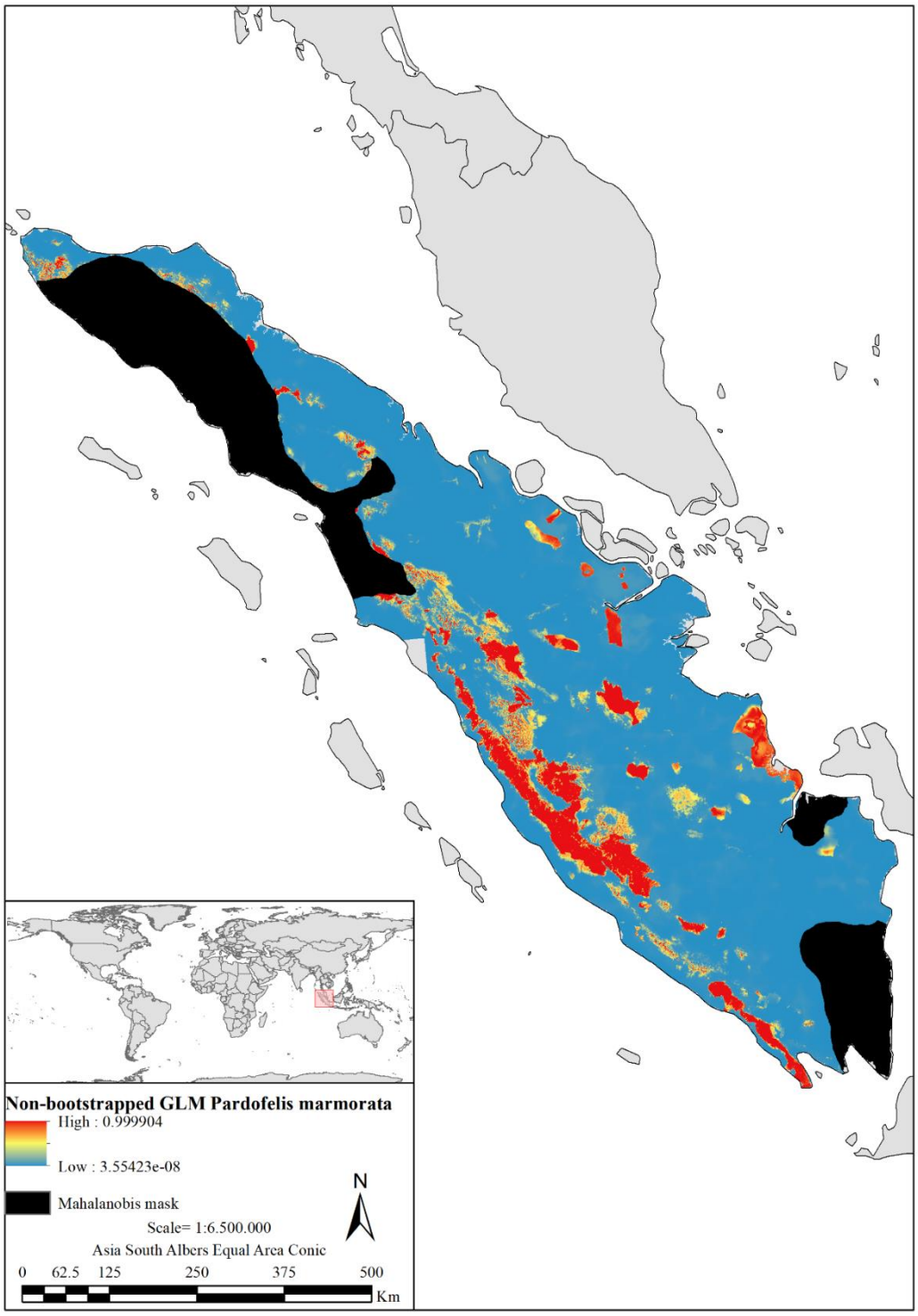



**Figure S46.** Probability surface for marbled cat (*Pardofelis marmorata*) in Sumatra, produced with non-bootstrapped GLM.

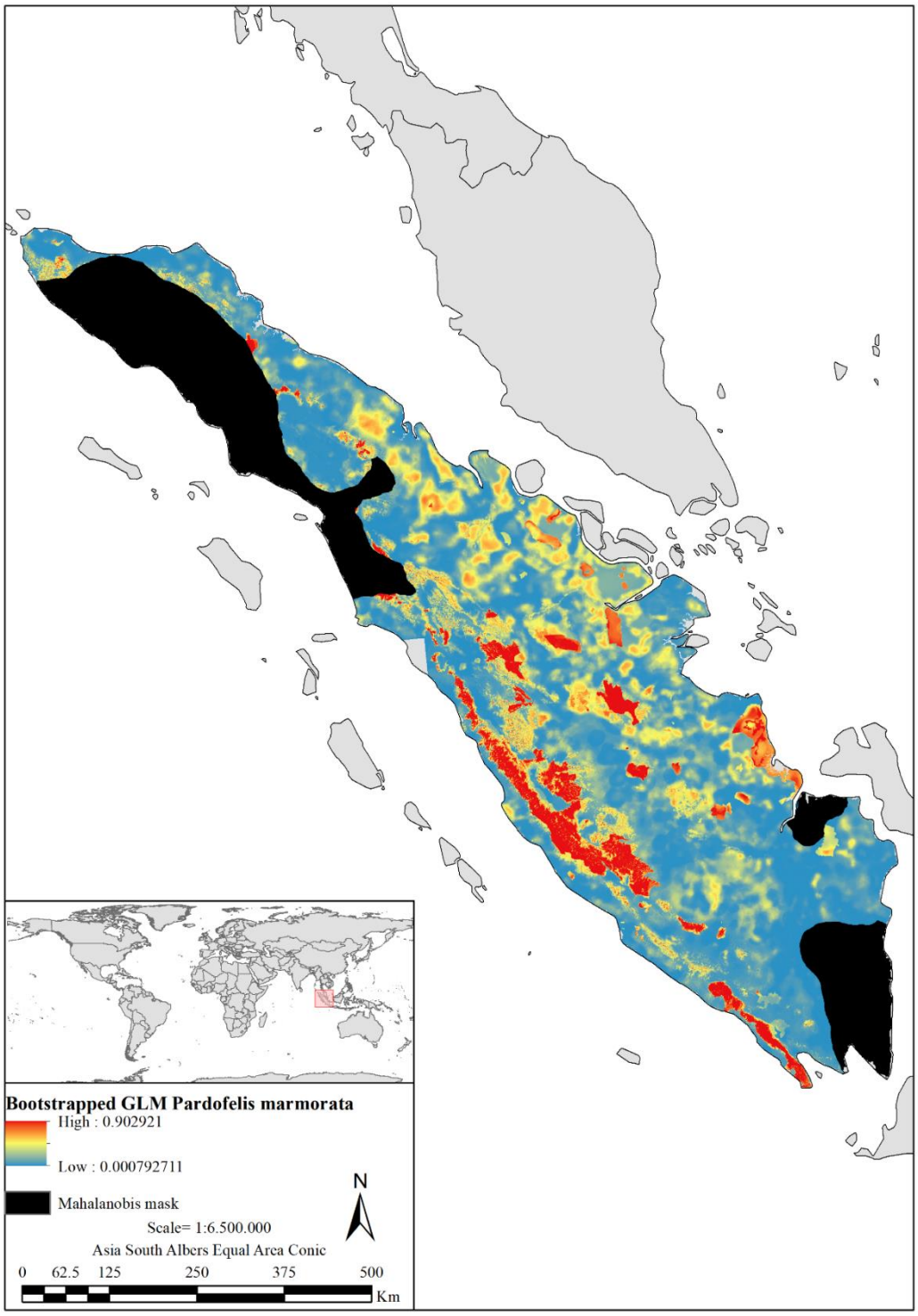



**Figure S47.** Probability surface for marbled cat (*Pardofelis marmorata*) in Sumatra, produced with bootstrapped GLM.

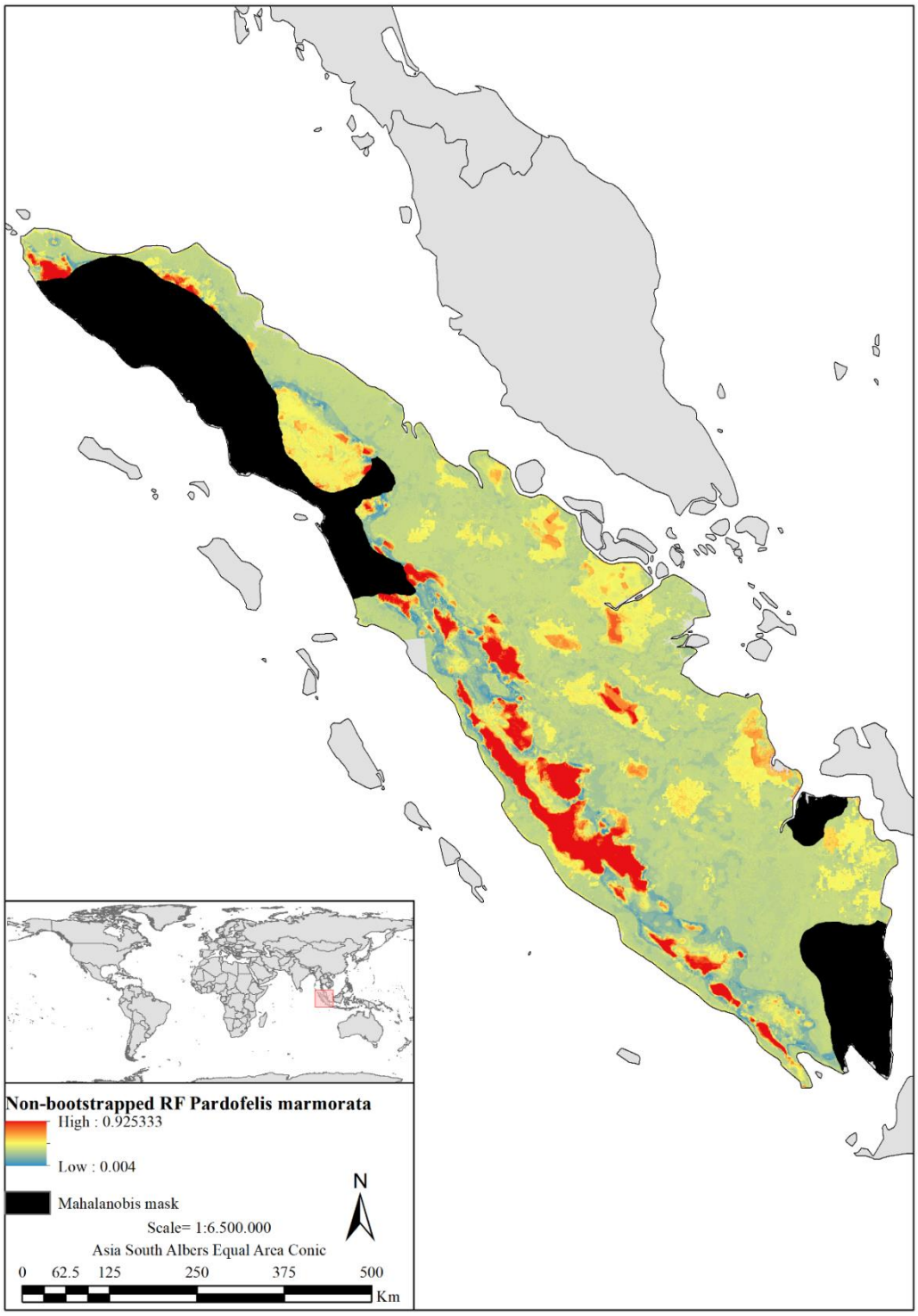



**Figure S48.** Probability surface for marbled cat (*Pardofelis marmorata*) in Sumatra, produced with non-bootstrapped RF.

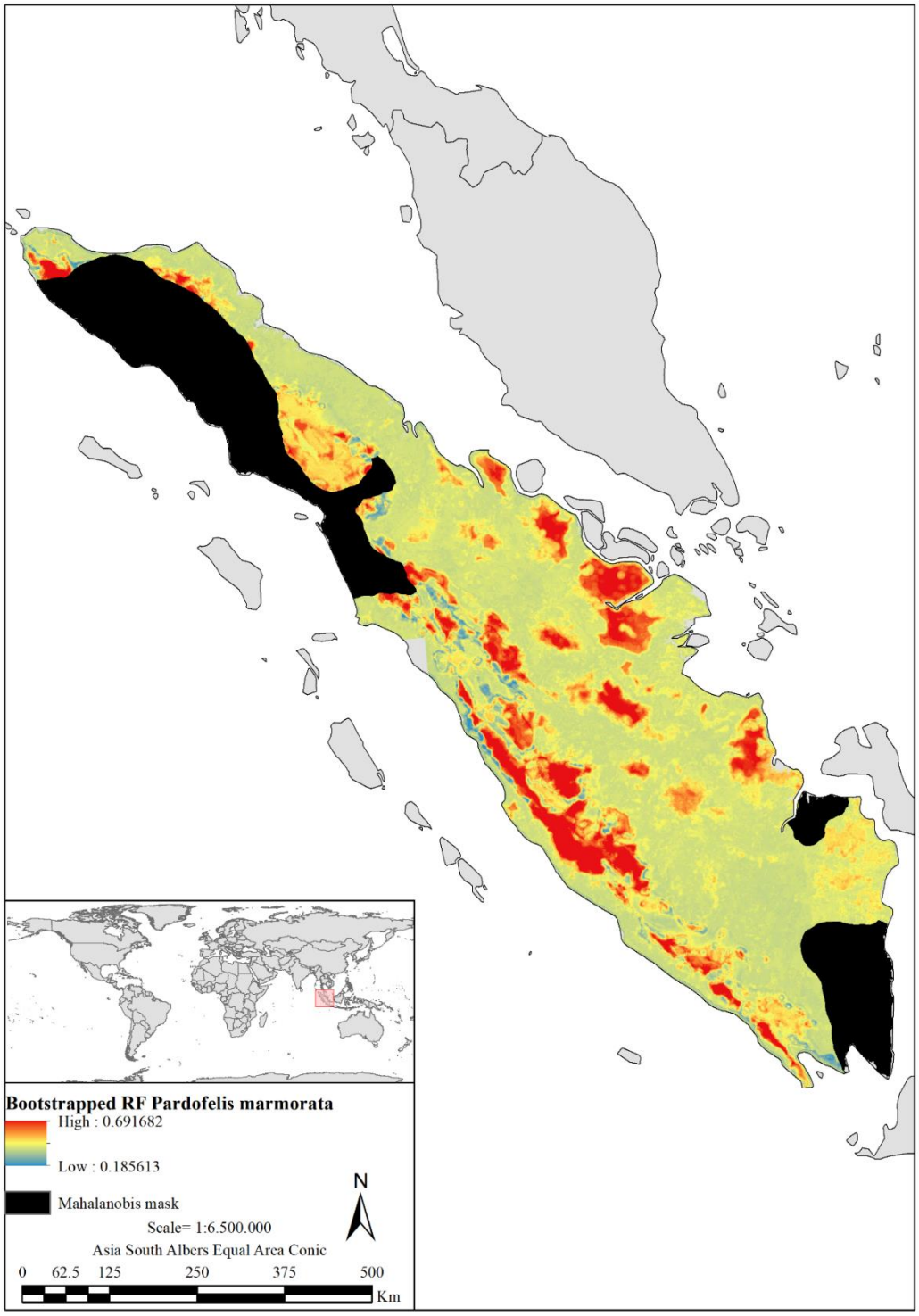



**Figure S49.** Probability surface for marbled cat (*Pardofelis marmorata*) in Sumatra, produced with bootstrapped RF.

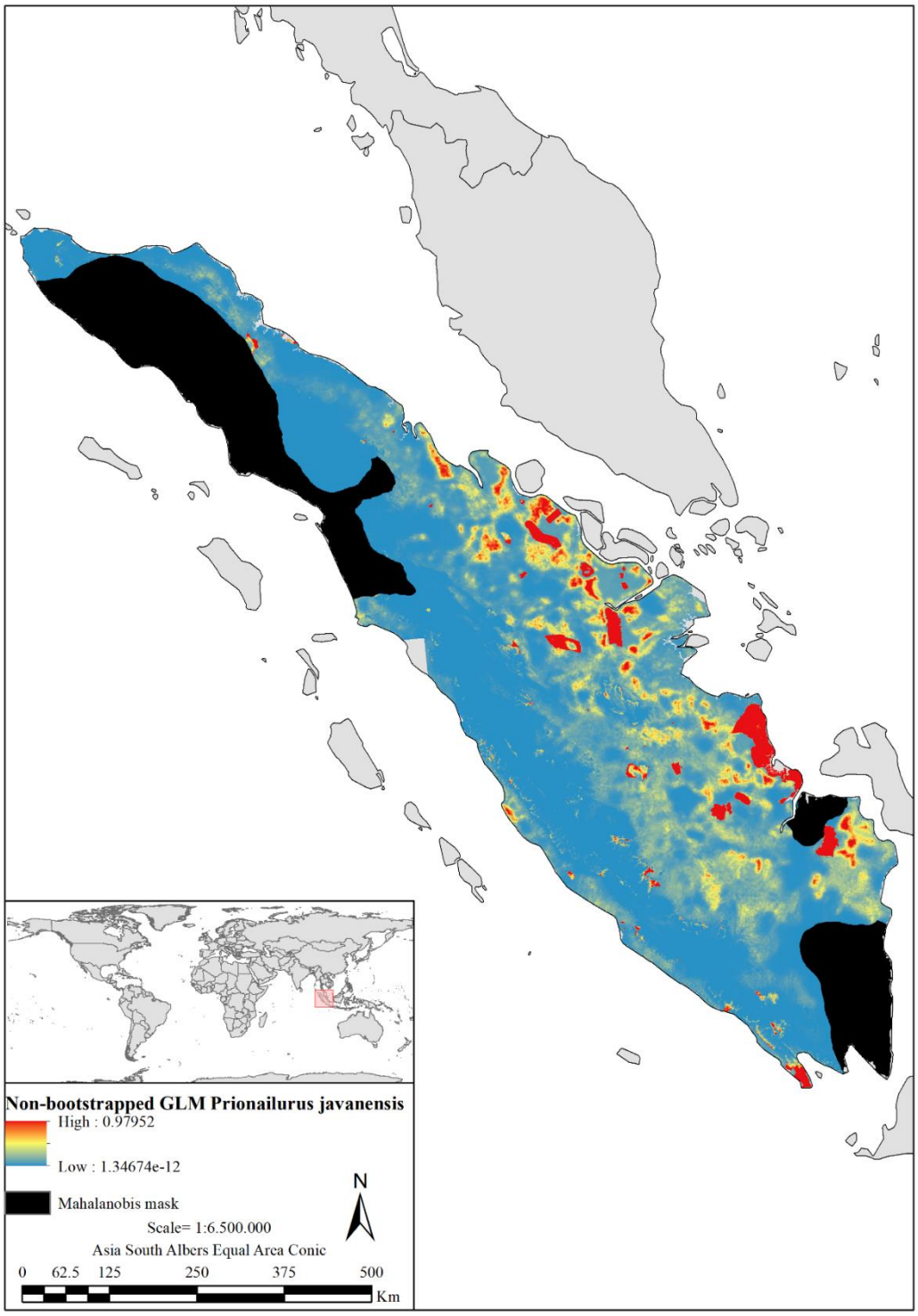



**Figure S50.** Probability surface for Sunda leopard cat (*Prionailurus javanensis*) in Sumatra, produced with non-bootstrapped GLM.

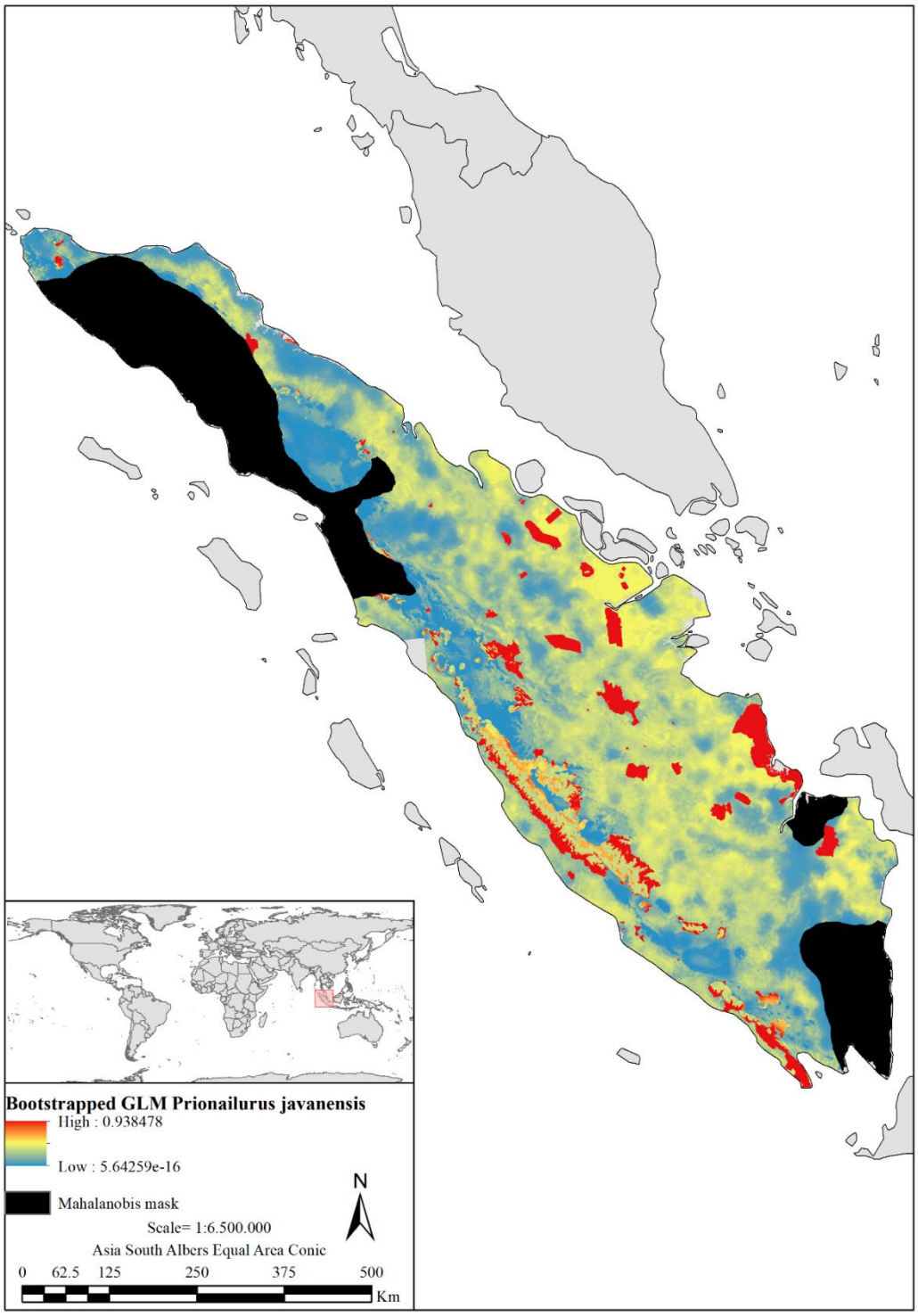



**Figure S51.** Probability surface for Sunda leopard cat (*Prionailurus javanensis*) in Sumatra, produced with bootstrapped GLM.

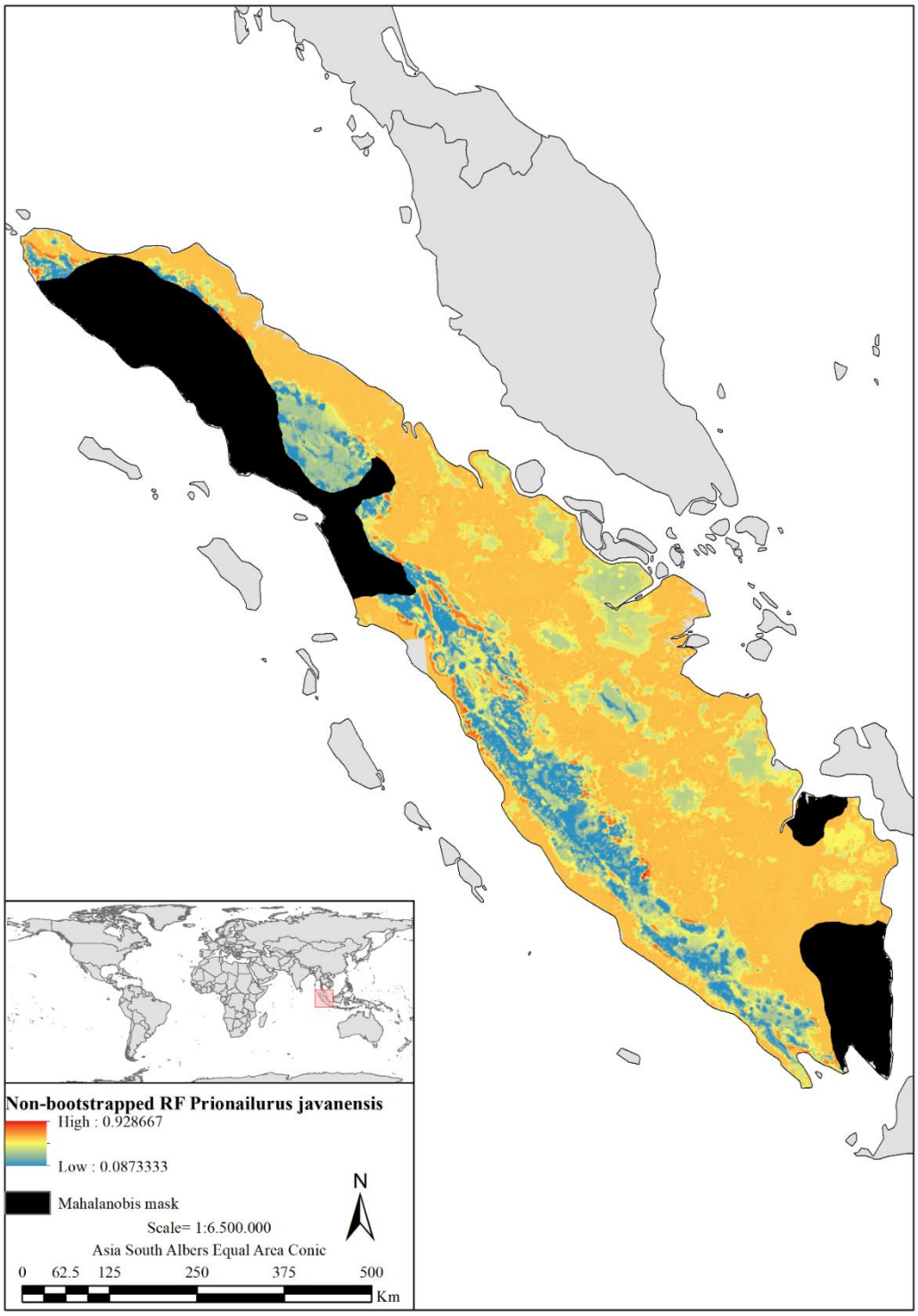



**Figure S52.** Probability surface for Sunda leopard cat (*Prionailurus javanensis*) in Sumatra, produced with non-bootstrapped RF.

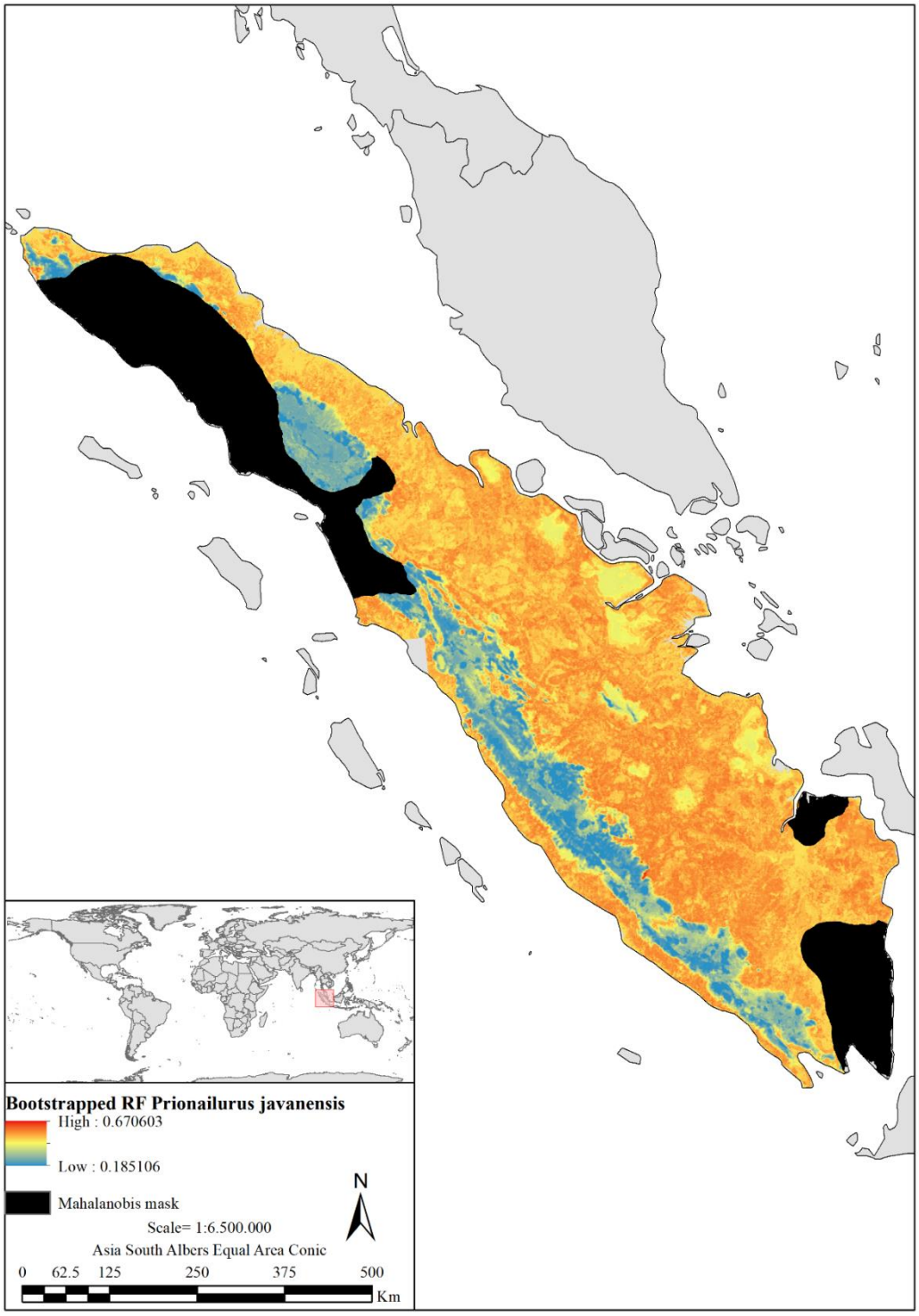



**Figure S53.** Probability surface for Sunda leopard cat (*Prionailurus javanensis*) in Sumatra, produced with bootstrapped RF.

**Table S1.** Preliminary set of covariates and corresponding representative scales obtained by applying univariate GLM and univariate RF for felids sampled in mainland Southeast Asia.

| | | | | GLM best scales (m) | | | | RF best scales (m) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Class | Covariate | Source | Metric | *Neofelis nebulosa* | *Catopuma temminckii* | *Pardofelis marmorata* | *Prionailurus bengalensis* | *Neofelis nebulosa* | *Catopuma temminckii* | *Pardofelis marmorata* | *Prionailurus bengalensis* |
| Environmental | Bare ground | ESA Land Cover CCI (2017) | PLAND | 32,000 | 32,000 | 32,000 | 32,000 | 32,000 | 32,000 | 16,000 | 32,000 |
| | Croplands | ESA Land | PLAND | 16,000 | 16,000 | 16,000 | 32,000 | 4,000 | 8,000 | 4,000 | 4,000 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Cover CCI (2017) | | | | | | | | | |
| Mosaic areas | ESA Land Cover CCI (2017) | PLAND | 8,000 | 8,000 | 250 | 32,000 | 250 | 250 | 250 | 250 |
| Forested areas | ESA Land Cover CCI (2017) | PLAND | 32,000 | 32,000 | 32,000 | 32,000 | 250 | 250 | 250 | 250 |
| Shrublands and | ESA Land Cover | PLAND | 16,000 | 32,000 | 1,000 | 2,000 | 250 | 250 | 250 | 250 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | grasslands | CCI (2017) | | | | | | | | | |
| | Urban areas | ESA Land Cover CCI (2017) | PLAND | 32,000 | 32,000 | 32,000 | 16,000 | 16,000 | 16,000 | 16,000 | 16,000 |
| | Water | ESA Land Cover CCI (2017) | PLAND | 16,000 | 8,000 | 16,000 | 16,000 | 8,000 | 8,000 | 8,000 | 8,000 |
| | Non-forested areas | Hansen et al. (2013) | PLAND | 32,000 | 32,000 | 32,000 | 32,000 | 1,000 | 8,000 | 1,000 | 1,000 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Open forests | Hansen et al. (2013) | PLAND | 16,000 | 32,000 | 32,000 | 32,000 | 500 | 4,000 | 500 | 500 |
| | Closed forests | Hansen et al. (2013) | PLAND | 16,000 | 32,000 | 32,000 | 32,000 | 500 | 500 | 500 | 250 |
| Anthropogenic | Forest loss | Hansen et al. (2013) | FM | 500 | 500 | 16,000 | 8,000 | 500 | 500 | 250 | 250 |
| | Human footprint | WCS and CIESIN (2005) | FM | 500 | 500 | 1,000 | 32,000 | 250 | 250 | 250 | 250 |
| | Human population density | CIESIN and CIAT (2016) | FM | 500 | 500 | 500 | 2,000 | 250 | 250 | 250 | 250 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Protected areas | UNEP-WCMC and IUCN (2022) | PLAND | 32,000 | 500 | 250 | 8,000 | 250 | 250 | 250 | 250 |
| Topographic | Elevation | Jarvis et al. (2008) | FM | 16,000 | 32,000 | 32,000 | 32,000 | 500 | 1,000 | 2,000 | 32,000 |
| | Compound topographic index | Jarvis et al. (2008); Evans et al. (2014) | FM | 500 | 32,000 | 250 | 32,000 | 16,000 | 2,000 | 500 | 16,000 |
| | Slope position | Jarvis et al. (2008); | FM | 8,000 | 8,000 | 250 | 4,000 | 1,000 | 1,000 | 250 | 16,000 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Climatic | Mean annual precipitation | Evans et al. (2014) Fick and Hijmans (2017) | FM | 1,000 | 32,000 | 32,000 | 8,000 | 32,000 | 1,000 | 250 | 2,000 |

**Table S2.** Preliminary set of covariates and corresponding representative scales obtained by applying univariate GLM and univariate RF for felids sampled in Borneo.

| | | | | GLM best scales (m) | | | | RF best scales (m) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Class | Covariate | Source | Metric | *Neofelis diardi* | *Catopuma badia* | *Pardofelis marmorata* | *Prionailurus javanensis* | *Neofelis diardi* | *Catopuma badia* | *Pardofelis marmorata* | *Prionailurus javanensis* |
| Environmental | Lowland forests | Miettinen et al. (2012) | PLAND | 8,000 | 4,000 | 8,000 | 8,000 | 250 | 500 | 500 | 250 |
| | Lowland mosaic areas | Miettinen et al. (2012) | PLAND | 8,000 | 8,000 | 8,000 | 8,000 | 250 | 500 | 500 | 250 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Lower-montane forests | Miettinen et al. (2012) | PLAND | 1,000 | 1,000 | 6,000 | 4,000 | 8,000 | 1,000 | 2,000 | 8,000 |
| | Lowland open areas | Miettinen et al. (2012) | PLAND | 10,000 | 4,000 | 10,000 | 10,000 | 500 | 2,000 | 500 | 4,000 |
| | Mangroves | Miettinen et al. (2012) | PLAND | 8,000 | 8,000 | NA⁂ | 6,000 | 4,000 | 4,000 | NA⁂ | 10,000 |
| | Montane mosaic areas | Miettinen et al. (2012) | PLAND | NA* | NA* | NA* | NA* | NA* | NA* | NA* | NA* |
| | Montane open areas | Miettinen et al. (2012) | PLAND | NA* | NA* | NA* | NA* | NA* | NA* | NA* | NA* |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Palm plantations | Miettinen et al. (2012) | PLAND | 2,000 | 4,000 | 2,000 | 2,000 | 2,000 | 2,000 | 2,000 | 2,000 |
| | Plantation and regrowth | Miettinen et al. (2012) | PLAND | 6,000 | 250 | 4,000 | 8,000 | 500 | 500 | 500 | 250 |
| | Upper-montane forests | Miettinen et al. (2012) | PLAND | NA* | NA* | NA* | NA* | NA* | NA* | NA* | NA* |
| | Water | Miettinen et al. (2012) | PLAND | 10,000 | NA‡ | 4,000 | 10,000 | 4,000 | NA‡ | 4,000 | 4,000 |
| | Continuous tree canopy cover | Hansen et al. (2013) | FM | 6,000 | 2,000 | 6,000 | 1,000 | 6,000 | 250 | 4,000 | 1,000 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Non-forested areas | Hansen et al. (2013) | PLAND | 6,000 | 4,000 | 8,000 | 1,000 | 500 | 1,000 | 500 | 1,000 |
| | Open forests | Hansen et al. (2013) | PLAND | 8,000 | 4,000 | 8,000 | 10,000 | 1,000 | 4,000 | 4,000 | 10,000 |
| | Closed forests | Hansen et al. (2013) | PLAND | 6,000 | 4,000 | 8,000 | 1,000 | 1,000 | 1,000 | 500 | 1,000 |
| Anthropogenic | Human footprint | WCS and CIESIN (2005) | FM | 250 | 10,000 | 10,000 | 500 | 250 | 250 | 250 | 250 |
| | Human population density | CIESIN and CIAT (2016) | FM | 6,000 | 250 | 10,000 | 250 | 500 | 4,000 | 2,000 | 4,000 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Forest loss | Hansen et al. (2013) | PLAND | 8,000 | 4,000 | 10,000 | 8,000 | 500 | 2,000 | 500 | 500 |
| | Protected areas | UNEP-WCMC and IUCN (2022) | PLAND | 8,000 | 4,000 | 8,000 | 10,000 | 250 | 1,000 | 500 | 1,000 |
| Topographic | Elevation | Jarvis et al. (2008) | FM | 250 | 250 | 250 | 10,000 | 10,000 | 8,000 | 8,000 | 10,000 |
| | Roughness | Jarvis et al. (2008); Evans et al. (2014) | FM | 1,000 | 2,000 | 1,000 | 10,000 | 8,000 | 4,000 | 8,000 | 6,000 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Slope position | Jarvis et al. (2008); Evans et al. (2014) | FM | 250 | 8,000 | 250 | 250 | 250 | 6,000 | 250 | 10,000 |
| | Compound topographic index | Jarvis et al. (2008); Evans et al. (2014) | FM | 500 | 2,000 | 500 | 10,000 | 1,000 | 2,000 | 1,000 | 6,000 |

* Covariates excluded from the models because occurring in <10% of camera trap stations.

⁑ Covariates excluded because the univariate models did not converge.

**Table S3.** Preliminary set of covariates and corresponding representative scales obtained by applying univariate GLM and univariate RF for felids sampled in Sumatra.

| | | | | GLM best scales (m) | | | | RF best scales (m) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Class | Covariate | Source | Metric | *Neofelis diardi* | *Catopuma temminckii* | *Pardofelis marmorata* | *Prionailurus javanensis* | *Neofelis diardi* | *Catopuma temminckii* | *Pardofelis marmorata* | *Prionailurus javanensis* |
| Environmental | Lowland forests | Miettinen et al. (2012) | PLAND | 4,000 | 6,000 | 500 | 6,000 | 500 | 500 | 500 | 500 |
| | Lowland mosaic areas | Miettinen et al. (2012) | PLAND | 8,000 | 10,000 | 8,000 | 4,000 | 2,000 | 2,000 | 4,000 | 2,000 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Lower-montane forests | Miettinen et al. (2012) | PLAND | 250 | 6,000 | 2,000 | 500 | 250 | 250 | 500 | 500 |
| | Lowland open areas | Miettinen et al. (2012) | PLAND | 6,000 | 10,000 | 6,000 | 4,000 | 4,000 | 4,000 | 6,000 | 6,000 |
| | Mangroves | Miettinen et al. (2012) | PLAND | NA* | NA* | NA* | NA* | NA* | NA* | NA* | NA* |
| | Montane mosaic areas | Miettinen et al. (2012) | PLAND | 2,000 | 10,000 | 6,000 | 10,000 | 2,000 | 2,000 | 4,000 | 6,000 |
| | Montane open areas | Miettinen et al. (2012) | PLAND | 8,000 | 8,000 | 8,000 | 10,000 | 6,000 | 10,000 | 6,000 | 8,000 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Palm plantations | Miettinen et al. (2012) | PLAND | 8,000 | 10,000 | 8,000 | NA‡‡ | 8,000 | 8,000 | 8,000 | NA‡‡ |
| | Plantation and regrowth | Miettinen et al. (2012) | PLAND | 1,000 | 2,000 | 2,000 | 2,000 | 250 | 500 | 500 | 500 |
| | Upper-montane forests | Miettinen et al. (2012) | PLAND | 6,000 | 10,000 | 10,000 | 10,000 | 4,000 | 4,000 | 4,000 | 4,000 |
| | Water | Miettinen et al. (2012) | PLAND | NA* | NA* | NA* | NA* | NA* | NA* | NA* | NA* |
| | Continuous tree canopy cover | Hansen et al. (2013) | FM | 500 | 1,000 | 6,000 | 2,000 | 500 | 2,000 | 6,000 | 6,000 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Non-forested areas | Hansen et al. (2013) | PLAND | 2,000 | 4,000 | 6,000 | 10,000 | 2,000 | 500 | 2,000 | 2,000 |
| | Open forests | Hansen et al. (2013) | PLAND | 6,000 | 8,000 | 8,000 | 10,000 | 8,000 | 2,000 | 4,000 | 4,000 |
| | Closed forests | Hansen et al. (2013) | PLAND | 6,000 | 4,000 | 6,000 | 10,000 | 1,000 | 1,000 | 2,000 | 2,000 |
| Anthropogenic | Human footprint | WCS and CIESIN (2005) | FM | 10,000 | 10,000 | 6,000 | 10,000 | 250 | 250 | 250 | 500 |
| | Human population density | CIESIN and CIAT (2016) | FM | 500 | 1,000 | 10,000 | 10,000 | 250 | 250 | 250 | 250 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Forest loss | Hansen et al. (2013) | PLAND | 6,000 | 4,000 | 4,000 | 4,000 | 500 | 1,000 | 1,000 | 1,000 |
| | Protected areas | UNEP-WCMC and IUCN (2022) | PLAND | 4,000 | 10,000 | 250 | 250 | 250 | 500 | 250 | 250 |
| Topographic | Elevation | Jarvis et al. (2008) | FM | 250 | 8,000 | 4,000 | 250 | 4,000 | 4,000 | 8,000 | 2,000 |
| | Roughness | Jarvis et al. (2008); Evans et al. (2014) | FM | 4,000 | 10,000 | 10,000 | 10,000 | 250 | 4,000 | 6,000 | 10,000 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Slope position | Jarvis et al. (2008); Evans et al. (2014) | FM | 8,000 | 250 | 1,000 | 2,000 | 250 | 250 | 1,000 | 4,000 |
| | Compound topographic index | Jarvis et al. (2008); Evans et al. (2014) | FM | 500 | 250 | 1,000 | 250 | 250 | 8,000 | 6,000 | 2,000 |

* Covariates excluded from the models because occurring in <10% of camera trap stations.

⁑ Covariates excluded because the univariate models did not converge.

**Table S4.** Pearson's correlations between the four models for clouded leopard (*Neofelis nebulosa*) in mainland Southeast Asia.

| *Neofelis nebulosa* | Non-bootstrapped GLM | Bootstrapped GLM | Non-bootstrapped RF | Bootstrapped RF |
|---|---|---|---|---|
| Non-bootstrapped GLM | 1.00 | 0.97 | 0.14 | 0.27 |
| Bootstrapped GLM | 0.97 | 1.00 | 0.10 | 0.23 |
| Non-bootstrapped RF | 0.14 | 0.10 | 1.00 | 0.81 |
| Bootstrapped RF | 0.27 | 0.23 | 0.81 | 1.00 |

**Table S5.** Pearson's correlations between the four models for Asiatic golden cat (*Catopuma temminckii*) in mainland Southeast Asia.

| *Catopuma temminckii* | Non-bootstrapped GLM | Bootstrapped GLM | Non-bootstrapped RF | Bootstrapped RF |
|---|---|---|---|---|
| Non-bootstrapped GLM | 1.00 | 0.49 | -0.13 | 0.58 |
| Bootstrapped GLM | 0.49 | 1.00 | -0.23 | 0.40 |
| Non-bootstrapped RF | -0.13 | -0.23 | 1.00 | 0.13 |
| Bootstrapped RF | 0.58 | 0.40 | 0.13 | 1.00 |

**Table S6.** Pearson's correlations between the four models for marbled cat (*Pardofelis marmorata*) in mainland Southeast Asia.

| *Pardofelis marmorata* | Non-bootstrapped GLM | Bootstrapped GLM | Non-bootstrapped RF | Bootstrapped RF |
|---|---|---|---|---|
| Non-bootstrapped GLM | 1.00 | 0.89 | 0.51 | 0.47 |
| Bootstrapped GLM | 0.89 | 1.00 | 0.36 | 0.31 |
| Non-bootstrapped RF | 0.51 | 0.36 | 1.00 | 0.75 |
| Bootstrapped RF | 0.47 | 0.31 | 0.75 | 1.00 |

**Table S7.** Pearson's correlations between the four models for leopard cat (*Prionailurus bengalensis*) in mainland Southeast Asia.

| *Prionailurus bengalensis* | Non-bootstrapped GLM | Bootstrapped GLM | Non-bootstrapped RF | Bootstrapped RF |
|---|---|---|---|---|
| Non-bootstrapped GLM | 1.00 | 0.93 | 0.30 | 0.31 |
| Bootstrapped GLM | 0.93 | 1.00 | 0.40 | 0.42 |
| Non-bootstrapped RF | 0.30 | 0.40 | 1.00 | 0.71 |
| Bootstrapped RF | 0.31 | 0.42 | 0.71 | 1.00 |

**Table S8.** Pearson's correlations between the four models for Sunda clouded leopard (*Neofelis diardi*) in Borneo.

| *Neofelis diardi* | Non-bootstrapped GLM | Bootstrapped GLM | Non-bootstrapped RF | Bootstrapped RF |
|---|---|---|---|---|
| Non-bootstrapped GLM | 1.00 | 0.86 | 0.64 | 0.66 |
| Bootstrapped GLM | 0.86 | 1.00 | 0.61 | 0.65 |
| Non-bootstrapped RF | 0.64 | 0.61 | 1.00 | 0.90 |
| Bootstrapped RF | 0.66 | 0.65 | 0.90 | 1.00 |

**Table S9.** Pearson's correlations between the four models for Borneo bay cat (*Catopuma badia*) in Borneo.

| *Catopuma badia* | Non-bootstrapped GLM | Bootstrapped GLM | Non-bootstrapped RF | Bootstrapped RF |
|---|---|---|---|---|
| Non-bootstrapped GLM | 1.00 | 0.57 | 0.09 | 0.12 |
| Bootstrapped GLM | 0.57 | 1.00 | 0.17 | 0.27 |
| Non-bootstrapped RF | 0.09 | 0.17 | 1.00 | 0.88 |
| Bootstrapped RF | 0.12 | 0.27 | 0.88 | 1.00 |

**Table S10.** Pearson's correlations between the four models for marbled cat (*Pardofelis marmorata*) in Borneo.

| *Pardofelis marmorata* | Non-bootstrapped GLM | Bootstrapped GLM | Non-bootstrapped RF | Bootstrapped RF |
|---|---|---|---|---|
| Non-bootstrapped GLM | 1.00 | 0.71 | 0.46 | 0.50 |
| Bootstrapped GLM | 0.71 | 1.00 | 0.60 | 0.74 |
| Non-bootstrapped RF | 0.46 | 0.60 | 1.00 | 0.87 |
| Bootstrapped RF | 0.50 | 0.74 | 0.87 | 1.00 |

**Table S11.** Pearson's correlations between the four models for Sunda leopard cat (*Prionailurus javanensis*) in Borneo.

| *Prionailurus javanensis* | Non-bootstrapped GLM | Bootstrapped GLM | Non-bootstrapped RF | Bootstrapped RF |
|---|---|---|---|---|
| Non-bootstrapped GLM | 1.00 | 0.97 | 0.60 | 0.61 |
| Bootstrapped GLM | 0.97 | 1.00 | 0.62 | 0.63 |
| Non-bootstrapped RF | 0.60 | 0.62 | 1.00 | 0.84 |
| Bootstrapped RF | 0.61 | 0.63 | 0.84 | 1.00 |

**Table S12.** Pearson's correlations between the four models for Sunda clouded leopard (*Neofelis diardi*) in Sumatra.

| *Neofelis diardi* | Non-bootstrapped GLM | Bootstrapped GLM | Non-bootstrapped RF | Bootstrapped RF |
|---|---|---|---|---|
| Non-bootstrapped GLM | 1.00 | 0.87 | 0.53 | 0.71 |
| Bootstrapped GLM | 0.87 | 1.00 | 0.58 | 0.84 |
| Non-bootstrapped RF | 0.53 | 0.58 | 1.00 | 0.81 |
| Bootstrapped RF | 0.71 | 0.84 | 0.81 | 1.00 |

**Table S13.** Pearson's correlations between the four models for Asiatic golden cat (*Catopuma temminckii*) in Sumatra.

| *Catopuma temminckii* | Non-bootstrapped GLM | Bootstrapped GLM | Non-bootstrapped RF | Bootstrapped RF |
|---|---|---|---|---|
| Non-bootstrapped GLM | 1.00 | 0.90 | 0.61 | 0.75 |
| Bootstrapped GLM | 0.90 | 1.00 | 0.64 | 0.71 |
| Non-bootstrapped RF | 0.61 | 0.64 | 1.00 | 0.86 |
| Bootstrapped RF | 0.75 | 0.71 | 0.86 | 1.00 |

**Table S14.** Pearson's correlations between the four models for marbled cat (*Pardofelis marmorata*) in Sumatra.

| *Pardofelis marmorata* | Non-bootstrapped GLM | Bootstrapped GLM | Non-bootstrapped RF | Bootstrapped RF |
|---|---|---|---|---|
| Non-bootstrapped GLM | 1.00 | 0.87 | 0.72 | 0.65 |
| Bootstrapped GLM | 0.87 | 1.00 | 0.64 | 0.64 |
| Non-bootstrapped RF | 0.72 | 0.64 | 1.00 | 0.83 |
| Bootstrapped RF | 0.65 | 0.64 | 0.83 | 1.00 |

**Table S15.** Pearson's correlations between the four models for Sunda leopard cat (*Prionailurus javanensis*) in Sumatra.

| *Prionailurus javanensis* | Non-bootstrapped GLM | Bootstrapped GLM | Non-bootstrapped RF | Bootstrapped RF |
|---|---|---|---|---|
| Non-bootstrapped GLM | 1.00 | 0.69 | 0.10 | 0.23 |
| Bootstrapped GLM | 0.69 | 1.00 | -0.02 | 0.11 |
| Non-bootstrapped RF | 0.10 | -0.02 | 1.00 | 0.89 |
| Bootstrapped RF | 0.23 | 0.11 | 0.89 | 1.00 |

**Table S16.** Binary matrix used for the Mantel test to investigate the effect of algorithm on the similarities between species distribution probability surfaces.

| Algorithm | Non-bootstrapped GLM | Bootstrapped GLM | Non-bootstrapped RF | Bootstrapped RF |
|---|---|---|---|---|
| Non-bootstrapped GLM | 1 | 1 | 0 | 0 |
| Bootstrapped GLM | 1 | 1 | 0 | 0 |
| Non-bootstrapped RF | 0 | 0 | 1 | 1 |
| Bootstrapped RF | 0 | 0 | 1 | 1 |

**Table S17.** Binary matrix used for the Mantel test to investigate the effect of bootstrapping on the similarities between species distribution probability surfaces.

| Bootstrapping | Non-bootstrapped GLM | Bootstrapped GLM | Non-bootstrapped RF | Bootstrapped RF |
|---|---|---|---|---|
| Non-bootstrapped GLM | 1 | 0 | 1 | 0 |
| Bootstrapped GLM | 0 | 1 | 0 | 1 |
| Non-bootstrapped RF | 1 | 0 | 1 | 0 |
| Bootstrapped RF | 0 | 1 | 0 | 1 |

**Table S18.** Sampling effort in mainland Southeast Asia. Shown are sampling locations per country, number of camera trap stations, total number of trap nights and average number of trap nights per camera trap station. Modified from Macdonald et al. (2020).

| Countries | Sampling locations | Camera trap stations | Total trap nights (mean) |
|---|---|---|---|
| Cambodia | Cardamom | 76 | 7,740 (101.84) |
| | PPWS | 77 | 5,283 (68.61) |
| India | Dampa | 74 | 4,962 (67.05) |
| Laos | NPNV | 78 | 3,729 (47.81) |
| | PD | 78 | 3,666 (47.00) |
| | PS | 72 | 3,446 (47.86) |
| Myanmar | Htamanthi | 162 | 14,546 (89.79) |
| Nepal | Langtang National Park | 83 | 7,105 (85.60) |
| Peninsular Malaysia | Ulu Muda Grid 1 | 76 | 9,251 (121.72) |
| | Ulu Muda Grid 2 | 54 | 5,995 (111.02) |
| Thailand | Dong Yai Wildlife Sanctuary | 58 | 2,709 (46.71) |
| | Khao Yai National Park | 67 | 6,266 (93.52) |
| | Pang Sida National Park | 163 | 17,750 (108.90) |

| | | | |
|---|---|---|---|
| | Ta Phraya National Park | 56 | 4,636 (82.79) |
| | Thap Lan National Park | 210 | 18,305 (87.17) |
| Total | | 1,384 | 115,389 (83.37) |

**Table S19.** Sampling effort in Sunda Islands. Shown are sampling locations per region, number of camera trap stations, total number of trap nights and average number of trap nights per camera trap station. Modified from Chiaverini et al. (2022).

| Regions | Sampling locations | Camera trap stations | Total trap nights (mean) |
|---|---|---|---|
| Kalimantan (Borneo) | Bawan | 63 | 4,070 (64.60) |
| | Belantikan | 51 | 3,846 (75.41) |
| | Kutai | 52 | 3,269 (62.87) |
| | Lesan | 73 | 10,057 (137.77) |
| | Sungai Wain | 78 | 4,729 (60.63) |
| Sabah (Borneo) | Crocker | 35 | 4,059 (115.97) |
| | Danum Palm | 23 | 2,214 (96.26) |
| | Danum Valley | 79 | 5,880 (74.43) |
| | IJM | 33 | 1,855 (56.21) |
| | Kinabatangan | 68 | 4,450 (65.44) |
| | Malua | 38 | 3,867 (101.76) |
| | Sepilok | 35 | 2,067 (59.06) |
| | Tabin | 74 | 6,877 (92.93) |
| | Tawau | 77 | 17,600 (228.57) |
| | Ulu Segama | 22 | 2,847 (129.41) |
| Sumatra | Bungo | 79 | 8,405 (106.39) |
| | Ipuh | 79 | 6,276 (79.44) |

| | | | |
|---|---|---|---|
| | Kambang | 149 | 12,599 (84.56) |
| | Linggau | 120 | 7,269 (60.58) |
| | Muara Hemat | 164 | 13,466 (82.11) |
| | RKE | 72 | 5,586 (77.58) |
| | Sipurak | 80 | 7,227 (90.34) |
| Total | | 1,544 | 138,515 (89.71) |

**Table S20.** Presences of felids sampled in mainland Southeast Asia.

| Countries | *Neofelis nebulosa* | *Catopuma temminckii* | *Pardofelis marmorata* | *Prionailurus bengalensis* |
|---|---|---|---|---|
| Cambodia | 11 | 11 | 18 | 60 |
| India | 28 | 22 | 14 | 25 |
| Laos | 39 | 45 | 28 | 32 |
| Myanmar | 67 | 56 | 31 | 45 |
| Nepal | 7 | 0 | 0 | 37 |
| Peninsular Malaysia | 21 | 2 | 15 | 60 |
| Thailand | 62 | 12 | 10 | 86 |
| Total | 235 | 148 | 116 | 345 |

**Table S21.** Presences of felids sampled in Borneo.

| Regions | *Neofelis diardi* | *Catopuma badia* | *Pardofelis marmorata* | *Prionailurus javanensis* |
|---|---|---|---|---|
| Kalimantan | 13 | 3 | 3 | 5 |
| Sabah | 141 | 22 | 82 | 197 |
| Total | 154 | 25 | 85 | 202 |

**Table S22.** Presences of felids sampled in Sumatra.

| Region | *Neofelis diardi* | *Catopuma temminckii* | *Pardofelis marmorata* | *Prionailurus javanensis* |
|---|---|---|---|---|
| Sumatra | 101 | 102 | 39 | 23 |

**Table S23.** Validation metrics of the models for felids sampled in mainland Southeast Asia, including Area Under the ROC Curve (AUC), Percent Correctly Classified (PCC), Sensitivity, Specificity and True Skill Statistic (TSS).

| Mainland | AUC | PCC | Sensitivity | Specificity | TSS |
|---|---|---|---|---|---|
| *Neofelis nebulosa* – non-bootstrapped GLM | 0.56 | 0.83 | 0.10 | 0.98 | 0.09 |
| *Neofelis nebulosa* – bootstrapped GLM | 0.59 | 0.70 | 0.38 | 0.76 | 0.14 |
| *Neofelis nebulosa* – non-bootstrapped RF | 0.55 | 0.70 | 0.38 | 0.77 | 0.15 |
| *Neofelis nebulosa* – bootstrapped RF | 0.53 | 0.63 | 0.40 | 0.68 | 0.08 |

| | | | | | |
|---|---|---|---|---|---|
| *Catopuma temminckii* – non-bootstrapped GLM | 0.79 | 0.69 | 0.92 | 0.66 | 0.58 |
| *Catopuma temminckii* – bootstrapped GLM | 0.63 | 0.30 | 0.91 | 0.23 | 0.14 |
| *Catopuma temminckii* – non-bootstrapped RF | 0.68 | 0.81 | 0.44 | 0.85 | 0.29 |
| *Catopuma temminckii* – bootstrapped RF | 0.67 | 0.73 | 0.50 | 0.76 | 0.25 |
| *Pardofelis marmorata* – non-bootstrapped GLM | 0.77 | 0.92 | 0.15 | 0.98 | 0.13 |
| *Pardofelis marmorata* – | 0.73 | 0.66 | 0.72 | 0.65 | 0.37 |

| | | | | | |
|---|---|---|---|---|---|
| bootstrapped GLM | | | | | |
| *Pardofelis marmorata* – non-bootstrapped RF | 0.59 | 0.92 | 0.15 | 0.98 | 0.13 |
| *Pardofelis marmorata* – bootstrapped RF | 0.62 | 0.79 | 0.31 | 0.83 | 0.15 |
| *Prionailurus bengalensis* – non-bootstrapped GLM | 0.65 | 0.70 | 0.54 | 0.75 | 0.29 |
| *Prionailurus bengalensis* – bootstrapped GLM | 0.64 | 0.66 | 0.53 | 0.70 | 0.23 |
| *Prionailurus bengalensis* – non- | 0.53 | 0.73 | 0.18 | 0.89 | 0.08 |

| | | | | | |
|---|---|---|---|---|---|
| bootstrapped RF | | | | | |
| *Prionailurus bengalensis* – bootstrapped RF | 0.54 | 0.57 | 0.50 | 0.59 | 0.09 |

**Table S24.** Validation metrics of the models for felids sampled in Borneo, including Area Under the ROC Curve (AUC), Percent Correctly Classified (PCC), Sensitivity, Specificity and True Skill Statistic (TSS).

| Borneo | AUC | PCC | Sensitivity | Specificity | TSS |
|---|---|---|---|---|---|
| *Neofelis diardi* – non-bootstrapped GLM | 0.76 | 0.81 | 0.68 | 0.84 | 0.52 |
| *Neofelis diardi* – bootstrapped GLM | 0.80 | 0.82 | 0.51 | 0.89 | 0.41 |
| *Neofelis diardi* – non-bootstrapped RF | 0.77 | 0.81 | 0.47 | 0.90 | 0.37 |
| *Neofelis diardi* – bootstrapped RF | 0.79 | 0.82 | 0.54 | 0.89 | 0.43 |
| *Catopuma badia* – non- | 0.87 | 0.88 | 0.80 | 0.89 | 0.69 |

| | | | | | |
|---|---|---|---|---|---|
| bootstrapped GLM | | | | | |
| *Catopuma badia* – bootstrapped GLM | 0.70 | 0.72 | 0.53 | 0.73 | 0.26 |
| *Catopuma badia* – non-bootstrapped RF | 0.77 | 0.82 | 0.50 | 0.85 | 0.35 |
| *Catopuma badia* – bootstrapped RF | 0.74 | 0.89 | 0.37 | 0.91 | 0.28 |
| *Pardofelis marmorata* – non-bootstrapped GLM | 0.80 | 0.89 | 0.38 | 0.95 | 0.33 |
| *Pardofelis marmorata* – bootstrapped GLM | 0.80 | 0.74 | 0.65 | 0.75 | 0.41 |
| *Pardofelis marmorata* – | 0.78 | 0.87 | 0.56 | 0.91 | 0.47 |

| | | | | | |
|---|---|---|---|---|---|
| non-bootstrapped RF | | | | | |
| *Pardofelis marmorata* – bootstrapped RF | 0.81 | 0.86 | 0.53 | 0.90 | 0.43 |
| *Prionailurus javanensis* – non-bootstrapped GLM | 0.76 | 0.72 | 0.93 | 0.64 | 0.57 |
| *Prionailurus javanensis* – bootstrapped GLM | 0.77 | 0.71 | 0.78 | 0.68 | 0.47 |
| *Prionailurus javanensis* – non-bootstrapped RF | 0.74 | 0.62 | 0.90 | 0.52 | 0.42 |
| *Prionailurus javanensis* – bootstrapped RF | 0.74 | 0.72 | 0.63 | 0.76 | 0.39 |

**Table S25.** Validation metrics of the models for felids sampled in Sumatra, including Area Under the ROC Curve (AUC), Percent Correctly Classified (PCC), Sensitivity, Specificity and True Skill Statistic (TSS).

| Sumatra | AUC | PCC | Sensitivity | Specificity | TSS |
|---|---|---|---|---|---|
| *Neofelis diardi* – non-bootstrapped GLM | 0.78 | 0.81 | 0.67 | 0.85 | 0.52 |
| *Neofelis diardi* – bootstrapped GLM | 0.72 | 0.75 | 0.51 | 0.79 | 0.31 |
| *Neofelis diardi* – non-bootstrapped RF | 0.82 | 0.82 | 0.47 | 0.91 | 0.37 |
| *Neofelis diardi* – bootstrapped RF | 0.75 | 0.80 | 0.48 | 0.85 | 0.33 |
| *Catopuma temminckii* – non- | 0.67 | 0.79 | 0.40 | 0.85 | 0.25 |

| | | | | | |
|---|---|---|---|---|---|
| bootstrapped GLM | | | | | |
| *Catopuma temminckii* – bootstrapped GLM | 0.62 | 0.68 | 0.44 | 0.72 | 0.16 |
| *Catopuma temminckii* – non-bootstrapped RF | 0.77 | 0.81 | 0.55 | 0.84 | 0.39 |
| *Catopuma temminckii* – bootstrapped RF | 0.67 | 0.73 | 0.48 | 0.77 | 0.25 |
| *Pardofelis marmorata* – non-bootstrapped GLM | 0.54 | 0.88 | 0.20 | 0.93 | 0.13 |
| *Pardofelis marmorata* – bootstrapped GLM | 0.66 | 0.71 | 0.51 | 0.72 | 0.23 |

| | | | | | |
|---|---|---|---|---|---|
| *Pardofelis marmorata* – non-bootstrapped RF | 0.67 | 0.88 | 0.30 | 0.92 | 0.22 |
| *Pardofelis marmorata* – bootstrapped RF | 0.69 | 0.83 | 0.38 | 0.86 | 0.24 |
| *Prionailurus javanensis* – non-bootstrapped GLM | 0.90 | 0.87 | 0.83 | 0.87 | 0.71 |
| *Prionailurus javanensis* – bootstrapped GLM | 0.73 | 0.59 | 0.71 | 0.58 | 0.30 |
| *Prionailurus javanensis* – non-bootstrapped RF | 0.81 | 0.91 | 0.67 | 0.92 | 0.58 |
| *Prionailurus javanensis* – | 0.73 | 0.90 | 0.38 | 0.91 | 0.30 |

| |
|---|
| bootstrapped<br><br>RF |

**Table S26.** Number of species with highest performing model based on AUC scores.

| AUC | Non-bootstrapped | Bootstrapped |
|---|---|---|
| GLM | 5 | 3 |
| RF | 2 | 2 |

**Table S27.** Number of species with highest performing model based on TSS scores.

| TSS | Non-bootstrapped | Bootstrapped |
|---|---|---|
| GLM | 7 | 1 |
| RF | 3 | 1 |

## References


Chiaverini, L., D. W. Macdonald, H. M. Bothwell, A. J. Hearn, S. M. Cheyne, I. Haidir, L. T. B. Hunter, Ż. Kaszta, E. A. Macdonald, J. Ross, and S. A. Cushman. 2022. Multi-scale, multivariate community models improve designation of biodiversity hotspots in the Sunda Islands. Animal Conservation.

CIESIN, and CIAT. 2016. Gridded Population of the World version 4 (GPWv4): Population density grids. Socioeconomic Data and Applications Center (SEDAC), Columbia University, Palisades, NY. http://sedac.ciesin.columbia.edu/gpw.

ESA Land Cover CCI. 2017. Product user guide version 2.0. http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf.

Evans, J. S., J. Oakleaf, S. A. Cushman, and D. Theobald. 2014. An ArcGIS toolbox for surface gradient and geomorphometric modeling, version 2.0-0. Available: http://evansmurphy.wix.com/evansspatial.

Fick, S. E., and R. J. Hijmans. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology **37**:4302-4315.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L.

Chini, C. O. Justice, and J. R. G. Townshend. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science **342**:850-853.

Hijmans, R. J., S. Phillips, J. Leathwick, and J. Elith. 2017. dismo: Species distribution modeling.

Jarvis, A., H. I. Reuter, A. Nelson, and E. Guevara. 2008. Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT). http://srtm.csi.cgiar.org.

Macdonald, D. W., L. Chiaverini, H. M. Bothwell, Z. Kaszta, E. Ash, G. Bolongon, O. E. Can, A. Campos-Arceiz, P. Channa, G. R. Clements, A. J. Hearn, L. Hedges, S. Htun, J. F. Kamler, E. A. Macdonald, J. Moore, H. Naing, M. Onuma, A. Rasphone, D. M. Rayan, J. Ross, P. Singh, C. K. W. Tan, J. Wadey, B. P. Yadav, and S. A. Cushman. 2020. Predicting biodiversity richness in rapidly changing landscapes: climate, low human pressure or protection as salvation? Biodiversity and Conservation.

Miettinen, J., C. H. Shi, W. J. Tan, and S. C. Liew. 2012. 2010 land cover map of insular Southeast Asia in 250-m spatial resolution. Remote Sensing Letters **3**:11-20.

UNEP-WCMC, and IUCN. 2022. Protected Planet: The World Database on Protected Areas (WDPA), Accessed April 2022, Cambridge, UK: UNEP-WCMC and IUCN. Available at: www.protectedplanet.net.

WCS, and CIESIN. 2005. Last of The Wild data version 2, (LTW-2): global human footprint dataset (geographic). http://sedac.ciesin.columbia.edu/wildareas/.

# Chapter 5

# Identifying gaps in the conservation of small wild cats of Southeast Asia.

Luca Chiaverini, David W. Macdonald, Andrew J. Hearn, Żaneta Kaszta, Eric Ash, Gilmoore Bolongon, Helen M. Bothwell, Ahimsa Campos-Arceiz, Özgün Emre Can, Phan Channa, Susan M. Cheyne, Gopalasamy Reuben Clements, Iding Achmad Haidir, Jan F. Kamler, Pyae Phyoe Kyaw, Jonathan H. Moore, Akchousanh Rasphone, Priya Singh, Cedric Kai Wei Tan, Bhupendra Yadav, Samuel A. Cushman

**Abstract**

Southeast Asia's biodiversity is severely threatened by habitat loss and degradation, which reduce the availability of intact habitats for a number of species. Carnivores are particularly vulnerable to anthropic pressure as they often occur at low population densities and require large home-ranges. Although Southeast Asia is one of the most biodiverse region globally, conservation is often catalysed by a few species, mainly represented by charismatic big carnivores. This is particularly evident for felids. Southeast Asia hosts more felids than any other region globally, but conservation actions are mostly focused on tigers and leopards, the two biggest felids of the region. Although smaller felids represent the majority of the Southeast Asian felids, information on their ecology and conservation is often lacking or limited to small geographic contexts. Building on the outcomes of a previous research that evaluated different modelling techniques to produce habitat suitability models for small felids in three study areas in Southeast Asia (mainland Southeast Asia, Borneo and Sumatra), we used the statistically most supported models for each species to produced predictive distribution surfaces for all the small felids together. We then evaluated the effectiveness of the current systems of protected areas in preserving large extents of highly suitable habitat, and highlighted the most suitable areas currently lacking protection. Finally, we assessed whether small felids in Southeast Asia are good proxies of the regional terrestrial vertebrate biodiversity. We highlighted that, in mainland Southeast Asia, the largest and most suitable habitat for small felids occurred in the Northern Forest Complex of Myanmar, where also the protected areas harbouring the largest extents of highly suitable habitat occurred. We also highlighted the importance for small felids of the region encompassed between East Myanmar, Laos and Vietnam. In Borneo, the most suitable habitat for small felids occurred in the central highlands of the islands and in Sabah. In these regions, protected areas had a fundamental role in preserving highly suitable habitats for small felids. In Sumatra, the

stronghold of habitat suitability for small felids was represented by the Barisan Mountains, in the western extent of the island. Also in Sumatra, protected areas were of fundamental importance to preserve suitable habitat, with the two largest protected areas harbouring more than one-third of the highly suitable habitat in the island. Additionally, we highlighted that small felids together correlated more strongly to terrestrial vertebrate biodiversity than single felid species individually. We provided the first regional assessment of the geographic distribution of small felids in Southeast Asia, highlighting the fundamental role of protected areas. Additionally, by highlighting the most suitable areas currently lacking protection, we provided important indications for conservation, and raised a red flag for the conservation of small felids.

## 1. Introduction

Among the regions where habitat loss and degradation severely threaten biodiversity, Southeast Asia is distinguished in having the highest deforestation rates globally (Hughes 2017). These are likely to persist and increase over the next decade at least (Cushman et al. 2017). If current rates of deforestation persist, Southeast Asia would lose three-quarter of its original forest cover by 2100 (Achard et al. 2002), with dramatic consequences for biodiversity, with a projected loss of between 24-63% of its terrestrial endemic species (Sodhi and Brook 2006).

Although extinction risks are ubiquitous across taxonomic groups, carnivores are particularly vulnerable to anthropic pressures, as their ecological role, atop the food pyramid, often dictates low population densities and vast home-ranges, requiring large amounts of ecological resources and leading to conflicts with humans (Cardillo et al. 2004). Additionally, habitat loss and fragmentation further reduce the availability of large, intact patches of suitable habitat for carnivores, which are inevitably pushed towards smaller, less suitable areas, often outside protected areas and embedded within anthropic matrices (Crooks et al. 2017, Cushman et al. 2018, Loveridge et al. 2022).

The urgency of providing information on Southeast Asian carnivores' conservation status is increased by the paucity of information for biodiversity in this region (Macdonald et al. 2020, Chiaverini et al. 2022), coupled with its very high level of biological diversity and rapid rates of habitat loss. While large charismatic species in this region have received most of the attention from conservationists (e.g., orangutans (Pandong et al. 2019), tigers (Ash et al. 2021), leopards (Rostro-Garcia et al. 2016) and clouded leopards (Macdonald et al. 2018, Macdonald et al. 2019)), smaller species are largely understudied (but see Macdonald et al.

(2020) and Chiaverini et al. (2022) for regional assessments of habitat suitability for terrestrial vertebrate biodiversity).

Southeast Asia is the richest region globally in terms of number of sympatric felids (Macdonald and Loveridge 2010). Of these, tigers and leopards have been the main focus of past conservation efforts and economic investments (Brodie 2009, Macdonald et al. 2015), and clouded leopards have recently gained attention (Macdonald et al. 2018, Macdonald et al. 2019). These are the biggest felids of the region and, unsurprisingly, information for smaller felids is often lacking or limited to small geographic contexts (Haidir et al. 2018, Hearn et al. 2018, Rasphone et al. 2021, Rostro-Garcia et al. 2021).

However, as secondary consumers, small felids have fundamental ecological roles in the trophic structure of ecosystems (Roemer et al. 2009), and they can also take on the role of apex predators through ecological release when larger carnivores are extirpated (de Oliveira et al. 2010). Additionally, small felids have been revealed to be important indicator and sentinel species (Cruz et al. 2019, Chiaverini et al. 2022), likely due to their relatively fast rates of population change and to their direct responses to both bottom-up and top-down ecological processes (Marneweck et al. 2022).

Therefore, we sought to develop the first regional assessment of habitat suitability and conservation status for Southeast Asia's small felids. By modelling habitat suitability individually for seven species of small felids sampled in three study areas (continental mainland, Borneo and Sumatra), our primary goal was to (1) define the geographic distribution of small felids in Southeast Asia. Then, we aimed to (2) evaluate the effectiveness, in terms of location, of the current systems of protected areas to preserve suitable habitats for small felids, and to (3) highlight the most suitable areas currently lacking

protection. Finally, we aimed to (4) assess whether small felids' suitable habitat approximated the geographic distribution of biodiversity richness in the region.

## 2. Materials and methods

### 2.1.Study area

The study area encompassed mainland Southeast Asia and the Sunda Islands of Borneo and Sumatra, extending between 5˚S-35˚N latitude and 80˚E-120˚E longitude, and spanning a broad altitudinal gradient ranging from coastal areas to Himalayan peaks above 8,000masl (Figure 1). Thirteen countries were included in the study area, of which Bangladesh, Bhutan, Cambodia, Laos, Malaysia, Myanmar, Nepal, Singapore, Thailand and Vietnam were entirely encompassed within the area, and China, India and Indonesia were partially included.

Due to the diversity of environments encompassed within this region, we masked out the areas that were ecologically highly divergent from the sampled locations by producing Mahalanobis distance surfaces (Chiaverini et al. 2022), using the package *dismo* (Hijmans et al. 2017) in R v4.2.0 (R Core Team 2018). We calculated the Mahalanobis distances using the environmental factors that determined an unevenly distributed survey: tree canopy cover (Hansen et al. 2013), protected areas (UNEP-WCMC and IUCN 2022), human footprint (WCS and CIESIN 2005), elevation (Jarvis et al. 2008), roughness (Evans et al. 2014), annual maximum temperature, annual minimum temperature and annual precipitations (PRISM Climate Group 2016) separately for mainland Southeast Asia, Borneo and Sumatra. The Mahalanobis surfaces were continuous layers of the multidimensional distances of each pixel across these regions where the camera traps were deployed. We excluded areas that were highly ecologically divergent from our sampled locations by applying thresholds for mainland Southeast Asia and for the Sunda Islands. Respectively, we used thresholds of the

40th and 20th percentile of the distribution of the multidimensional distances, accounting for the different extents of the mainland and offshore regions.

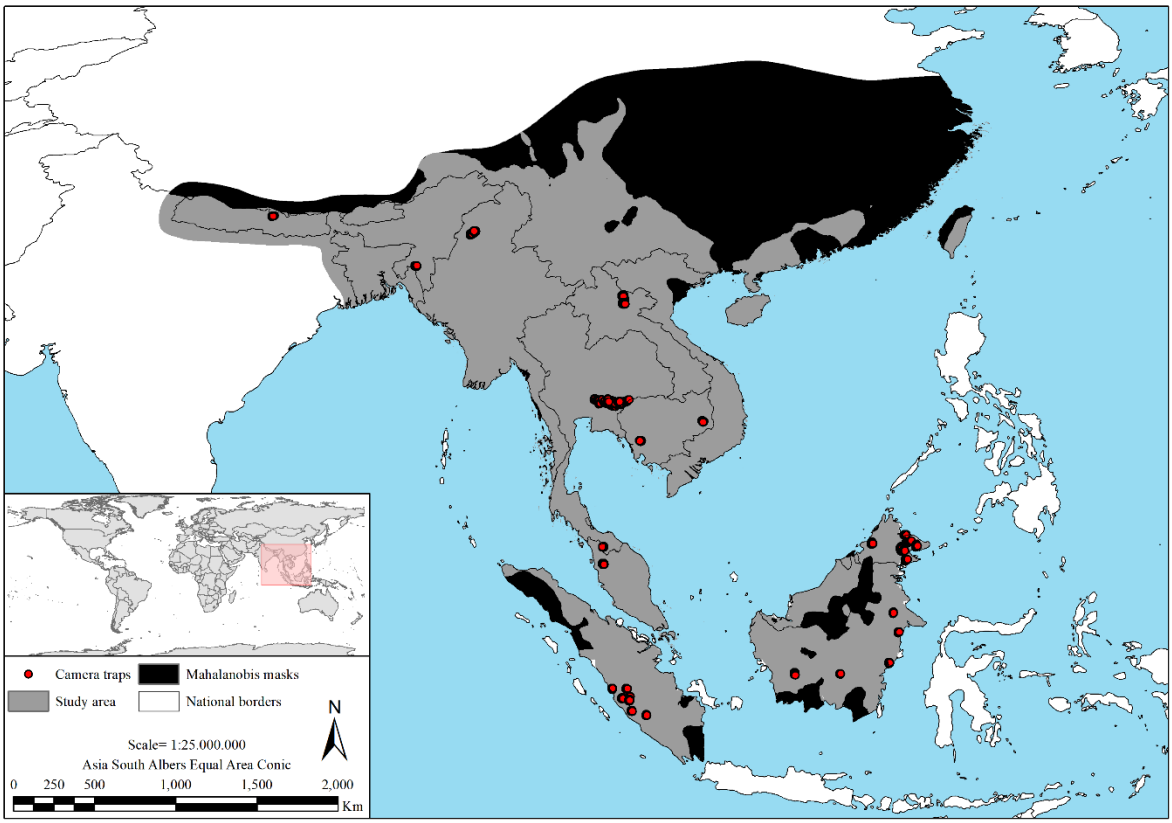



**Figure 1.** Map of the study area. Shown are the locations of the camera traps and the Mahalanobis masks produced to cover the regions too different from the sampling locations.

### 2.2.Data collection

Data on felid detections were collected with camera trap surveys carried out by our field teams between 2007 and 2016 across Southeast Asia. We deployed camera trap grids in 37 landscapes across Southeast Asia, of which 15 were on the mainland, 15 in Borneo and 7 in Sumatra, yielding a sampling network of 2,928 camera trap stations (1,384 in mainland, 801

in Borneo and 743 in Sumatra). The total sampling effort produced 253,904 trap nights, of which 115,389 were in mainland, 77,687 in Borneo and 60,828 in Sumatra (Figure 1; Tables S1-S3).

Primarily, the surveys were designed to sample mainland and Sunda clouded leopards (*Neofelis nebulosa* and *Neofelis diardi*, respectively), whose geographic ranges shaped the extents of the mainland and offshore study areas. However, the surveys also collected data for other members of the felids guilds. Specifically, on the mainland, we recorded clouded leopard, Asiatic golden cat (*Catopuma temminckii*), marbled cat (*Pardofelis marmorata*) and leopard cat (*Prionailurus bengalensis*) respectively at 235, 148, 116 and 345 stations. In Borneo, we recorded Sunda clouded leopard, Borneo bay cat (*Catopuma badia*), marbled cat and Sunda leopard cat (*Prionailurus javanensis*) respectively at 154, 25, 85 and 202 stations. In Sumatra, we recorded Sunda clouded leopard, Asiatic golden cat, marbled cat and Sunda leopard cat respectively at 101, 102, 39 and 23 stations (Tables S1-S3).

Camera trap grids were deployed mainly in protected areas, and each consisted of approximately 80 camera trap stations placed 1.0-2.0km apart. For each station, a pair of camera traps was positioned at ~40cm above the ground, along forest trails, natural ridgelines and disused logging roads, allowing greater visibility to maximise felids' detection success (Hearn et al. 2018, Macdonald et al. 2018, Macdonald et al. 2019).

### 2.3.Habitat covariates

To assess the relationships of the sampled felids with habitat factors, we selected two sets of covariates, one for the mainland and another for the Sunda Islands, reflecting the different habitats of the two regions. Covariates were chosen based on knowledge on felid ecology accrued from previous research in Southeast Asia (Hearn et al. 2018, Macdonald et al. 2018,

Macdonald et al. 2019), and included landscape, anthropic, topographic and climatic factors (Tables S4-S5).

For each covariate, we calculated ecologically informative metrics in FRAGSTATS (McGarigal et al. 2012) and ArcMap v10.6.1 (ESRI 2018). Specifically, in FRAGSTATS we implemented percentage of landscape (PLAND) for discrete layers, which calculates, for each pixel, the percentage of landscape occupied by a covariate within a defined circular buffer. In ArcMap we implemented focal mean (FM) for continuous layers, which calculates, for each pixel, the ratio between the extent of a covariate within a circular buffer, and the total area of the buffer.

Additionally, we calculated the aforementioned metrics by using circular buffers with varying radii already demonstrated to be relevant for Southeast Asian biodiversity (Macdonald et al. 2020, Chiaverini et al. 2022): for the mainland, we used radii of 250m, 500m, 1km, 2km, 4km, 8km, 16km and 32km, while for the Sunda Islands we investigated radii of 250m, 500m, 1km, 2km, 4km, 6km, 8km and 10km, accounting for the different extents of the two regions. We then applied a multi-scale optimisation approach (*sensu* McGarigal et al. (2016)) using the methods described in Macdonald et al. (2018) and Macdonald et al. (2019) to produce predictive occurrence models for each species reflecting the multi-scale nature of the species-habitat relationships.

### 2.4.Modelling frameworks

Building on the methodologies applied and discussed in (Chiaverini et al. 2023), we produced habitat suitability models for the sampled felids by applying two different algorithms and, for each of them, two data re-sampling techniques, resulting in four different modelling frameworks. Specifically, we produced habitat suitability models by applying

generalised linear model (GLM) and random forest (RF) and, for each of them, we trained the models by bootstrapping or not the training data.

The modelling frameworks described below have been applied independently for each species, for each study area, and they have been implemented in R v4.2.0.

### 2.4.1. GLM – Scales optimisation and covariates selection

We evaluated the most representative spatial scales by performing univariate GLMs, independently for each covariate at each scale (Wasserman et al. 2012, Hearn et al. 2018). Then, for each covariate, we retained the scale in which the univariate GLM presented the lowest Akaike Information Criterion (AIC). We trained the univariate GLMs with a subset of the original dataset, composed of a randomly selected 80% of the data, and used the remaining 20% for model validation. Then, we checked the scale-optimised covariates for multicollinearity by calculating Pearson's correlation index between all covariates. When a covariates' dyad was found to be highly correlated (i.e., $|r| \geq 0.7$), we discarded the covariate whose univariate GLM showed the lowest adjusted-$R^2$ (Guisan and Zimmermann 2000).

### 2.4.2. GLM – Model training

Using the final set of standardised, uncorrelated, scale-optimised covariates as explanatory variables and the felids' presence-absence data as response variables, we trained GLMs by applying two frameworks. For the first approach (hereafter non-bootstrapped GLM), we trained the models by using the same dataset used for the univariate GLMs, composed of 80% of the original dataset. We first conducted an all-subset GLM analysis using the *MuMIn* package (Bartoń 2013), and we then ranked the candidate models using

AIC, retaining only those with ΔAIC≤ 2 (Burnham and Anderson 2002). Finally, by using the covariates' coefficients averaged from the suite of competing models, we created a layer predicting the species' habitat suitability.

For the second approach (hereafter bootstrapped GLM), we subsampled the original full dataset by randomly selecting 20% of the presence and of the absence data. We reiterated this selection 100 times with replacement, producing as many subsets as we used to train 100 independent GLMs by using the same framework described above for the non-bootstrapped GLM. We then produced the final habitat suitability layer by averaging the outcomes of the 100 iterations.

### 2.4.3. GLM – Model validation

We evaluated the performances of the non-bootstrapped GLM by calculating Area Under the ROC Curve (AUC) using the 20% of the data held-out from the training dataset, using the package *PresenceAbsence* (Freeman and Moisen 2007).

To evaluate the performances of the bootstrapped GLM, we calculated AUC independently for each of the 100 iterations, using the data held-out from the training datasets, composed of 80% of the presence and of the absence data. We then obtained the final AUC by averaging the AUC from each iteration.

### 2.4.4. RF – Scales optimisation and covariates selection

To investigate the most representative spatial scales with RF, we performed univariate RF independently for each covariate at each scale using the original full dataset and selecting, for each covariate, the scale whose univariate RF showed the lowest out-of-bag (OOB) error. We

trained the univariate RF for scale-optimisation by tuning 9,999 bootstrap replicates. Bradter et al. (2013) proposed a different method to perform spatial scales model optimisation with random forest, allowing the simultaneous evaluation of several scales of multiple predictors. However, we decided to implement a univariate spatial scale optimisation to make the RF modelling framework comparable to the framework implemented for GLM. To perform RF, we did not check for multicollinearity among scale-optimised covariates, since the algorithm is usually considered more robust to correlated explanatory variables than linear regression (Cutler et al. 2007). However, other authors argued that RF can be negatively affected by correlated variables, unlike what usually believed (Genuer et al. 2010, Bradter et al. 2013). These authors highlighted that correlated variables provide redundant information, potentially leading to over-fitting and decreasing model performance. As a result, if correlated variables are included in the model, the information they provide would be redundant, and their variable importance might decrease since it would be divided among them, and none of them might show high importance individually. While acknowledging this potential issue, we decided not to check for multicollinearity. We implemented the analyses using the package *randomForest* (Liaw and Wiener 2002).

### 2.4.5. RF – Model training

We applied two modelling frameworks using the final set of scale-optimised covariates and the felid presence-absence data. Both the modelling frameworks were composed of two steps: 1) variables selection, implemented on the scale-optimised covariates by tuning 2,000 bootstrap replicates, and 2) RF training, performed by tuning 500 bootstrap replicates and by setting the number of randomly drawn candidate variables (*mtry*) to 10. For the first approach (hereafter non-bootstrapped RF), we trained the RF by using the same dataset used to train

the non-bootstrapped GLM, composed of randomly selected 80% of the original dataset, and produced the species' habitat suitability layers.

For the second approach (hereafter bootstrapped RF), we independently produced 100 subsets of the original full dataset by randomly sampling, with replacement, 20% of the presence and of the absence data. Then, for each subset, we trained RF and produced 100 habitat suitability layers, which we eventually averaged to obtain the final habitat suitability layer.

### 2.4.6. RF – Model validation

To evaluate the performances of the non-bootstrapped RF, we calculated AUC using the 20% of the hold-out data subset.

To evaluate the performances of the bootstrapped RF, for each of the 100 iterations we calculated AUC using the data held-out from the training dataset, composed of 80% of the presence and absence data. We then averaged the validation metrics from each iteration to calculate the final AUC.

## 2.5.Felid species habitat suitability

For each sampled felid species, we selected the best model among the four competing modelling frameworks by choosing the one showing the highest AUC in the independent validation (Chiaverini et al. 2023). While AUC is a widely used method to evaluate the performance of binary models, limitations and drawbacks have been highlighted. Namely, AUC is highly susceptible to imbalanced datasets, and misleading values could be provided when the number of samples in one class significantly outweighs the other. In such cases, the

AUC value might be high even if the model is not performing well on the minority class, as it tends to prioritise the majority class. This case is particularly relevant in our work, since the number of absences in the camera trap sampling outnumbers the presences. However, in this case, the biased AUC values would have determined a more cautious model validation, with AUC values lower than the case with a balanced presence/absence ratio. Additionally, since models for the same species have been trained with the same dataset, the AUC values would have been similarly biased, and the relative comparison of the AUC values between models would have led to the same results. In addition, independently for mainland Southeast Asia, Borneo and Sumatra, we highlighted the most suitable areas for all the sampled felid species together by combining by addition the single-species habitat suitability layers.

### 2.6.Protected areas effectiveness

We evaluated the effectiveness of the current protected area systems in mainland Southeast Asia, Borneo and Sumatra in preserving the hotspots of suitability for all sampled felids together by reclassifying the habitat suitability layers based on the percentile distribution of their suitability values. Specifically, we classified as low suitability the pixels whose values ranged between the $50^{th}$ and the $70^{th}$ percentile, as medium suitability those with values ranging between the $70^{th}$ and the $90^{th}$ percentile and as high suitability the pixels with suitability values higher than the $90^{th}$ percentile (Macdonald et al. 2018, Chiaverini et al. 2022). Then, we defined the effectiveness of the protected areas by calculating the ratio between the extent of each suitability class (i.e., low, medium and high suitability, as specified above) encompasses within each protected area, and the extent of that class in the entire study area, independently for mainland Southeast Asia, Borneo and Sumatra. The protected area layer was obtained from the World Database on Protected Areas (UNEP-

WCMC and IUCN 2022), and we consistently followed its structure without deleting, merging or altering the protected areas, thereby ensuring that our assessment is reproducible throughout the study area.

### 2.7.Gap analysis

Focusing only on the medium and high suitability classes to emphasise the areas with the highest conservation value, we produced a gap analysis by highlighting the areas predicted to be important to felid diversity occurring outside the current systems of protected areas. We performed gap analyses independently for mainland Southeast Asia, Borneo and Sumatra.

### 2.8.Felids as biodiversity indicators

To evaluate the effectiveness of small felids as biodiversity indicators, independently for each study area, we calculated Pearson's correlation coefficients between the most supported model for each felid and the species richness surfaces produced for mainland Southeast Asia and Sunda Islands, respectively, by Macdonald et al. (2020) and Chiaverini et al. (2022). Additionally, for each study area, we calculated the Pearson's correlation coefficients also between the sum of habitat suitability layers across felids altogether and the species richness surfaces.

## 3. Results

### 3.1.Model training and validation

Using the final sets of scale-optimised covariates, we produced habitat suitability models for the sampled felids by applying the four modelling frameworks, and evaluated the best

model for each species based on the AUC scores. The species-habitat relationships for the best models, described either by GLM's coefficients or by RF's mean decrease accuracies, are provided in Tables S9-S20.

We evaluated the predictive performances of the four modelling techniques by calculating the AUC scores of the resulting habitat suitability layers choosing, for each felid, the model showing the highest AUC. In mainland Southeast Asia, the best models for clouded leopard, Asiatic golden cat, marbled cat and leopard cat showed, respectively, AUC= 0.59, 0.79, 0.77 and 0.65. In Borneo, the best models for Sunda clouded leopard, Borneo bay cat, marbled cat and Sunda leopard cat showed, respectively, AUC= 0.80, 0.87, 0.81 and 0.77. In Sumatra, the best models for Sunda clouded leopard, Asiatic golden cat, marbled cat and Sunda leopard cat showed, respectively, AUC= 0.82, 0.77, 0.69 and 0.90.

### 3.2.Felid habitat suitability

Independently for each study area, we combined by addition the statistically most supported single-species habitat suitability model for each felid species to produce a layer representing the predicted species richness across felids. Here we present the assessment of the protected areas effectiveness and gap analysis based on this summed suitability across felid species (Figures 2-4). In addition, we provide the single-species habitat suitability layers in Figures S1-S12.

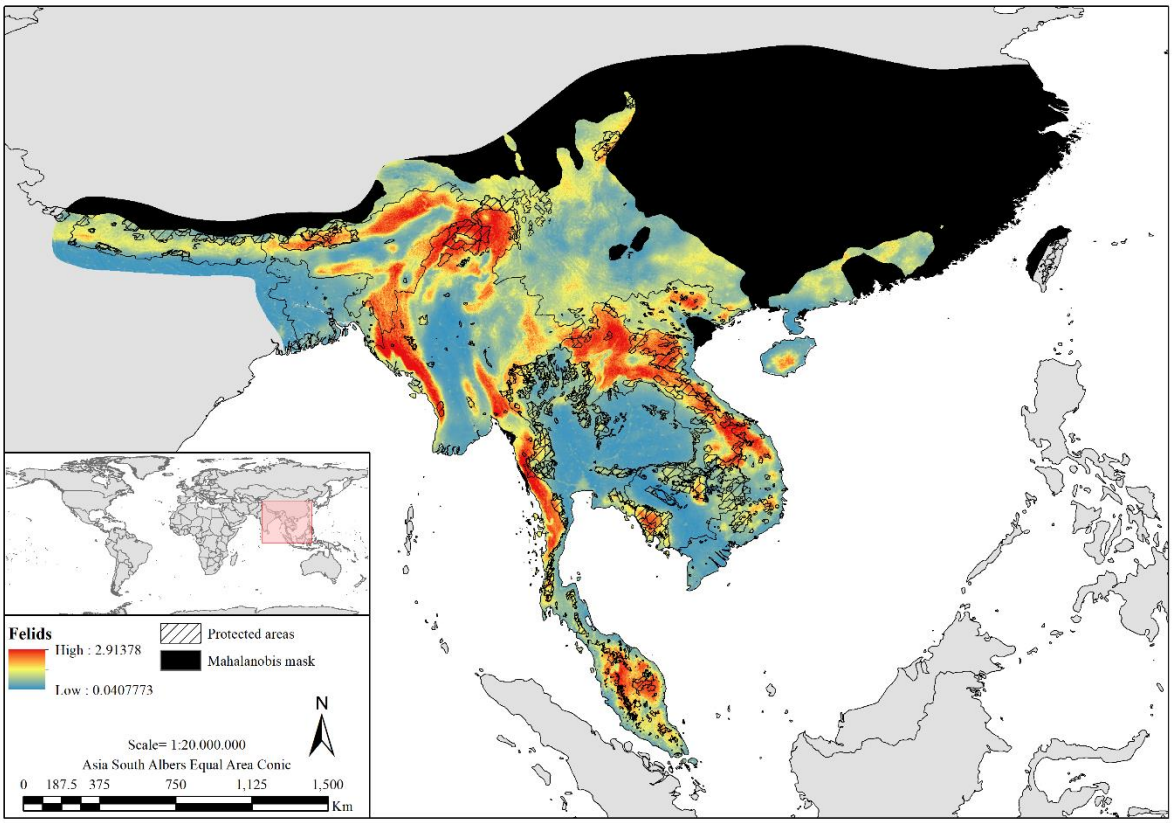



**Figure 2.** Species distribution probability surface of the sampled felid species in mainland Southeast Asia.

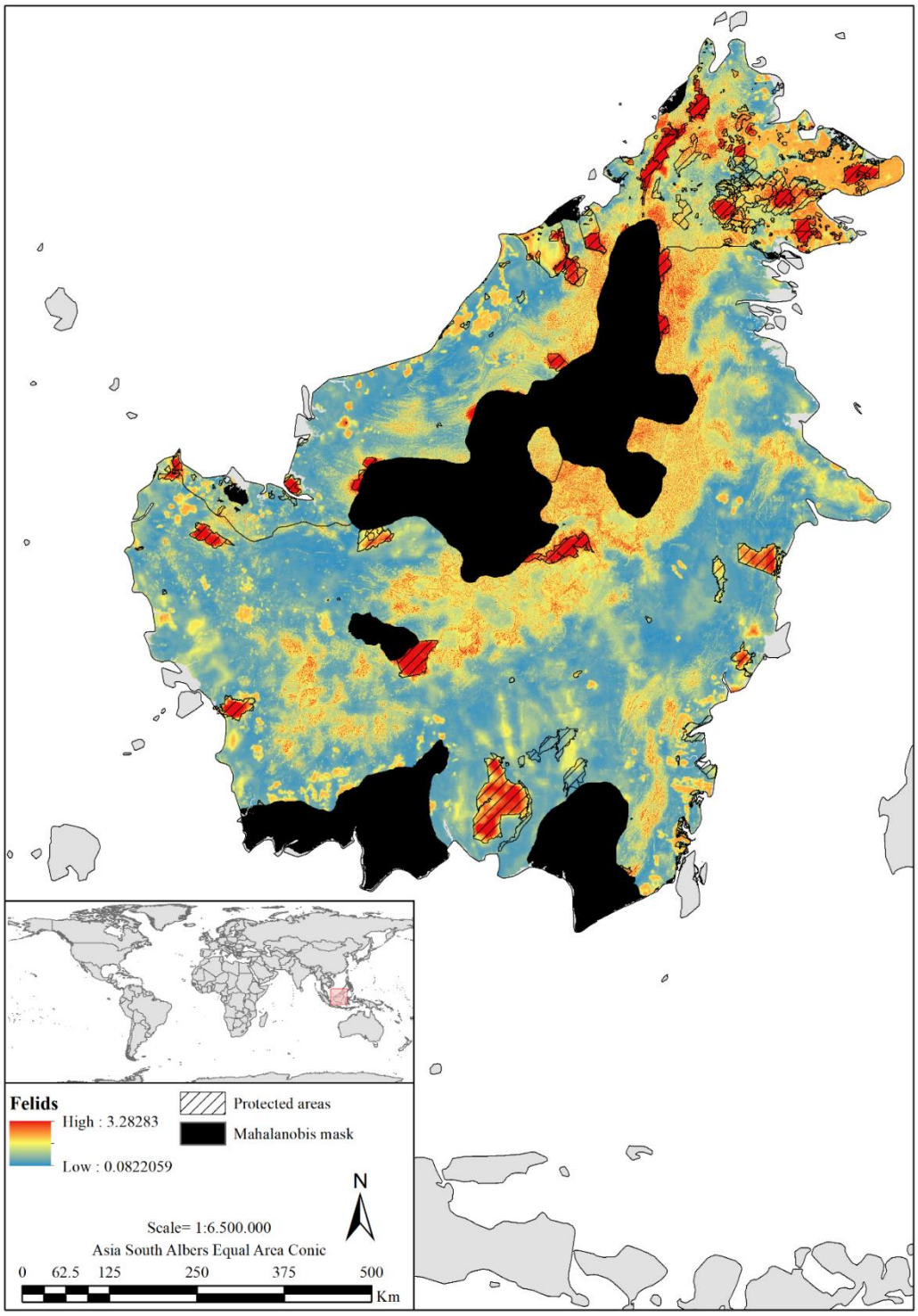



**Figure 3.** Species distribution probability surface of the sampled felid species in Borneo.

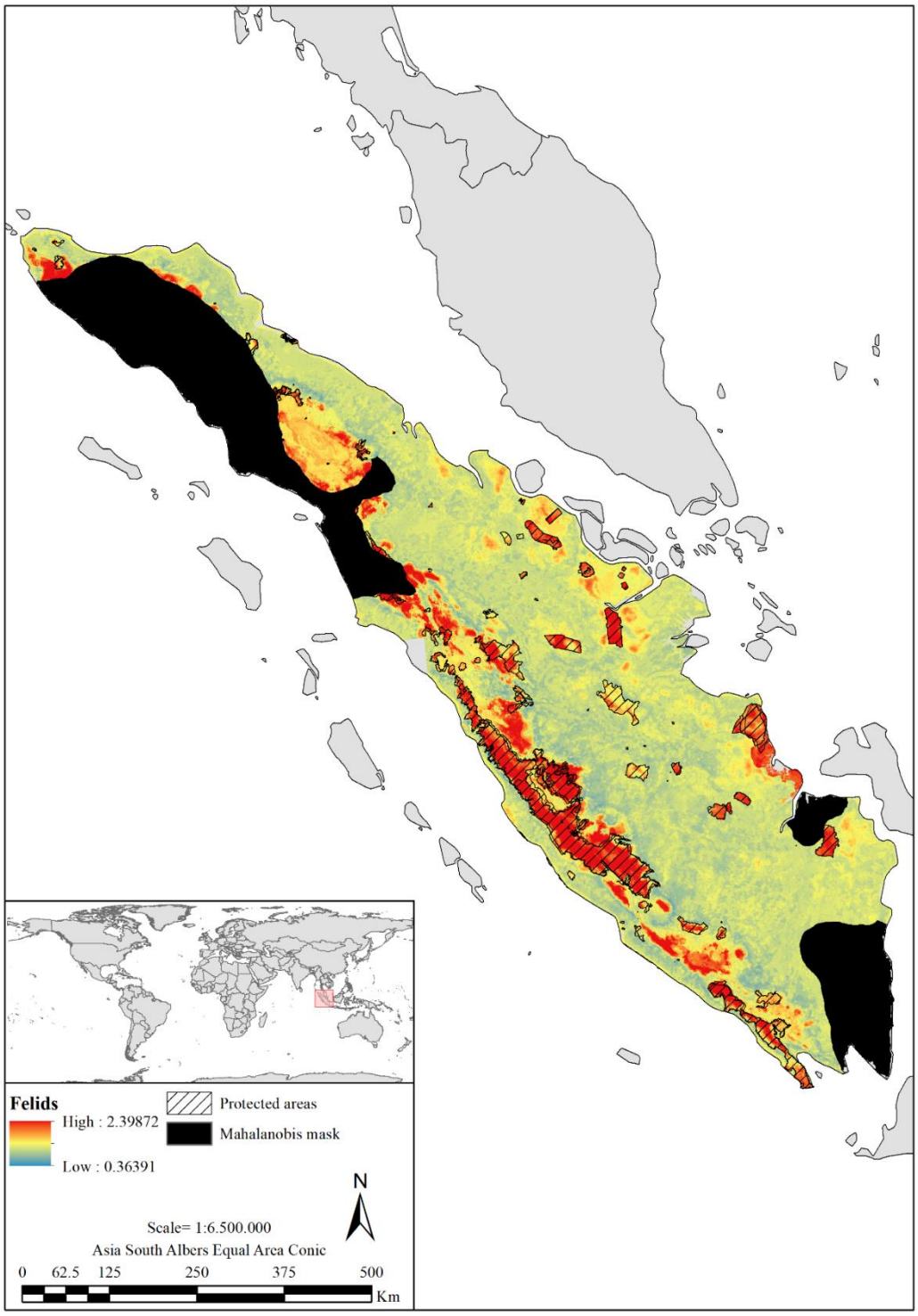



**Figure 4.** Species distribution probability surface of the sampled felid species in Sumatra.

### 3.3.Protected areas effectiveness

In mainland Southeast Asia, several of the protected areas that were most effective in preserving highly suitable habitat for small felid conservation were located in the Northern Forest Complex of Myanmar (Figure 5). Specifically, the most effective protected area in terms of containing highly suitable habitat for multiple felid species was Hukaung Valley Wildlife Sanctuary (extension), while Hukaung Valley Wildlife Sanctuary and Bumpha Bum Wildlife Sanctuary ranked respectively as the third and fourth most effective protected areas (Table 1). By this criterion of congruity between habitat suitability and protected areas placement, the second most important protected area was Western Nghe An UNESCO-MAB Biosphere Reserve, occurring in Vietnam. In Malaysia, we highlighted the effectiveness of the Taman Negara (Pahang) National Park (fifth highest effectiveness rank), followed by Virajay National Park in Cambodia (sixth rank), where also Central Kravanh and Southern Kravanh National Parks occurred (seventh and ninth rank, respectively). Among the most effective protected areas were also Taninthayi Nature Reserve in Myanmar (eighth rank) and Dong Ampham National Protected Area in Laos (tenth rank).

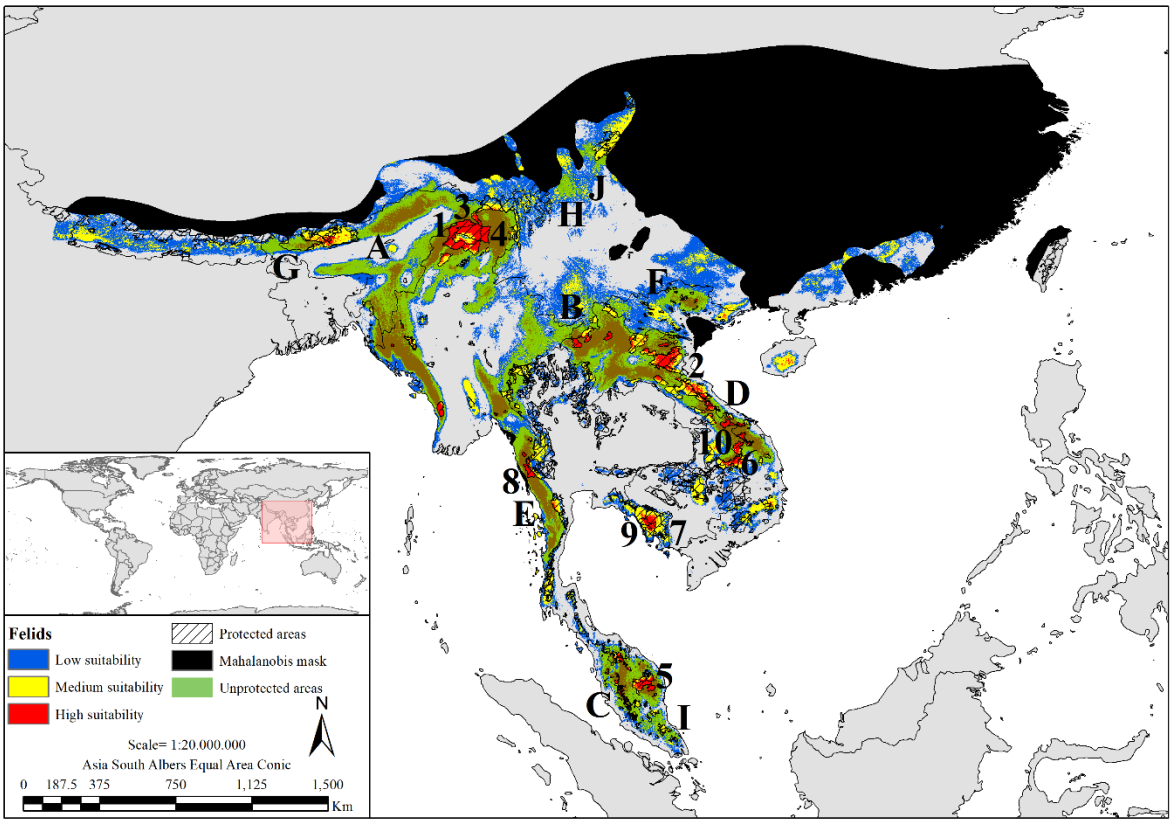



**Figure 5.** Reclassified probability surface of the sampled felid species in mainland Southeast Asia, showing areas with low, medium and high habitat suitability. Codes of the protected areas and of the unprotected suitable areas are reported in Tables 1 and 4, respectively.

**Table 1.** Main protected areas in terms of extent of high suitability habitat for the sampled felid species in mainland Southeast Asia. Reported are the extent of the low, medium and high suitability habitats occurring in each protected area, and the proportion of the extent of the suitability class over the extent of that class in mainland Southeast Asia. Codes refer to the main protected areas displayed in Figure 5.

| Protected areas | Code | Country | Low suitability habitat ($Km^2$) | Low suitability habitat (%) | Medium suitability habitat ($Km^2$) | Medium suitability habitat (%) | High suitability habitat ($Km^2$) | High suitability habitat (%) |
|---|---|---|---|---|---|---|---|---|
| Hukaung Valley Wildlife Sanctuary (extension) | 1 | Myanmar | 3.94 | 0.00 | 1,305.31 | 0.19 | 9,992.06 | 3.00 |
| Western Nghe An | 2 | Vietnam | 679.75 | 0.09 | 3,278.00 | 0.47 | 5,648.63 | 1.69 |
| Hukaung Valley Wildlife Sanctuary | 3 | Myanmar | 6.13 | 0.00 | 929.88 | 0.13 | 5,553.88 | 1.67 |
| Bumpha Bum Wildlife Sanctuary | 4 | Myanmar | 0.00 | 0.00 | 5.69 | 0.00 | 2,923.00 | 0.88 |
| Taman Negara (Pahang) | 5 | Malaysia | 0.00 | 0.00 | 614.94 | 0.09 | 2,012.06 | 0.60 |
| Virajay | 6 | Cambodia | 138.13 | 0.02 | 1,391.06 | 0.20 | 1,774.13 | 0.53 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Central Kravanh | 7 | Cambodia | 58.81 | 0.01 | 2,271.94 | 0.33 | 1,675.81 | 0.50 |
| Taninthayi Nature Reserve | 8 | Myanmar | 0.00 | 0.00 | 5.25 | 0.00 | 1,605.25 | 0.48 |
| Southern Kravanh | 9 | Cambodia | 486.88 | 0.07 | 1,967.44 | 0.28 | 1,583.75 | 0.47 |
| Dong Ampham | 10 | Laos | 0.00 | 0.00 | 413.06 | 0.06 | 1,570.69 | 0.47 |

In Borneo, we found that one-fourth of the highly suitable felid habitat was encompassed within the first ten protected areas for extent of highly suitable habitat encompassed (Table 2). The two most important protected areas for felid habitat quality in Borneo were Sebangau National Park and Bukit Baka – Bukit Raya National Park, in the Indonesian provinces of Central and West Kalimantan (Figure 6). Next in importance based on our habitat criteria, we highlighted Bukit Sapat Hawung Nature Reserve in the highlands of the central Borneo and Crocker Range National Park in Sabah (third and fourth rank, respectively). Tabin Wildlife Reserve and Kinabalu Park, also in Sabah, were ranked in sixth and eight places among the most effective Bornean protected areas. Among the areas protecting the largest extents of highly suitable habitat, we also identified Kutai National Park in East Kalimantan (fifth rank), Kayan Mentarang National Park in North Kalimantan (seventh rank), and Gunung Nyiut Penrissen Nature Reserve (ninth rank) and Gunung Palung National Park in West Kalimantan (tenth rank).

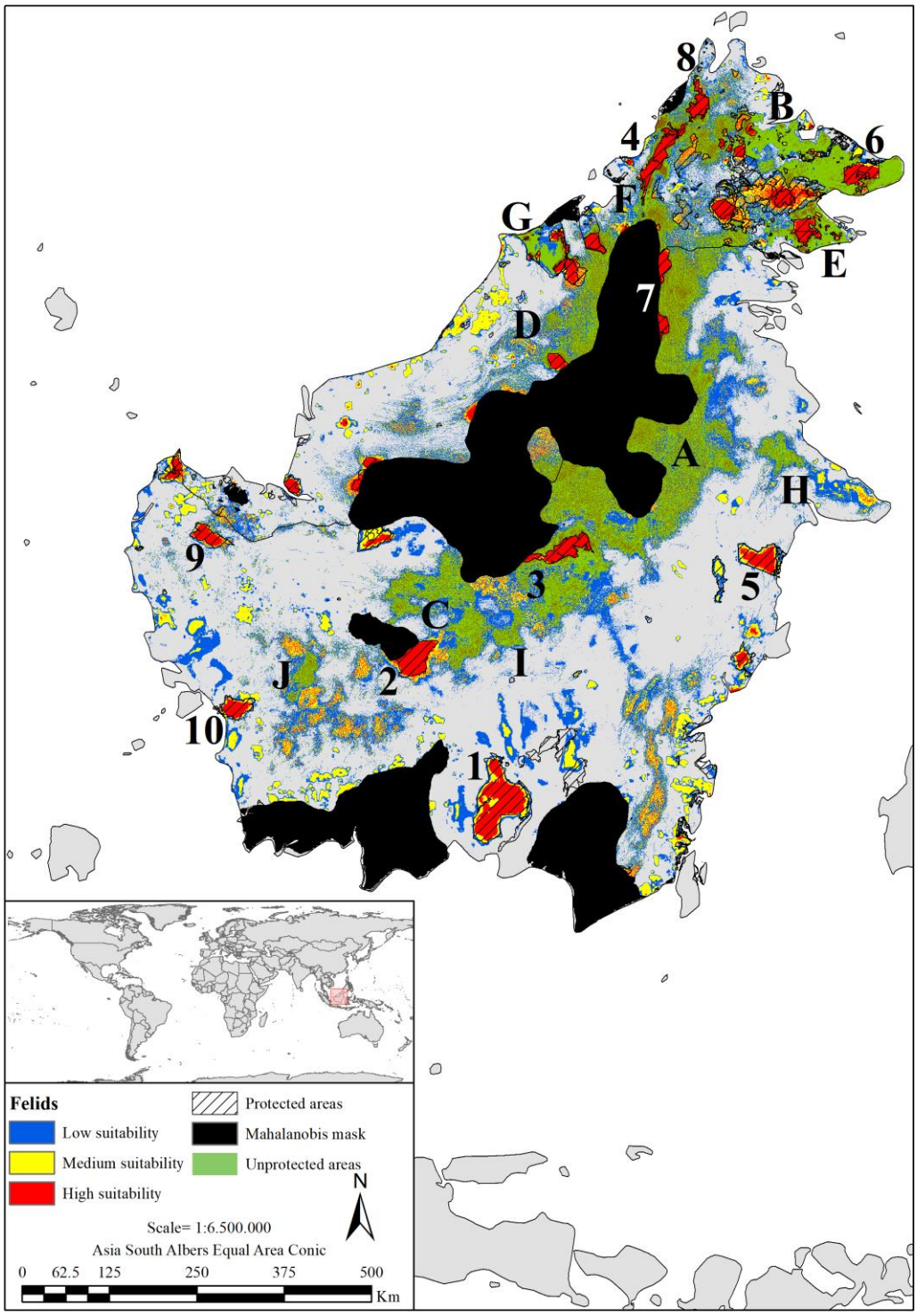



**Figure 6.** Reclassified probability surface of the sampled felid species in Borneo, showing areas with low, medium and high habitat suitability. Codes of the protected areas and of the unprotected suitable areas are reported in Tables 2 and 5, respectively.

**Table 2.** Main protected areas in terms of extent of high suitability habitat for the sampled felid species in Borneo. Reported are the extent of the low, medium and high suitability habitats occurring in each protected area, and the proportion of the extent of the suitability class over the extent of that class in Borneo. Codes refer to the main protected areas displayed in Figure 6.

| Protected areas | Code | Low suitability habitat ($Km^2$) | Low suitability habitat (%) | Medium suitability habitat ($Km^2$) | Medium suitability habitat (%) | High suitability habitat ($Km^2$) | High suitability habitat (%) |
|---|---|---|---|---|---|---|---|
| Sebangau | 1 | 443.81 | 0.39 | 1,274.38 | 1.12 | 3,784.88 | 7.11 |
| Bukit Baka - Bukit Raya | 2 | 18.94 | 0.02 | 69.31 | 0.06 | 1,625.50 | 3.05 |
| Bukit Sapat Hawung | 3 | 0.00 | 0.00 | 67.94 | 0.06 | 1,491.06 | 2.80 |
| Crocker Range National Park | 4 | 2.25 | 0.00 | 26.31 | 0.02 | 1,367.75 | 2.57 |
| Kutai | 5 | 233.06 | 0.20 | 505.81 | 0.45 | 1,096.44 | 2.06 |
| Tabin Wildlife Reserve | 6 | 4.56 | 0.00 | 155.88 | 0.14 | 955.44 | 1.79 |
| Kayan Mentarang | 7 | 0.81 | 0.00 | 32.94 | 0.03 | 840.94 | 1.58 |
| Kinabalu Park | 8 | 3.44 | 0.00 | 51.38 | 0.05 | 781.50 | 1.47 |
| Gunung Nyiut Penrissen | 9 | 31.56 | 0.03 | 167.69 | 0.15 | 711.75 | 1.34 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Gunung Palung | 10 | 93.13 | 0.08 | 231.44 | 0.20 | 704.81 | 1.32 |

In Sumatra, we highlighted the overwhelming importance, based on our criteria, of Kerinci Seblat National Park, followed by Bukit Barisan Selatan National Park, respectively the first and third largest protected areas in the island (Figure 7; Table 3). Next in importance, we highlighted Kerumutan Wildlife Reserve (third rank), Padang Sugihan Wildlife Reserve (fourth rank) and Berbak National Park (fifth rank), all in the eastern coastal region of Sumatra, where we identified also Giam Siak Kecil Wildlife Reserve (seventh rank) as an important protected area for felid diversity. Bukit Rimbang Bukit Baling Wildlife Reserve and Arau Hilir dan Air Terusan Wildlife Reserve, both in the Barisan Mountains, ranked respectively as the sixth and eight most effective protected areas. In the interior of Sumatra, we found that Air Sawan National Park and Dangku Wildlife Reserve were of considerable importance in terms of habitat for the felid species (ninth and tenth rank, respectively).

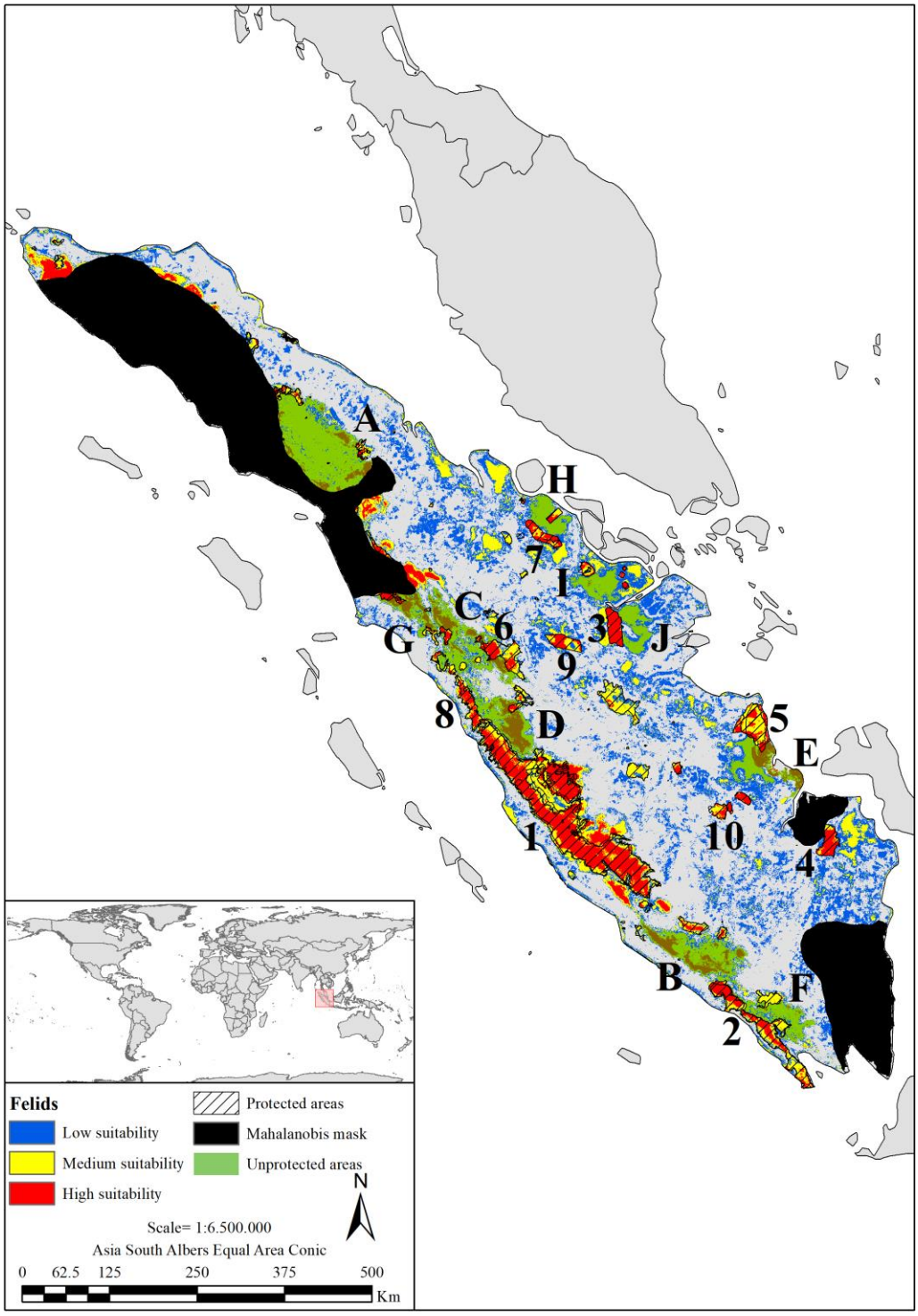



**Figure 7.** Reclassified probability surface of the sampled felid species in Sumatra, showing areas with low, medium and high habitat suitability. Codes of the protected areas and of the unprotected suitable areas are reported in Tables 3 and 6, respectively.

**Table 3.** Main protected areas in terms of extent of high suitability habitat for the sampled felid species in Sumatra. Reported are the extent of the low, medium and high suitability habitats occurring in each protected area, and the proportion of the extent of the suitability class over the extent of that class in Sumatra. Codes refer to the main protected areas displayed in Figure 7.

| Protected areas | Code | Low suitability habitat ($Km^2$) | Low suitability habitat (%) | Medium suitability habitat ($Km^2$) | Medium suitability habitat (%) | High suitability habitat ($Km^2$) | High suitability habitat (%) |
|---|---|---|---|---|---|---|---|
| Kerinci Seblat | 1 | 246.69 | 0.36 | 3,340.13 | 5.40 | 9,753.06 | 32.58 |
| Bukit Barisan Selatan | 2 | 166.88 | 0.24 | 1,463.38 | 2.36 | 1,406.00 | 4.70 |
| Kerumutan | 3 | 0.50 | 0.00 | 37.00 | 0.06 | 909.06 | 3.04 |
| Padang Sugihan | 4 | 0.88 | 0.00 | 167.38 | 0.27 | 680.06 | 2.27 |
| Berbak | 5 | 0.31 | 0.00 | 1,080.56 | 1.75 | 643.50 | 2.15 |
| Bukit Rimbang Bukit Baling | 6 | 120.44 | 0.17 | 642.50 | 1.04 | 602.88 | 2.01 |
| Giam Siak Kecil | 7 | 0.50 | 0.00 | 205.81 | 0.33 | 576.31 | 1.93 |
| Arau Hilir dan Air Terusan | 8 | 22.06 | 0.03 | 340.94 | 0.55 | 536.31 | 1.79 |
| Air Sawan | 9 | 54.94 | 0.08 | 331.38 | 0.54 | 429.81 | 1.44 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Dangku | 10 | 0.69 | 0.00 | 204.75 | 0.33 | 241.75 | 0.81 |

### 3.4.Gap analysis

In mainland Southeast Asia, the largest unprotected area characterised by medium and high habitat suitability for small felid species occurred largely in northern Myanmar, but extended also into the easternmost Indian states, Bangladesh, Bhutan and China (area A in Figure 5; Table 4). Almost contiguous to this area, another highly important area for felid habitat was represented by the Himalayan foothills of Bhutan (area G). In the Chinese province of Sichuan, we highlighted two smaller areas important for felid habitat suitability (areas H and J). Along the border between Myanmar and Thailand we identified the second largest suitable unprotected area, which extended across much of two countries, but also included portions of Laos, China and Vietnam (area B). Adjacent to this area, we identified another area of high felid habitat suitability extending across the border between Laos and Vietnam (area D). In the northern parts of Vietnam, we found an additional unprotected area of suitable habitat for multiple felid species (area F). We also highlighted suitable areas in southern Myanmar (area E), and in the southernmost extent of the Malay Peninsula, where we highlighted two adjacent areas across Malaysia and Thailand (areas C and I).

**Table 4.** Largest unprotected suitable areas for the sampled felid species in mainland Southeast Asia, including medium and high suitability habitats. Codes refer to the largest unprotected suitable areas displayed in Figure 5.

| Code | Country | Area ($Km^2$) |
|---|---|---|
| A | Bangladesh, Bhutan, China, India, Myanmar | 289,168.00 |
| B | China, Laos, Myanmar, Thailand, Vietnam | 192,444.06 |
| C | Malaysia, Thailand | 49,414.38 |
| D | Cambodia, Laos, Vietnam | 43,892.63 |
| E | Myanmar, Thailand | 42,650.38 |
| F | China, Vietnam | 18,395.50 |
| G | Bhutan, China, India, Nepal | 13,684.19 |
| H | China | 7,198.88 |
| I | Malaysia | 6,559.31 |
| J | China | 6,332.38 |

In Borneo, the largest area characterised by unprotected high and medium suitability habitat extended across the central highlands of the Indonesian provinces of North, East and Central Kalimantan (area A in Figure 6; Table 5). South of the region excluded by the Mahalanobis mask, we identified two areas in Central and West Kalimantan (areas C and I). Additionally, in the Malaysian state of Sarawak we highlighted an important unprotected area of highly suitable felid habitat (area D). In Sabah, we highlighted the second largest suitable unprotected area (area B) and, adjacent to this one, we identified another area occurring in the eastern parts of Sabah (area E). Importantly, we mapped an unprotected area between the central highlands and Sabah playing a crucial role connecting these two strongholds (area F).

We also mapped an area of unprotected area for felid conservation extending within the borders of Brunei Darussalam (area G). Additionally, we highlighted suitable areas occurring in East Kalimantan (area H) and West Kalimantan (area J).

**Table 5.** Largest unprotected suitable areas for the sampled felid species in Borneo, including medium and high suitability habitats. Codes refer to the largest unprotected suitable areas displayed in Figure 6.

| Code | Area ($Km^2$) |
|---|---|
| A | 39,547.81 |
| B | 14,983.38 |
| C | 7,511.06 |
| D | 6,461.69 |
| E | 3,100.25 |
| F | 2,693.06 |
| G | 1,819.81 |
| H | 1,507.44 |
| I | 1,182.94 |
| J | 1,173.81 |

In Sumatra, the largest suitable unprotected area covered a large portion of the mountainous parts of North Sumatra (area A in Figure 7; Table 6), while the second largest suitable unprotected area was identified in the southern Barisan Mountains, connecting Kerinci Seblat and Bukit Barisan Selatan National Parks (area B). Additionally, we

highlighted another area of highly important unprotected felid habitat extending outwards from Bukit Barisan Selatan National Park (area F). Similarly, we found a suitable unprotected area contiguous with Kerinci Seblat National Park (area D). In the Barisan Mountains, we highlighted two additional areas (areas C and G). We identified unprotected suitable areas also in eastern Sumatra, in close proximity with several protected areas. Specifically, we mapped an area of suitable felid habitat surrounding Giam Siak Kecil Wildlife Reserve (area H), and two suitable areas adjacent to Kerumutan Wildlife Reserve (areas I and J). We also found suitable unprotected habitat adjacent to Berbak National Park (area E).

**Table 6.** Largest unprotected suitable areas for the sampled felid species in Sumatra, including medium and high suitability habitats. Codes refer to the largest unprotected suitable areas displayed in Figure 7.

| Code | Area ($Km^2$) |
|---|---|
| A | 10,274.38 |
| B | 4,952.38 |
| C | 4,435.88 |
| D | 3,511.38 |
| E | 3,013.75 |
| F | 2,308.25 |
| G | 1,737.38 |
| H | 1,666.94 |
| I | 1,651.00 |
| J | 1,639.56 |

### 3.5.Felids as biodiversity indicators

The assessment of the role of small felids as biodiversity indicators revealed that the multi-species predictive layers for sampled felids together showed, overall, higher Pearson's correlation coefficients than the single-species habitat suitability layers. Therefore, the predictive layers for small felids together were more correlated to the geographic distribution of terrestrial vertebrate biodiversity modelled by Macdonald et al. (2020) and by Chiaverini et al. (2022), than the predictive layers for small felids individually.

In mainland Southeast Asia, Borneo and Sumatra, the Pearson's correlation coefficients of the predictive layers for small felids together were, respectively, 0.31, 0.44 and 0.55. Only in mainland Southeast Asia the modelled distribution of a single felid was more correlated to the terrestrial vertebrate biodiversity than the distribution of small felids together, with Asiatic golden cat showing Pearson's correlation coefficient= 0.36 (Table 7).

**Table 7.** Pearson's correlation coefficients of the single-species habitat suitability models and of the predictive surface of the sampled felids altogether, with the layer of vertebrate terrestrial biodiversity for mainland Southeast Asia, Borneo and Sumatra.

| | Clouded leopard | Sunda clouded leopard | Asiatic golden cat | Borneo bay cat | Marbled cat | Leopard cat | Sunda leopard cat | All felids |
|---|---|---|---|---|---|---|---|---|
| Biodiversity mainland | 0.30 | - | 0.36 | - | 0.25 | -0.06 | - | 0.31 |
| Biodiversity Borneo | - | 0.37 | - | 0.33 | 0.31 | - | 0.26 | 0.44 |
| Biodiversity Sumatra | - | 0.33 | 0.50 | - | 0.51 | - | 0.04 | 0.55 |

## 4. Discussion

Our study provides the first regional distribution maps and multi-species hotspot maps based on the habitat preferences of small felids across Southeast Asia, assessing independently three different study areas on the basis of their environments and felid guilds composition: mainland Southeast Asia, Borneo and Sumatra. We identified the most suitable areas for small felids, assessed the effectiveness of protected areas in preserving highly suitable habitat for the species, and highlighted the most suitable areas still lacking formal protection. Finally, we evaluated the effectiveness of small felids as indicator species of regional biodiversity. Our results draw attention to the conservation of small felids in Southeast Asia, which have rarely been the focus of conservation actions (Brodie 2009,

Macdonald et al. 2015), despite their ecological importance (de Oliveira et al. 2010, Chiaverini et al. 2022) and potential ambassadorial role (Macdonald et al. 2017). To our knowledge, this is the first assessment of the habitat suitability for these species at this broad scale and in comparing protected area effectiveness and gap analysis at a regional scale.

### 4.1.Mainland Southeast Asia

We highlighted the most suitable habitats for small felids sampled in mainland Southeast Asia: these emerged as defined hotspots, surrounded by areas of lower suitability. Specifically, one stronghold of habitat suitability occurred in the Northern Forest Complex of Myanmar, a vast system of contiguous forests extending in northern Myanmar, at the border with China and India. Macdonald et al. (2020) found this area, characterised by contiguous forests and wide altitudinal range, to be crucial also for the overall vertebrate terrestrial biodiversity, and it was identified as the most important node in the regional habitat and connectivity network for clouded leopard (Kaszta et al. 2020b). Unsurprisingly, we found that the predictive layer produced from the aggregated single species' habitat suitability models was a better predictor of terrestrial biodiversity richness in mainland Southeast Asia, as modelled by (Macdonald et al. 2020), than the single species habitat suitability models (except for the predictive surface of Asiatic golden cat). This result was expected, since the suitable habitat for multiple felids included broader gradients of habitat characteristics, which would be more closely associated with the ecological needs of the modelled terrestrial biodiversity as a whole (Cushman and Landguth 2012, Macdonald et al. 2012). The topographic heterogeneity along the altitudinal gradient in this region enhances environmental variety, promoting niche diversity that supports the coexistence of several potential competitors in a relatively small area (Wikramanayake 2002, Macdonald et al.

2020). The importance of the Northern Forest Complex of Myanmar for small felids was corroborated by the fact that here we identified the most important protected areas in mainland Southeast Asia. Our results confirmed that protected areas in northern Myanmar are of immense importance to felid biodiversity conservation, corroborating the importance of vast and intact patches of forests for regional biodiversity (Macdonald et al. 2020). However, Kaszta et al. (2020a) highlighted the critical threats that clouded leopard in northern Myanmar face from the Belt and Road Initiative (Yu 2017) and other regional developments, both in terms of loss of core habitats and dispersal corridors, and these risks can likely be generalised to other felids if protection of the remaining suitable habitats is not enforced. We highlighted not only the critical role of the existing protected areas, but also the importance of the suitable unprotected areas for small felids: northern Myanmar and adjacent areas of neighbouring countries encompasses the largest area of highly suitable habitat currently lacking protection.

Beyond the Northern Forest Complex of Myanmar, suitable habitats in the mainland study region were heavily constrained by the surrounding disturbed and degraded landscapes occurring in the Indian state of Assam and in the central parts of Myanmar. While the border between Myanmar and India showed high habitat suitability, driven by the large forested areas in the region, the central and more developed portions of Myanmar had much lower habitat suitability. We identified another stronghold of suitability in eastern Myanmar, extending across most of Laos and southward toward the Malay Peninsula. This region was only partially highlighted as a biodiversity hotspot by Macdonald et al. (2020), and differences between that more broadly based analysis and our current research focused on small felids were particularly evident in northern Laos. The highly suitable habitats extended from this region towards the border between Laos and Vietnam. In this area we found some of the most important protected areas in mainland Southeast Asia, based on habitat potential

for small felids. These promising results for small felids in Laos and Vietnam come in the wider context of the extirpation of tigers (*Panthera tigris*) and leopards (*Panthera pardus*) from northern Laos (Rasphone et al. 2019, Rasphone et al. 2021). If the rampant snaring in this region is not staunched, other species will soon follow (Johnson et al. 2016). Additionally, and as in northern Myanmar, the Belt and Road Initiative is likely to impact biodiversity negatively also in northern Laos (Ng et al. 2020).

Landscapes degraded by intense agricultural activities were a major limit to the extensiveness of suitable habitat for small felids in Cambodia, where the only suitable areas remained in the protected areas on the northern border with Laos and in the Cardamom Mountains. Although we identified suitable habitats in Cambodian protected areas, biodiversity in the country is critically threatened by hunting and snaring (Harrison et al. 2016). In Thailand, the negative effects of land degradation were particularly evident: along the northern portion of the Malay Peninsula, the boundary between suitable and unsuitable habitats could be traced along the border with Myanmar. On the Myanmar side of that border, we identified both protected and unprotected suitable habitats in forested areas along the peninsular extent of the country, while on the Thai side little suitable habitat was identified.

### 4.2.Borneo

In Borneo, we highlighted the critical role of protected areas for small felids, given that we found that one-fourth of highly suitable habitat was encompassed within the ten most effective protected areas. This result was unexpected, since Chiaverini et al. (2022) found that less than 10% of Bornean terrestrial biodiversity was encompassed within protected areas. This difference suggests that small felids are more specialised than the terrestrial Bornean biodiversity generally and may depend more heavily on protected areas, although we also

highlighted that the aggregated predictive layer produced from the single species habitat suitability models was a much better descriptor of terrestrial biodiversity than the individual single species habitat suitability models suggesting, as expectable, that the ecological requirements for multiple felids species together are more correlated to the ecological requirements for the terrestrial biodiversity as a whole (Cushman and Landguth 2012, Macdonald et al. 2012). Chiaverini et al. (2022) found the highest numbers of vertebrate terrestrial species in the Malaysian state of Sabah, in the north of the island, while the largest and most suitable habitat for small felids occurred in the central highlands of Borneo. Additionally, our predicted distribution for small felids collectively, which closely matched the distribution of Bornean carnivores by Mathai et al. (2016), was more correlated to the terrestrial biodiversity mapped by Chiaverini et al. (2022) than was any of the single-species habitat suitability models individually.

The central highlands of Borneo, although partly covered by the Mahalanobis mask, was the region where we modelled the largest and most suitable habitats for small felids in the island. This area is less impacted by forest loss than the surrounding regions (Cushman et al. 2017). It represents a vital refuge for felids, and for other specialised species, compared to the more disturbed lowlands (Scriven et al. 2015). Importantly, this region overlapped with the Heart of Borneo initiative, a transboundary conservation area signed between Brunei Darussalam, Indonesia and Malaysia to manage the last undisturbed forests of the island (Keong and Onuma 2021). In the central highlands our modelled distribution for small felids matched the Heart of Borneo's proposed outline, but extended beyond it. Extending from the Heart of Borneo, we modelled suitable habitats in the mountainous region along the Mangkalihat Peninsula in East Kalimantan, along the range of the Meratus Mountains in South Kalimantan, and in West and Central Kalimantan. In these regions, we highlighted scattered unprotected suitable areas, but also the most important protected areas in Borneo.

Additionally, we highlighted also the northernmost extent of Sabah as harbouring expansive suitable habitat for small felids both within and outside protected areas. Although the carnivore distribution maps produced by Mathai et al. (2016) differed from our prediction in this area, our predictions corresponded to those for Sunda leopard cat by Hearn et al. (2018). Sunda leopard cat is a generalist species, occurring also in disturbed landscapes, such as logged forests and oil palm plantations (Rajaratnam et al. 2007). Supporting our conclusions, Chiaverini et al. (2022) found this area to be a hotspot for Bornean terrestrial biodiversity.

### 4.3.Sumatra

In Sumatra, we highlighted the overwhelming importance of the Barisan Mountains, extending along the western coastal region, and representing the stronghold of suitability for small felids. Our predicted multispecies suitability for small felids closely matched the modelled habitat suitability for Sunda clouded leopard on the island (Macdonald et al. 2018). Additionally, our predicted multispecies suitability layer was also strongly correlated with the modelled terrestrial biodiversity (Macdonald et al. 2020) supporting, as seen also in mainland Southeast Asia and in Borneo that, when considered together, the ecological requirement of the sampled felids were more similar to those of the sampled terrestrial biodiversity (Cushman and Landguth 2012, Macdonald et al. 2012). Although the entire mountain range harboured suitable habitats, the most suitable areas largely corresponded with Kerinci Seblat and Bukit Barisan Selatan National Parks, two of the largest protected areas in Sumatra. Notably, both these areas are part of the Tropical Rainforest Heritage of Sumatra, a UNESCO World Heritage site that includes also Gunung Leuser National Park, occurring in the northern parts of the island that were almost entirely excluded by the Mahalanobis mask.

Importantly, along the Barisan Mountains we also mapped extensive regions of suitable but unprotected areas, often occurring in close association with the most important protected areas.

In the eastern regions of Sumatra we also identified highly suitable habitats in association with protected areas, mirroring the findings of Chiaverini et al. (2022) for terrestrial vertebrate biodiversity. The eastern provinces of Sumatra are among the most affected by deforestation for oil palm and wood fibre plantations (Singh and Yan 2021), and protected areas likely represent the only remaining refuge for small felids in this region, except for leopard cat, which have higher tolerance for disturbed ecosystems (Mohamed et al. 2013). Importantly, we found suitable unprotected areas in close association with most of the main protected areas in eastern Sumatra as well. These areas are notable candidates for additional protection, with the capacity to buffer existing protected areas from deforestation and land degradation (Singh and Yan 2021).

## 5. References

Achard, F., H. D. Eva, H. J. Stibig, P. Mayaux, J. Gallego, T. Richards, and J. P. Malingreau. 2002. Determination of deforestation rates of the world's humid tropical forests. Science **297**:999-1002.

Ash, E., Z. Kaszta, A. Noochdumrong, T. Redford, P. Chanteap, C. Hallam, B. Jaroensuk, S. Raksat, K. Srinoppawan, and D. W. Macdonald. 2021. Opportunity for Thailand's forgotten tigers: assessment of the Indochinese tiger Panthera tigris corbetti and its prey with camera-trap surveys. Oryx **55**:204-211.

Bartoń, K. 2013. MuMIn: Multi-model inference.

Bradter, U., W. E. Kunin, J. D. Altringham, T. J. Thom, and T. G. Benton. 2013. Identifying appropriate spatial scales of predictors in species distribution models with the random forest algorithm. Methods in Ecology and Evolution **4**:167-174.

Brodie, J. F. 2009. Is research effort allocated efficiently for conservation? Felidae as a global case study. Biodiversity and Conservation **18**:2927-2939.

Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer, New York, NY, USA.

Cardillo, M., A. Purvis, W. Sechrest, J. L. Gittleman, J. Bielby, and G. M. Mace. 2004. Human population density and extinction risk in the world's carnivores. Plos Biology **2**:909-914.

Chiaverini, L., D. W. Macdonald, H. M. Bothwell, A. J. Hearn, S. M. Cheyne, I. Haidir, L. T. B. Hunter, Ż. Kaszta, E. A. Macdonald, J. Ross, and S. A. Cushman. 2022. Multi-scale, multivariate community models improve designation of biodiversity hotspots in the Sunda Islands. Animal Conservation.

Chiaverini, L., D. W. Macdonald, A. J. Hearn, Z. Kaszta, E. Ash, H. M. Bothwell, Ö. E. Can, P. Channa, G. R. Clements, I. A. Haidir, P. P. Kyaw, J. H. Moore, A. Rasphone, C. K. W. Tan, and S. A. Cushman. 2023. Not seeing the forest for the trees: Generalised linear model out-performs random forest in species distribution modelling for Southeast Asian felids. Ecological Informatics **75**.

Crooks, K. R., C. L. Burdett, D. M. Theobald, S. R. B. King, M. Di Marco, C. Rondinini, and L. Boitani. 2017. Quantification of habitat fragmentation reveals extinction risk in terrestrial mammals. Proceedings of the National Academy of Sciences of the United States of America **114**:7635-7640.

Cruz, P., C. De Angelo, J. M. Pardo, M. E. Iezzi, D. Varela, M. S. Di Bitetti, and A. Paviolo. 2019. Cats under cover: Habitat models indicate a high dependency on woodlands by Atlantic Forest felids. Biotropica **51**:266-278.

Cushman, S. A., N. B. Elliott, D. Bauer, K. Kesch, L. Bahaa-el-din, H. Bothwell, M. Flyman, G. Mtare, D. W. Macdonald, and A. J. Loveridge. 2018. Prioritizing core areas, corridors and conflict hotspots for lion conservation in southern Africa. Plos One **13**.

Cushman, S. A., and E. L. Landguth. 2012. Multi-taxa population connectivity in the Northern Rocky Mountains. Ecological Modelling **231**:101-112.

Cushman, S. A., E. Macdonald, E. Landguth, Y. Malhi, and D. Macdonald. 2017. Multiple-scale prediction of forest loss risk across Borneo. Landscape Ecology **32**:1581-1598.

Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, and K. T. Hess. 2007. Random forests for classification in ecology. Ecology **88**:2783-2792.

de Oliveira, T. G., M. A. Tortato, L. Silveira, C. B. Kasper, F. D. Mazim, M. Lucherini, A. T. Jácomo, J. B. G. Soares, R. V. Marques, and M. Sunquist. 2010. Ocelot ecology and its effect on the small-felid guild in the lowland Neotropics. Pages 559-580 *in* D. W. Macdonald and A. J. Loveridge, editors. The Biology and Conservation of Wild Felids. Oxford University Press, Oxford, UK.

ESRI. 2018. ArcGIS Desktop: Release 10.6.1. Environmental System Research Institute. Redlands, CA, USA.

Evans, J. S., J. Oakleaf, S. A. Cushman, and D. Theobald. 2014. An ArcGIS toolbox for surface gradient and geomorphometric modeling, version 2.0-0. Available: http://evansmurphy.wix.com/evansspatial.

Freeman, E., and G. Moisen. 2007. PresenceAbsence: An R Package for Presence Absence Analysis. Journal of Statistical Software **23**.

Genuer, R., J. M. Poggi, and C. Tuleau-Malot. 2010. Variable selection using random forests. Pattern Recognition Letters **31**:2225-2236.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling **135**:147-186.

Haidir, I. A., D. W. Macdonald, and M. Linkie. 2018. Assessing the spatiotemporal interactions of mesopredators in Sumatra's tropical rainforest. Plos One **13**.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science **342**:850-853.

Harrison, R. D., R. Sreekar, J. F. Brodie, S. Brook, M. Luskin, H. O'Kelly, M. Rao, B. Scheffers, and N. Velho. 2016. Impacts of hunting on tropical forests in Southeast Asia. Conservation Biology **30**:972-981.

Hearn, A. J., S. A. Cushman, J. Ross, B. Goossens, L. T. B. Hunter, and D. W. Macdonald. 2018. Spatio-temporal ecology of sympatric felids on Borneo. Evidence for resource partitioning? Plos One **13**.

Hijmans, R. J., S. Phillips, J. Leathwick, and J. Elith. 2017. dismo: Species distribution modeling.

Hughes, A. C. 2017. Understanding the drivers of Southeast Asian biodiversity loss. Ecosphere **8**.

Jarvis, A., H. I. Reuter, A. Nelson, and E. Guevara. 2008. Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT). http://srtm.csi.cgiar.org.

Johnson, A., J. Goodrich, T. Hansel, A. Rasphone, S. Saypanya, C. Vongkhamheng, Venevongphet, and S. Strindberg. 2016. To protect or neglect? Design, monitoring,

and evaluation of a law enforcement strategy to recover small populations of wild tigers and their prey. Biological Conservation **202**:99-109.

Kaszta, Z., S. A. Cushman, S. Htun, H. Naing, D. Burnham, and D. W. Macdonald. 2020a. Simulating the impact of Belt and Road initiative and other major developments in Myanmar on an ambassador felid, the clouded leopard, Neofelis nebulosa. Landscape Ecology **35**:727-746.

Kaszta, Z., S. A. Cushman, and D. W. Macdonald. 2020b. Prioritizing habitat core areas and corridors for a large carnivore across its range. Animal Conservation.

Keong, C. Y., and A. Onuma. 2021. Transboundary Ecological Conservation, Environmental Value, and Environmental Sustainability: Lessons from the Heart of Borneo. Sustainability **13**.

Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. R News **2**:18-22.

Loveridge, A. J., L. L. Sousa, S. Cushman, Z. Kaszta, and D. W. Macdonald. 2022. Where have all the lions gone? Establishing realistic baselines to assess decline and recovery of African lions. Diversity and Distributions **28**:2388-2402.

Macdonald, D. W., H. M. Bothwell, A. J. Hearn, S. M. Cheyne, I. Haidir, L. T. B. Hunter, A. Kaszta, M. Linkie, E. A. Macdonald, J. Ross, and S. A. Cushman. 2018. Multi-scale habitat selection modeling identifies threats and conservation opportunities for the Sunda clouded leopard (Neofelis diardi). Biological Conservation **227**:92-103.

Macdonald, D. W., H. M. Bothwell, Ż. Kaszta, E. Ash, G. Bolongon, D. Burnham, Ö. E. Can, A. Campos-Arceiz, P. Channa, G. R. Clements, A. J. Hearn, L. Hedges, S. Htun, J. F. Kamler, K. Kawanishi, E. A. Macdonald, S. W. Mohamad, J. Moore, H. Naing, M. Onuma, U. Penjor, A. Rasphone, D. Mark Rayan, J. Ross, P. Singh, C. K. W. Tan, J. Wadey, B. P. Yadav, and S. A. Cushman. 2019. Multi-scale habitat modelling

identifies spatial conservation priorities for mainland clouded leopards (Neofelis nebulosa). Diversity and Distributions **0**:1-16.

Macdonald, D. W., D. Burnham, A. E. Hinks, and R. Wrangham. 2012. A problem shared is a problem reduced: seeking efficiency in the conservation of felids and primates. Folia Primatol (Basel) **83**:171-215.

Macdonald, D. W., L. Chiaverini, H. M. Bothwell, Z. Kaszta, E. Ash, G. Bolongon, O. E. Can, A. Campos-Arceiz, P. Channa, G. R. Clements, A. J. Hearn, L. Hedges, S. Htun, J. F. Kamler, E. A. Macdonald, J. Moore, H. Naing, M. Onuma, A. Rasphone, D. M. Rayan, J. Ross, P. Singh, C. K. W. Tan, J. Wadey, B. P. Yadav, and S. A. Cushman. 2020. Predicting biodiversity richness in rapidly changing landscapes: climate, low human pressure or protection as salvation? Biodiversity and Conservation.

Macdonald, D. W., and A. J. Loveridge. 2010. Biology and conservation of wild felids. Oxford University Press, Oxford ; New York.

Macdonald, E. A., D. Burnham, A. E. Hinks, A. J. Dickman, Y. Malhi, and D. W. Macdonald. 2015. Conservation inequality and the charismatic cat: Felis felicis. Global Ecology and Conservation **3**:851-866.

Macdonald, E. A., A. Hinks, D. J. Weiss, A. Dickman, D. Burnham, C. J. Sandom, Y. Malhi, and D. W. Macdonald. 2017. Identifying ambassador species for conservation marketing. Global Ecology and Conservation **12**:204-214.

Marneweck, C. J., B. L. Allen, A. R. Butler, E. Do Linh San, S. N. Harris, A. J. Jensen, E. A. Saldo, M. J. Somers, K. Titus, M. Muthersbaugh, A. Vanak, and D. S. Jachowski. 2022. Middle-out ecology: small carnivores as sentinels of global change. Mammal Review **52**:471-479.

Mathai, J., J. W. Duckworth, E. Meijaard, G. Fredriksson, J. Hon, A. Sebastian, M. Ancrenaz, A. J. Hearn, J. Ross, S. Cheyne, A. Wilting, and B. C. Consortium. 2016. Carnivore

conservation planning on Borneo: identifying key carnivore landscapes, research priorities and conservation interventions. Raffles Bulletin of Zoology:186-217.

McGarigal, K., S. A. Cushman, and E. Ene. 2012. FRAGSTATS v4: spatial pattern analysis program for categorical and continuous maps. Computer software program produced by the authors at the University of Massachusetts, Amherst, MA. Available: http://www.umass.edu/landeco/research/fragstats/fragstats.html.

McGarigal, K., H. Y. Wan, K. A. Zeller, B. C. Timm, and S. A. Cushman. 2016. Multi-scale habitat selection modeling: a review and outlook. Landscape Ecology **31**:1161-1175.

Mohamed, A., R. Sollmann, H. Bernard, L. N. Ambu, P. Lagan, S. Mannan, H. Hofer, and A. Wilting. 2013. Density and habitat use of the leopard cat (Prionailurus bengalensis) in three commercial forest reserves in Sabah, Malaysian Borneo. Journal of Mammalogy **94**:82-89.

Ng, L., A. Campos-Arceiz, S. Sloan, A. C. Hughes, D. C. F. Tiang, B. V. Li, and A. M. Lechner. 2020. The scale of biodiversity impacts of the Belt and Road Initiative in Southeast Asia. Biological Conservation **248**.

Pandong, J., M. Gumal, Z. M. Aton, M. S. Sabki, and L. P. Koh. 2019. Threats and lessons learned from past orangutan conservation strategies in Sarawak, Malaysia. Biological Conservation **234**:56-63.

PRISM Climate Group. 2016. Oregon State University. http://prism.oregonstate.edu.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rajaratnam, R., M. Sunquist, L. Rajaratnam, and L. Ambu. 2007. Diet and habitat selection of the leopard cat (Prionailurus bengalensis borneoensis) in an agricultural landscape in Sabah, Malaysian Borneo. Journal of Tropical Ecology **23**:209-217.

Rasphone, A., J. F. Kamler, M. Tobler, and D. W. Macdonald. 2021. Density trends of wild felids in northern Laos. Biodiversity and Conservation **30**:1881-1897.

Rasphone, A., M. Kery, J. F. Kamler, and D. W. Macdonald. 2019. Documenting the demise of tiger and leopard, and the status of other carnivores and prey, in Lao PDR's most prized protected area: Nam Et - Phou Louey. Global Ecology and Conservation **20**.

Roemer, G. W., M. E. Gompper, and B. Van Valkenburgh. 2009. The Ecological Role of the Mammalian Mesocarnivore. Bioscience **59**:165-173.

Rostro-Garcia, S., J. F. Kamler, E. Ash, G. R. Clements, L. Gibson, A. J. Lynam, R. McEwing, H. Naing, and S. Paglia. 2016. Endangered leopards: Range collapse of the Indochinese leopard (Panthera pardus delacouri) in Southeast Asia. Biological Conservation **201**:293-300.

Rostro-Garcia, S., J. F. Kamler, C. Minge, A. Caragiulo, R. Crouthers, M. Groenenberg, T. N. E. Gray, V. In, C. Pin, P. Sovanna, M. Kery, and D. W. Macdonald. 2021. Small cats in big trouble? Diet, activity, and habitat use of jungle cats and leopard cats in threatened dry deciduous forests, Cambodia. Ecology and Evolution **11**:4205-4217.

Scriven, S. A., J. A. Hodgson, C. J. McClean, and J. K. Hill. 2015. Protected areas in Borneo may fail to conserve tropical forest biodiversity under climate change. Biological Conservation **184**:414-423.

Singh, M., and S. H. Yan. 2021. Spatial-temporal variations in deforestation hotspots in Sumatra and Kalimantan from 2001-2018. Ecology and Evolution **11**:7302-7314.

Sodhi, N. S., and B. W. Brook. 2006. Southeast Asian biodiversity in crisis. Cambridge University Press, Cambridge.

UNEP-WCMC, and IUCN. 2022. Protected Planet: The World Database on Protected Areas (WDPA), Accessed April 2022, Cambridge, UK: UNEP-WCMC and IUCN. Available at: www.protectedplanet.net.

Wasserman, T. N., S. A. Cushman, D. O. Wallin, and J. Hayden. 2012. Multi Scale Habitat Relationships of Martes americana in Northern Idaho, USA. Environmental Sciences Faculty and Staff Publications. 20. https://cedar.wwu.edu/esci_facpubs/20.

WCS, and CIESIN. 2005. Last of The Wild data version 2, (LTW-2): global human footprint dataset (geographic). http://sedac.ciesin.columbia.edu/wildareas/.

Wikramanayake, E. D. 2002. Terrestrial ecoregions of the Indo-Pacific : a conservation assessment. Island Press, Washington D.C.; London.

Yu, H. 2017. Motivation behind China's "One Belt, One Road' Initiatives and Establishment of the Asian Infrastructure Investment Bank. Journal of Contemporary China **26**:353-368.

# Supplementary material for Chapter 5

## Results

### Scales optimisation and covariates selection

We tested four modelling frameworks for each species and chose the best one based on the AUC scores. In Tables S6-S8 we provided the comprehensive sets of representative scales selected for each felid species.

In mainland Southeast Asia, the scale optimisation revealed that most of the covariates had their strongest effects on felids at broad scales. With 42.2% of the covariates selected at 32km, the broadest assessed scale was the most selected one, followed by 16km (17.8% of the covariates) and 8km (13.3% of the covariates). The two smallest scales 250m and 500m were both selected for 8.9% of the covariates. Medium scales were the least selected ones, with 2km selected for 4.4% of the covariates, and 1km and 4km both selected for 2.2% of the covariates.

In Borneo, the scale optimisation revealed that felids selected fine and broad scales at similar rates. The two most selected scales were 500m and 8km, both selected for 20.4% of the covariates, followed by 250m (16.3% of the covariates), 2km (14.3% of the covariates) and 4km (12.2% of the covariates). The broadest assessed scale, 10km, was selected for 10.2% of the covariates, while 1km for 4.1% of the covariates and 6km for 2.0% of the covariates.

In Sumatra, most of the covariates were selected at fine and medium scales. The most selected scale was 250m, being the most representative scale for 26.8% of the covariates, followed by 4km (19.6% of the covariates) and 2km (12.5% of the covariates). Both 500m and 6km were selected for 10.7% of the covariates. Broad scaled were less representative, with 8km being selected for 8.9% of the covariates and 10km for 7.1% of the covariates. Finally, 1km was selected for 3.6% of the covariates.

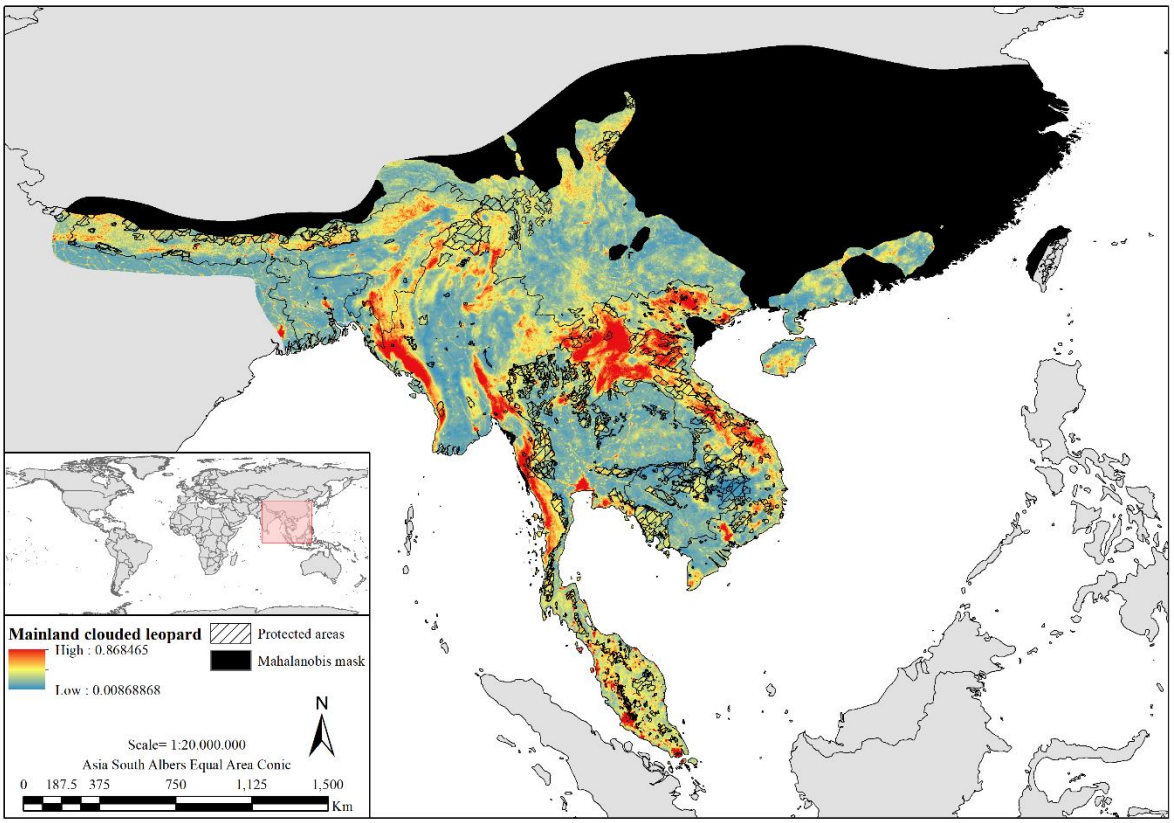



**Figure S1.** Species distribution probability surface of clouded leopard in mainland Southeast Asia.

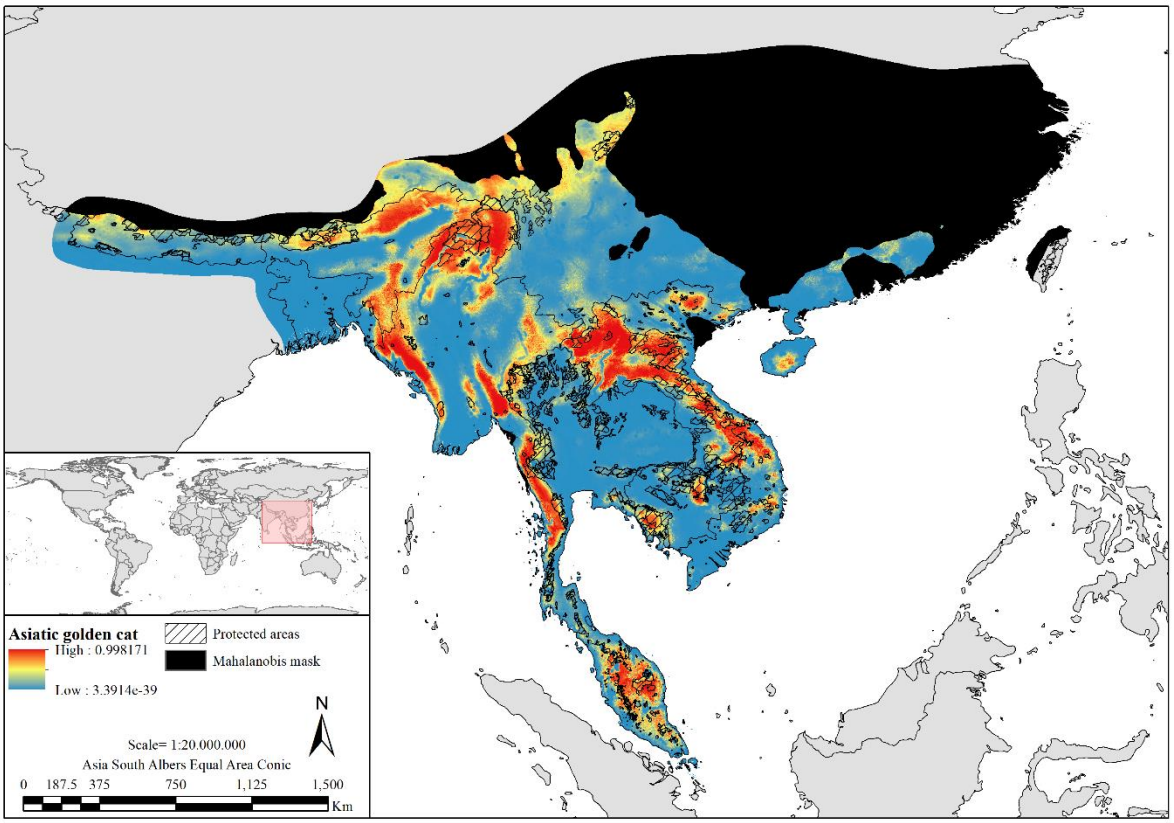



**Figure S2.** Species distribution probability surface of Asiatic golden cat in mainland Southeast Asia.

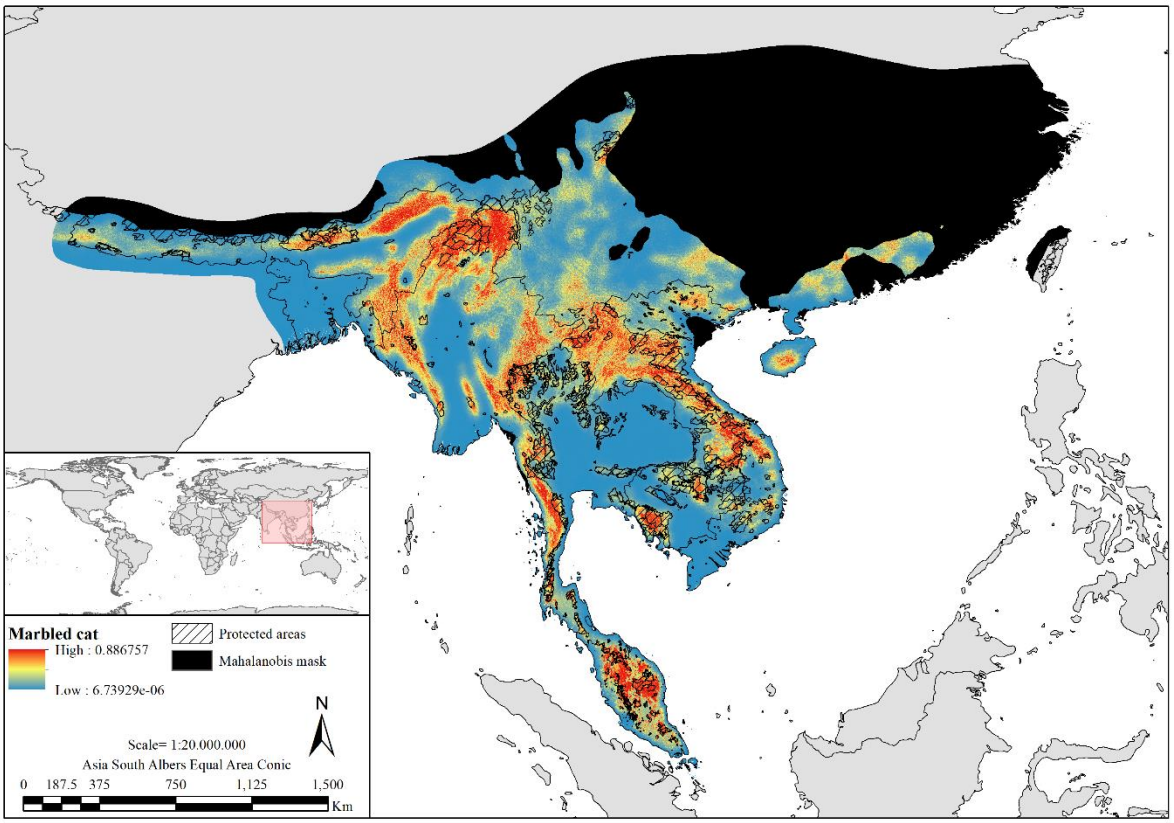



**Figure S3.** Species distribution probability surface of marbled cat in mainland Southeast Asia.

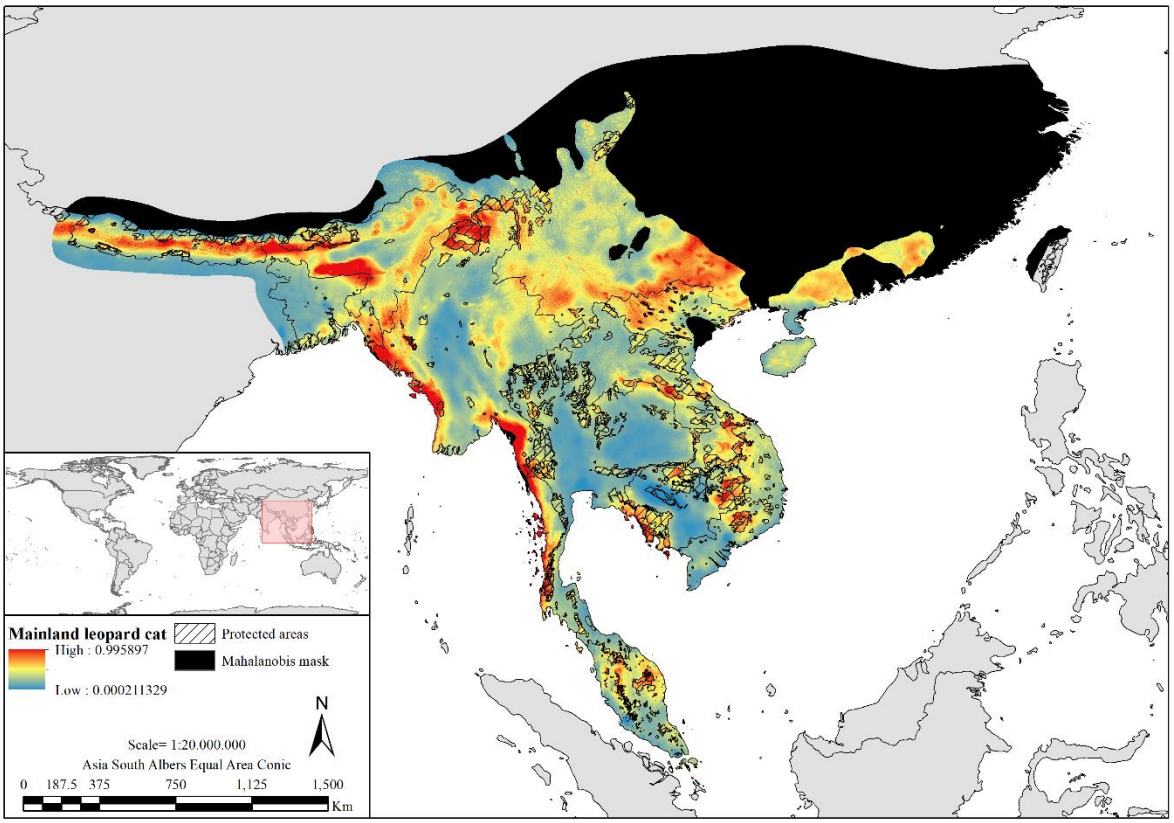



**Figure S4.** Species distribution probability surface of leopard cat in mainland Southeast Asia.

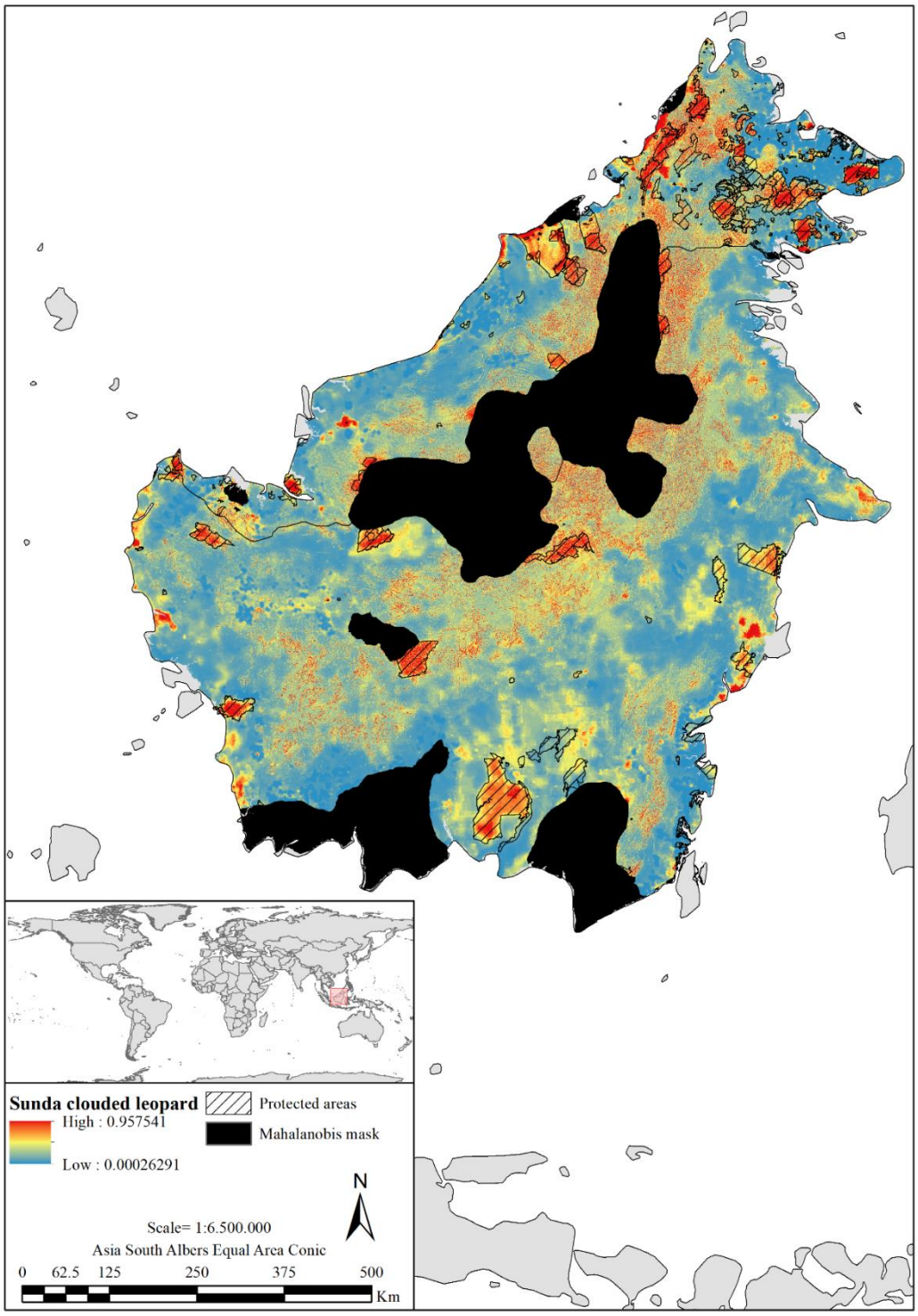



**Figure S5.** Species distribution probability surface of Sunda clouded leopard in Borneo.

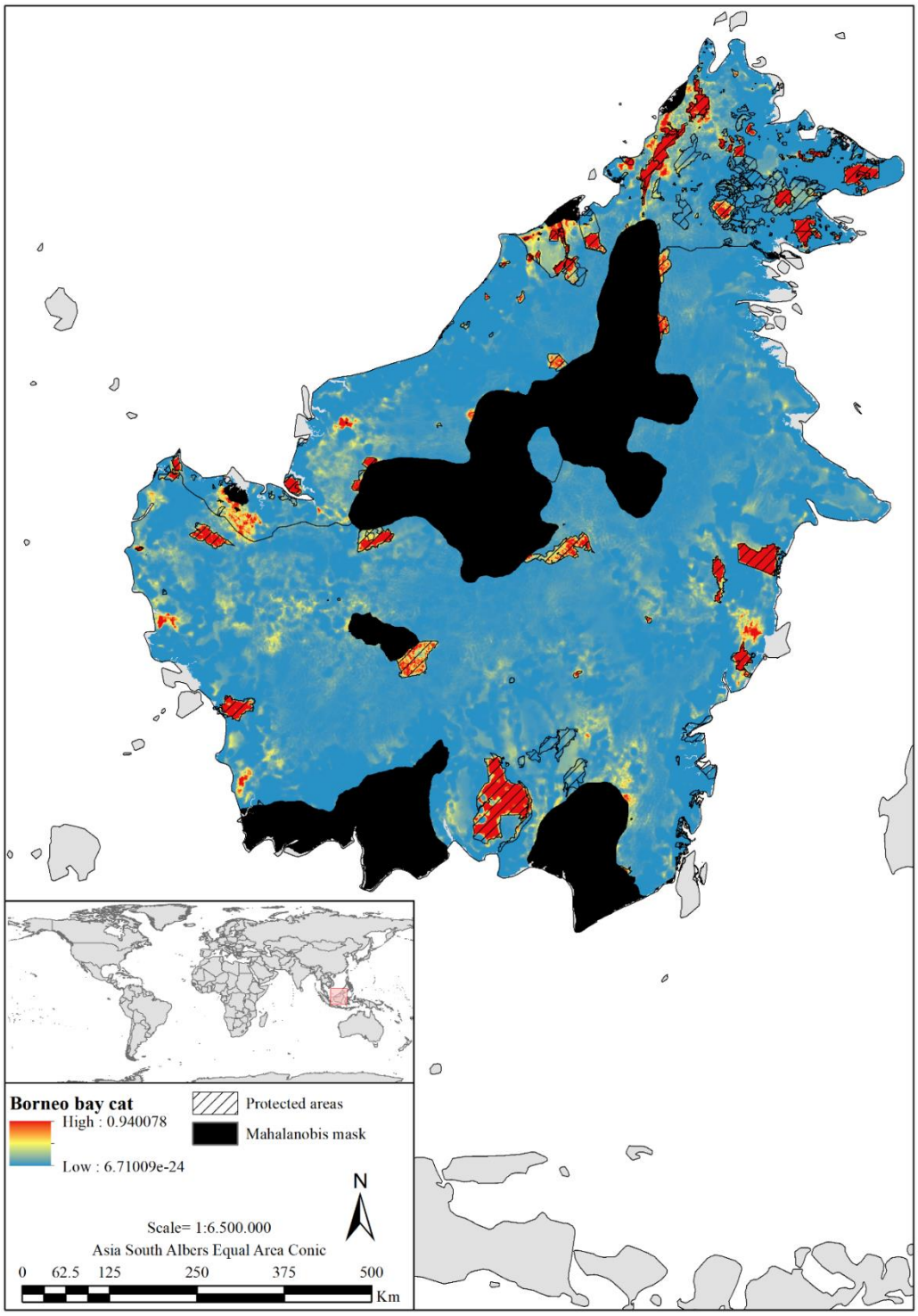



**Figure S6.** Species distribution probability surface of Borneo bay cat in Borneo.

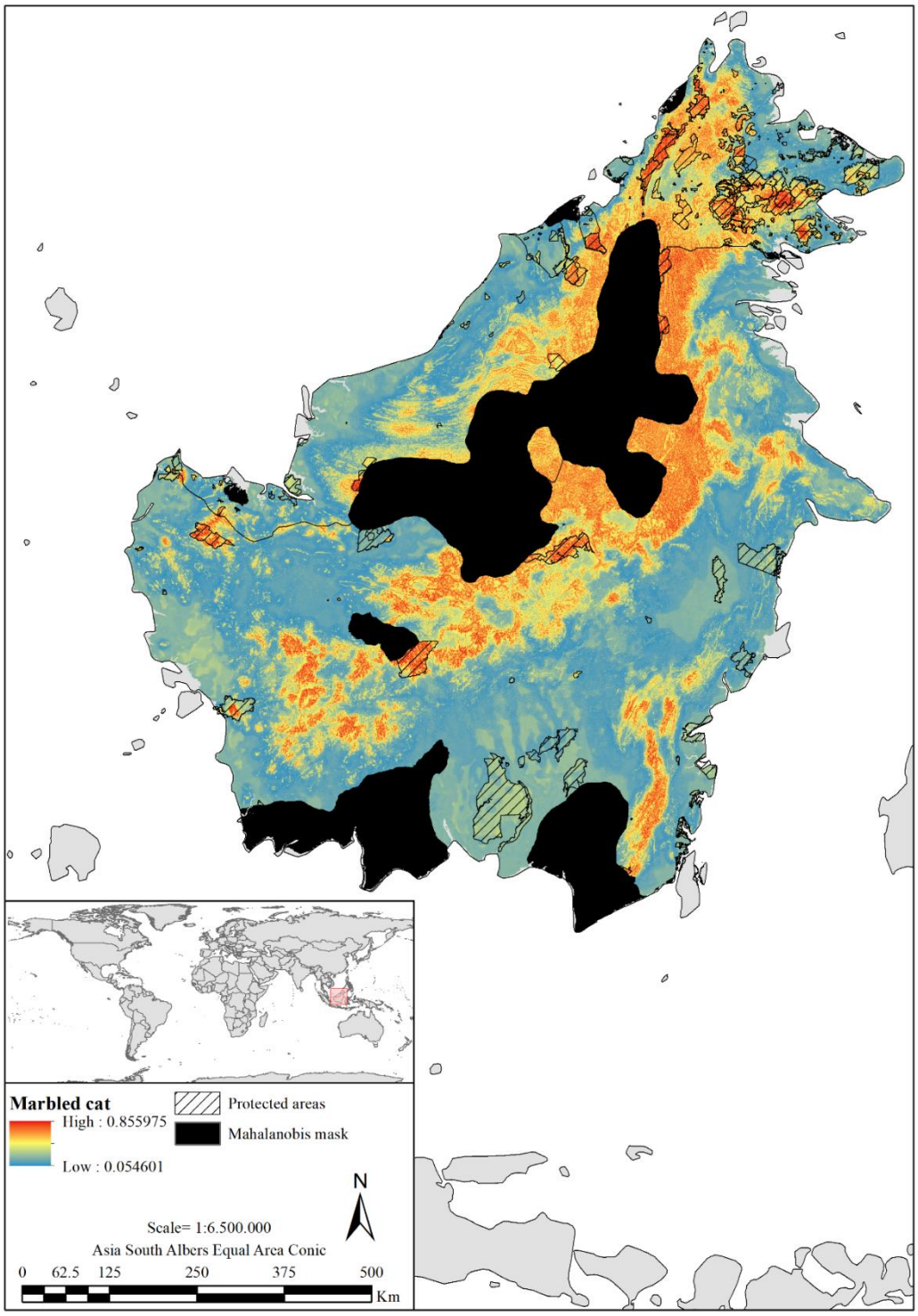



**Figure S7.** Species distribution probability surface of marbled cat in Borneo.

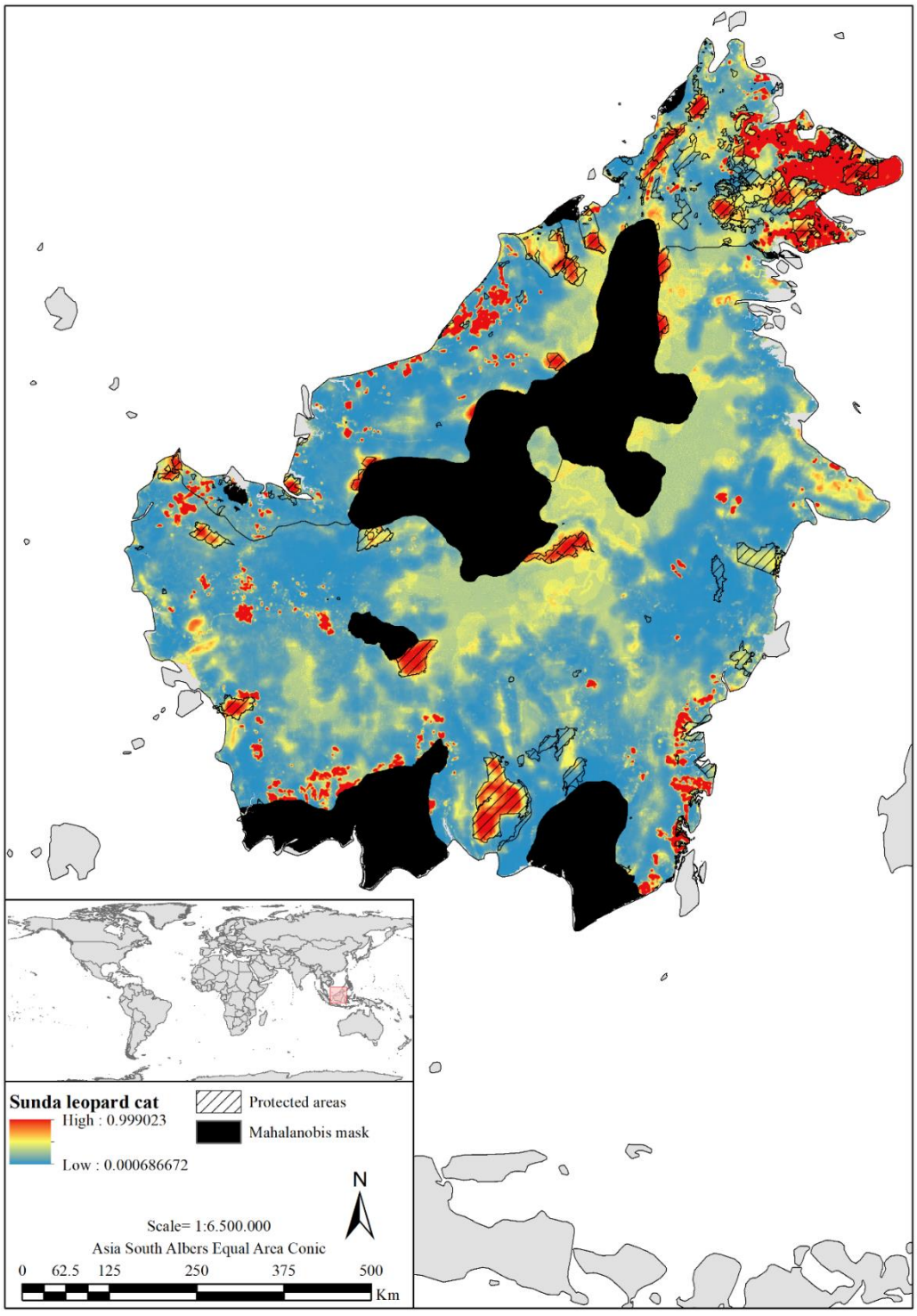



**Figure S8.** Species distribution probability surface of Sunda leopard cat in Borneo.

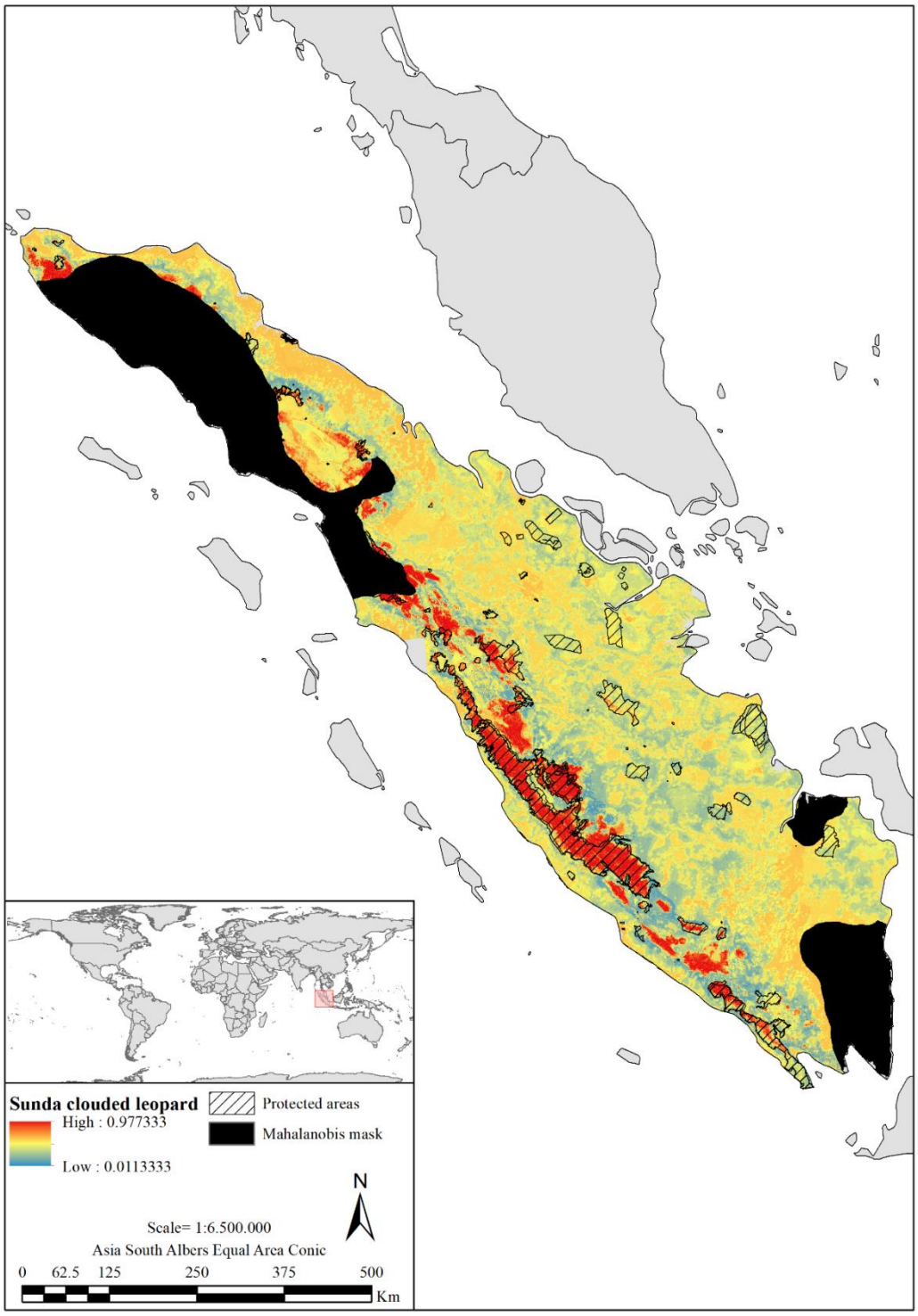



**Figure S9.** Species distribution probability surface of Sunda clouded leopard in Sumatra.

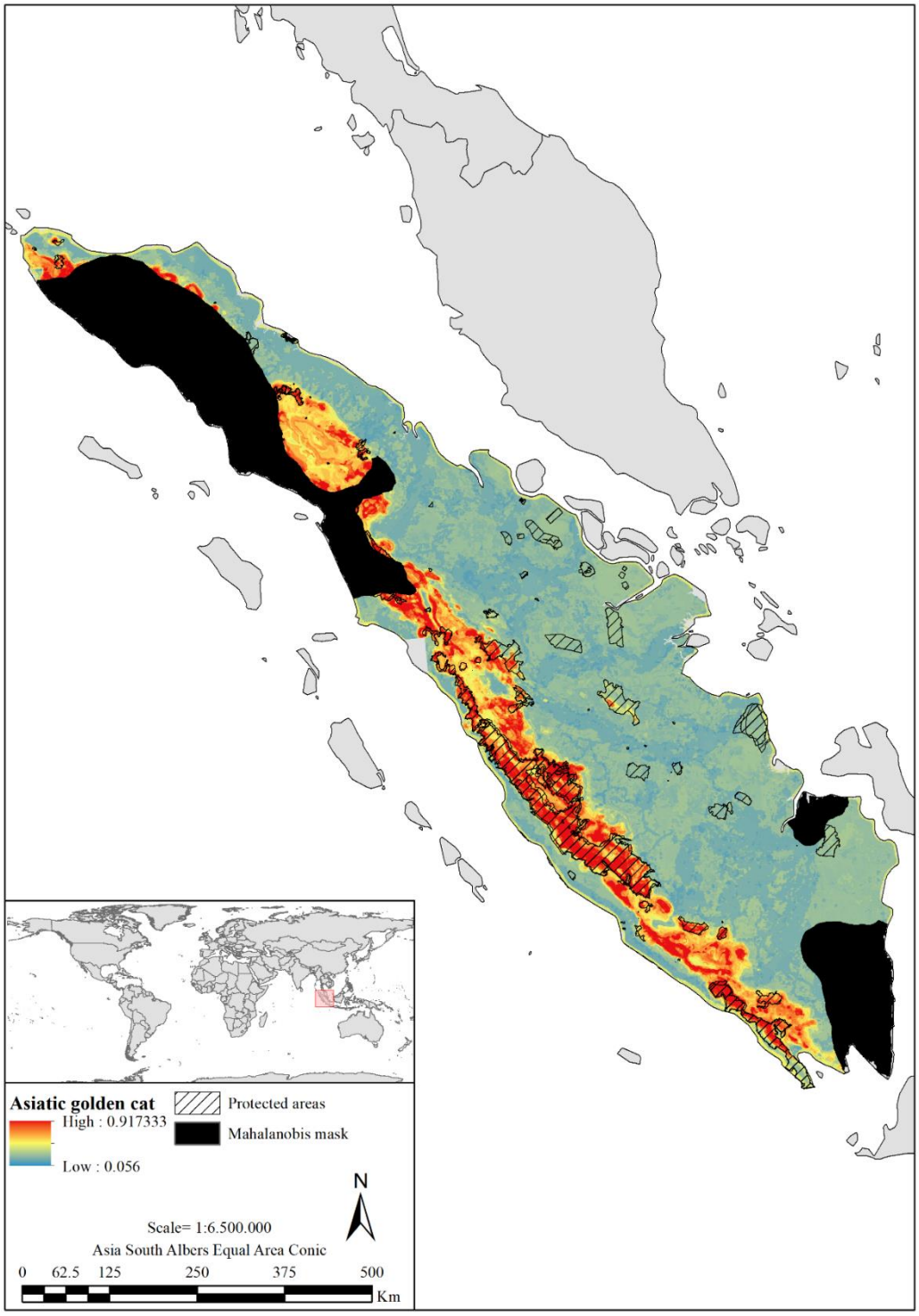



**Figure S10.** Species distribution probability surface of Asiatic golden cat in Sumatra.

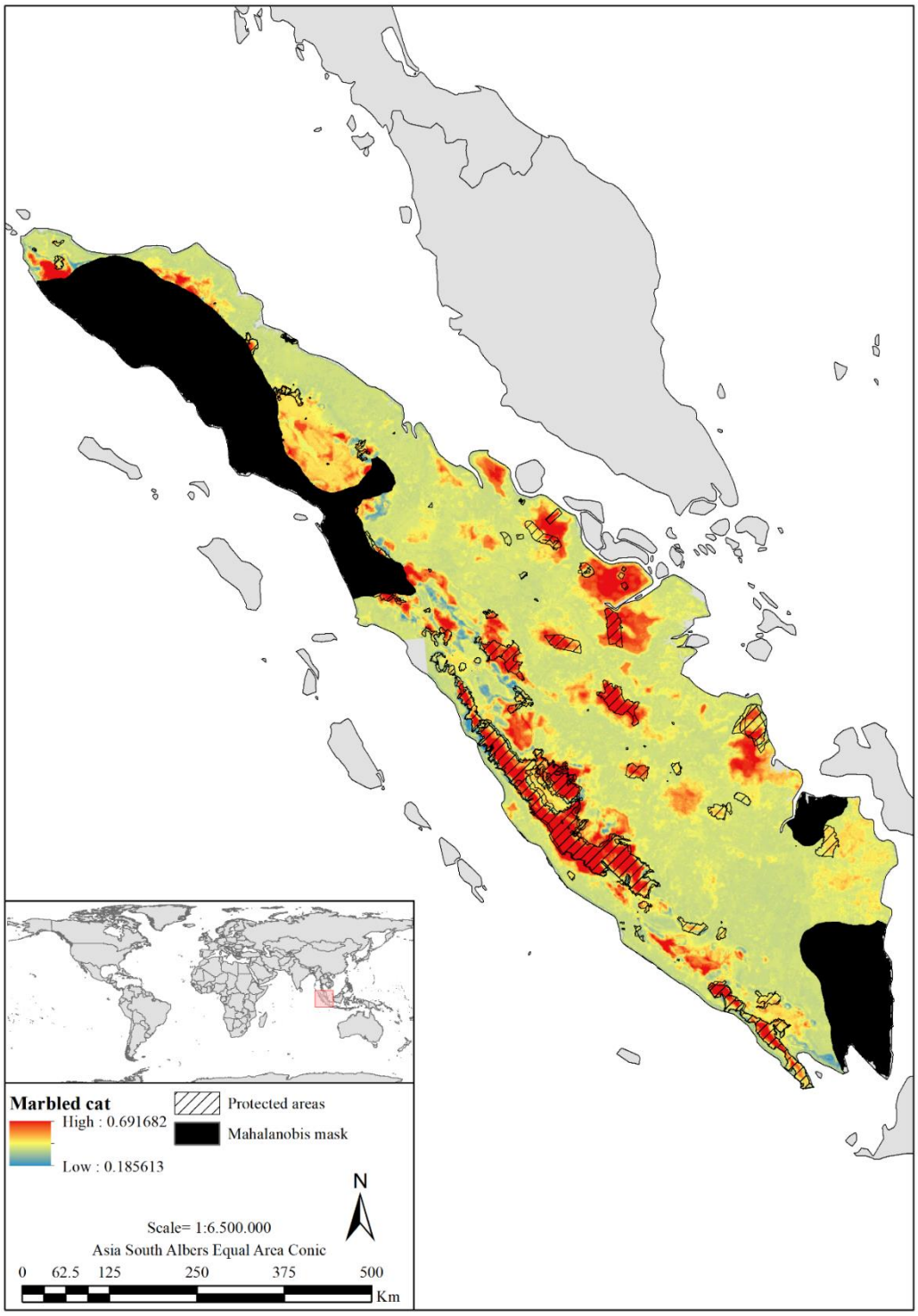



**Figure S11.** Species distribution probability surface of marbled cat in Sumatra.

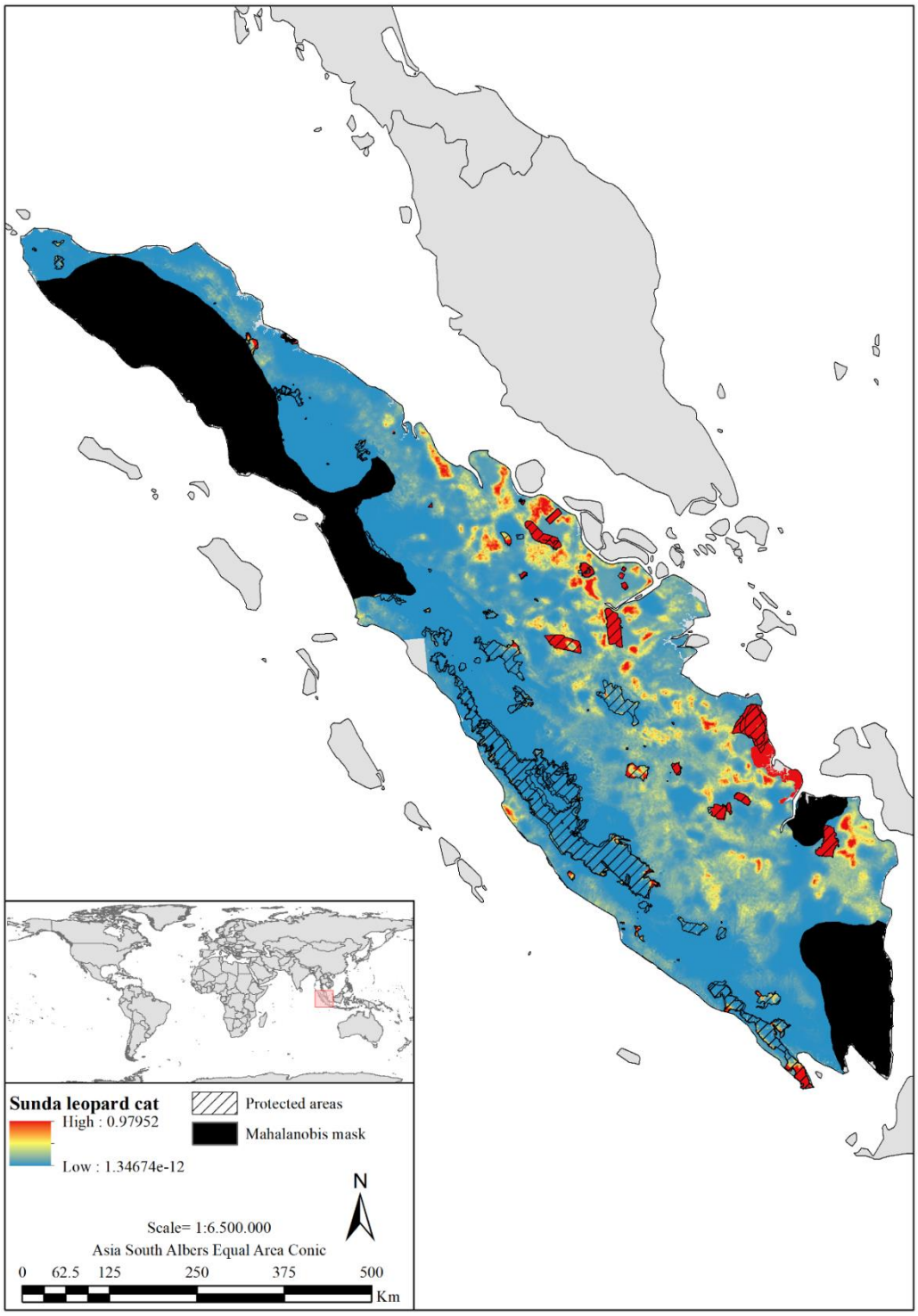



**Figure S12.** Species distribution probability surface of Sunda leopard cat in Sumatra.

**Table S1.** Synoptic table showing country, number of camera trap stations and number of camera trap nights for each sampling location in mainland Southeast Asia. Additionally, we provided the number of detections for each sampled felid.

| Country | Sampling location | Number of camera trap stations | Number of camera trap nights | *Neofelis nebulosa* detections | *Catopuma temminckii* detections | *Pardofelis marmorata* detections | *Prionailurus bengalensis* detections |
|---|---|---|---|---|---|---|---|
| Cambodia | Cardamom | 76 | 7,740 | 10 | 11 | 17 | 29 |
| | PPWS | 77 | 5,283 | 1 | 0 | 1 | 31 |
| India | Dampa | 74 | 4,962 | 28 | 22 | 14 | 25 |
| Laos | NPNV | 78 | 3,729 | 8 | 13 | 10 | 13 |
| | PD | 78 | 3,666 | 15 | 25 | 15 | 15 |
| | PS | 72 | 3,446 | 16 | 7 | 3 | 4 |
| Malaysia (peninsular) | Ulu Muda Grid 1 | 76 | 9,251 | 10 | 0 | 10 | 39 |
| | Ulu Muda Grid 2 | 54 | 5,995 | 11 | 2 | 5 | 21 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Myanmar | Htamanthi | 162 | 14,546 | 67 | 56 | 31 | 45 |
| Nepal | Langtang NP | 83 | 7,105 | 7 | 0 | 0 | 37 |
| Thailand | Dong Yai Wildlife Sanctuary | 58 | 2,709 | 3 | 0 | 0 | 3 |
| | Khao Yai National Park | 67 | 6,266 | 10 | 2 | 1 | 20 |
| | Pang Sida National Park | 163 | 17,750 | 24 | 5 | 3 | 28 |
| | Ta Phraya National Park | 56 | 4,636 | 5 | 0 | 1 | 8 |
| | Thap Lan National Park | 210 | 18,305 | 20 | 5 | 5 | 27 |
| Total | | 1,384 | 115,389 | 235 | 148 | 116 | 345 |

**Table S2.** Synoptic table showing country, number of camera trap stations and number of camera trap nights for each sampling location in Borneo. Additionally, we provided the number of detections for each sampled felid.

| Country | Sampling location | Number of camera trap stations | Number of camera trap nights | *Neofelis diardi* detections | *Catopuma badia* detections | *Pardofelis marmorata* detections | *Prionailurus javanensis* detections |
|---|---|---|---|---|---|---|---|
| Indonesia (Kalimantan) | Bawan | 63 | 4,070 | 0 | 0 | 1 | 1 |
| | Belantikan | 51 | 3,846 | 2 | 0 | 0 | 3 |
| | Kutai | 52 | 3,269 | 1 | 2 | 1 | 0 |
| | Lesan | 73 | 10,057 | 1 | 0 | 1 | 1 |
| | Sungai Wain | 78 | 4,729 | 9 | 1 | 0 | 0 |
| Malaysia (Sabah) | Crocker | 35 | 4,059 | 13 | 0 | 8 | 12 |
| | Danum Palm | 23 | 2,214 | 0 | 0 | 1 | 23 |
| | Danum Valley | 79 | 5,880 | 30 | 1 | 20 | 14 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | IJM | 33 | 1,855 | 0 | 0 | 0 | 26 |
| | Kinabatangan | 68 | 4,450 | 10 | 0 | 4 | 7 |
| | Malua | 38 | 3,867 | 7 | 2 | 3 | 32 |
| | Sepilok | 35 | 2,067 | 0 | 0 | 0 | 6 |
| | Tabin | 74 | 6,877 | 25 | 3 | 11 | 26 |
| | Tawau | 77 | 17,600 | 38 | 14 | 28 | 30 |
| | Ulu Segama | 22 | 2,847 | 18 | 2 | 7 | 21 |
| Total | | 801 | 77,687 | 154 | 25 | 85 | 202 |

**Table S3.** Synoptic table showing country, number of camera trap stations and number of camera trap nights for each sampling location in Sumatra. Additionally, we provided the number of detections for each sampled felid.

| Country | Sampling location | Number of camera trap stations | Number of camera trap nights | *Neofelis diardi* detections | *Catopuma temminckii* detections | *Pardofelis marmorata* detections | *Prionailurus javanensis* detections |
|---|---|---|---|---|---|---|---|
| Indonesia (Sumatra) | Bungo | 79 | 8,405 | 18 | 14 | 7 | 1 |
| | Ipuh | 79 | 6,276 | 13 | 13 | 7 | 2 |
| | Kambang | 149 | 12,599 | 12 | 9 | 2 | 1 |
| | Linggau | 120 | 7,269 | 0 | 2 | 3 | 17 |
| | Muara Hemat | 164 | 13,466 | 18 | 24 | 3 | 1 |
| | RKE | 72 | 5,586 | 27 | 16 | 6 | 1 |
| | Sipurak | 80 | 7,227 | 13 | 24 | 11 | 0 |
| Total | | 743 | 60,828 | 101 | 102 | 39 | 23 |

**Table S4.** Preliminary set of covariates evaluated to model habitat suitability for felids sampled in mainland Southeast Asia.

| Class | Covariate | Source | Metric |
|---|---|---|---|
| Landscape | Bare ground | ESA Land Cover CCI (2017) | PLAND |
| | Croplands | ESA Land Cover CCI (2017) | PLAND |
| | Mosaic areas | ESA Land Cover CCI (2017) | PLAND |
| | Forested areas | ESA Land Cover CCI (2017) | PLAND |
| | Shrublands and grasslands | ESA Land Cover CCI (2017) | PLAND |
| | Permanent snow and ice | ESA Land Cover CCI (2017) | PLAND |
| | Urban areas | ESA Land Cover CCI (2017) | PLAND |
| | Water | ESA Land Cover CCI (2017) | PLAND |
| | Non-forested areas (0-20%) | Hansen et al. (2013) | PLAND |
| | Open forests (20-40%) | Hansen et al. (2013) | PLAND |
| | Closed forests (>40%) | Hansen et al. (2013) | PLAND |
| Anthropic | Forest loss | Hansen et al. (2013) | FM |
| | Human footprint | WCS and CIESIN (2005) | FM |

| | | | |
|---|---|---|---|
| | Human population density | CIESIN and CIAT (2016) | FM |
| | Protected areas | UNEP-WCMC and IUCN (2022) | PLAND |
| Topographic | Elevation | Jarvis et al. (2008) | FM |
| | Compound Topographic Index | Jarvis et al. (2008); Evans et al. (2014) | FM |
| | Slope position | Jarvis et al. (2008); Evans et al. (2014) | FM |
| Climatic | Mean annual precipitation | Fick and Hijmans (2017) | FM |

**Table S5.** Preliminary set of covariates evaluated to model habitat suitability for felids sampled in Borneo and Sumatra.

| Class | Covariate | Source | Metric |
|---|---|---|---|
| Landscape | Lowland forests | Miettinen et al. (2012) | PLAND |
| | Lowland mosaic areas | Miettinen et al. (2012) | PLAND |
| | Lower-montane forests | Miettinen et al. (2012) | PLAND |
| | Lowland open areas | Miettinen et al. (2012) | PLAND |
| | Mangroves | Miettinen et al. (2012) | PLAND |
| | Montane mosaic areas | Miettinen et al. (2012) | PLAND |
| | Montane open areas | Miettinen et al. (2012) | PLAND |
| | Palm plantations | Miettinen et al. (2012) | PLAND |
| | Peat swamp forests | Miettinen et al. (2012) | PLAND |
| | Plantation and regrowth | Miettinen et al. (2012) | PLAND |
| | Upper-montane forests | Miettinen et al. (2012) | PLAND |
| | Urban areas | Miettinen et al. (2012) | PLAND |
| | Water | Miettinen et al. (2012) | PLAND |

| | | | |
|---|---|---|---|
| | Continuous tree canopy cover | Hansen et al. (2013) | FM |
| | Non-forested areas (0-20%) | Hansen et al. (2013) | PLAND |
| | Open forests (20-40%) | Hansen et al. (2013) | PLAND |
| | Closed forests (>40%) | Hansen et al. (2013) | PLAND |
| Anthropic | Human footprint | WCS and CIESIN (2005) | FM |
| | Human population density | CIESIN and CIAT (2016) | FM |
| | Forest loss | Hansen et al. (2013) | PLAND |
| | Protected areas | UNEP-WCMC and IUCN (2022) | PLAND |
| Topographic | Elevation | Jarvis et al. (2008) | FM |
| | Roughness | Jarvis et al. (2008); Evans et al. (2014) | FM |
| | Slope position | Jarvis et al. (2008); Evans et al. (2014) | FM |
| | Compound topographic index | Jarvis et al. (2008); Evans et al. (2014) | FM |

**Table S6.** Most representative spatial scales for felids sampled in mainland Southeast Asia, obtained by applying univariate GLM and univariate RF modelling frameworks.

| | | | GLM best scales (m) | | | | RF best scales (m) | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Class | Covariate | Metric | *Neofelis nebulosa* | *Catopuma temminckii* | *Pardofelis marmorata* | *Prionailurus bengalensis* | *Neofelis nebulosa* | *Catopuma temminckii* | *Pardofelis marmorata* | *Prionailurus bengalensis* |
| Landscape | Bare ground | PLAND | 32,000 | 32,000 | 32,000 | 32,000 | 32,000 | 32,000 | 16,000 | 32,000 |
| | Croplands | PLAND | 16,000 | 16,000 | 16,000 | 32,000 | 4,000 | 8,000 | 4,000 | 4,000 |
| | Mosaic areas | PLAND | 8,000 | 8,000 | 250 | 32,000 | 250 | 250 | 250 | 250 |
| | Forested areas | PLAND | 32,000 | 32,000 | 32,000 | 32,000 | 250 | 250 | 250 | 250 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Shrublands and grasslands | PLAND | 16,000 | 32,000 | 1,000 | 2,000 | 250 | 250 | 250 | 250 |
| | Urban areas | PLAND | 32,000 | 32,000 | 32,000 | 16,000 | 16,000 | 16,000 | 16,000 | 16,000 |
| | Water | PLAND | 16,000 | 8,000 | 16,000 | 16,000 | 8,000 | 8,000 | 8,000 | 8,000 |
| | Non-forested areas | PLAND | 32,000 | 32,000 | 32,000 | 32,000 | 1,000 | 8,000 | 1,000 | 1,000 |
| | Open forests | PLAND | 16,000 | 32,000 | 32,000 | 32,000 | 500 | 4,000 | 500 | 500 |
| | Closed forests | PLAND | 16,000 | 32,000 | 32,000 | 32,000 | 500 | 500 | 500 | 250 |
| Anthropic | Forest loss | FM | 500 | 500 | 16,000 | 8,000 | 500 | 500 | 250 | 250 |
| | Human footprint | FM | 500 | 500 | 1,000 | 32,000 | 250 | 250 | 250 | 250 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Human population density | FM | 500 | 500 | 500 | 2,000 | 250 | 250 | 250 | 250 |
| | Protected areas | PLAND | 32,000 | 500 | 250 | 8,000 | 250 | 250 | 250 | 250 |
| Topographic | Elevation | FM | 16,000 | 32,000 | 32,000 | 32,000 | 500 | 1,000 | 2,000 | 32,000 |
| | Compound topographic index | FM | 500 | 32,000 | 250 | 32,000 | 16,000 | 2,000 | 500 | 16,000 |
| | Slope position | FM | 8,000 | 8,000 | 250 | 4,000 | 1,000 | 1,000 | 250 | 16,000 |
| Climatic | Mean annual precipitation | FM | 1,000 | 32,000 | 32,000 | 8,000 | 32,000 | 1,000 | 250 | 2,000 |

**Table S7.** Most representative spatial scales for felids sampled in Borneo, obtained by applying univariate GLM and univariate RF modelling frameworks.

| | | | GLM best scales (m) | | | | RF best scales (m) | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Class | Covariate | Metric | *Neofelis diardi* | *Catopuma badia* | *Pardofelis marmorata* | *Prionailurus javanensis* | *Neofelis diardi* | *Catopuma badia* | *Pardofelis marmorata* | *Prionailurus javanensis* |
| Landscape | Lowland forests | PLAND | 8,000 | 4,000 | 8,000 | 8,000 | 250 | 500 | 500 | 250 |
| | Lowland mosaic areas | PLAND | 8,000 | 8,000 | 8,000 | 8,000 | 250 | 500 | 500 | 250 |
| | Lower-montane forests | PLAND | 1,000 | 1,000 | 6,000 | 4,000 | 8,000 | 1,000 | 2,000 | 8,000 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Lowland open areas | PLAND | 10,000 | 4,000 | 10,000 | 10,000 | 500 | 2,000 | 500 | 4,000 |
| Mangroves | PLAND | 8,000 | 8,000 | NA⁂ | 6,000 | 4,000 | 4,000 | NA⁂ | 10,000 |
| Montane mosaic areas | PLAND | NA* | NA* | NA* | NA* | NA* | NA* | NA* | NA* |
| Montane open areas | PLAND | NA* | NA* | NA* | NA* | NA* | NA* | NA* | NA* |
| Palm plantations | PLAND | 2,000 | 4,000 | 2,000 | 2,000 | 2,000 | 2,000 | 2,000 | 2,000 |
| Plantation and regrowth | PLAND | 6,000 | 250 | 4,000 | 8,000 | 500 | 500 | 500 | 250 |
| Upper-montane forests | PLAND | NA* | NA* | NA* | NA* | NA* | NA* | NA* | NA* |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Water | PLAND | 10,000 | NA‡ | 4,000 | 10,000 | 4,000 | NA‡ | 4,000 | 4,000 |
| | Continuous tree canopy cover | FM | 6,000 | 2,000 | 6,000 | 1,000 | 6,000 | 250 | 4,000 | 1,000 |
| | Non-forested areas | PLAND | 6,000 | 4,000 | 8,000 | 1,000 | 500 | 1,000 | 500 | 1,000 |
| | Open forests | PLAND | 8,000 | 4,000 | 8,000 | 10,000 | 1,000 | 4,000 | 4,000 | 10,000 |
| | Closed forests | PLAND | 6,000 | 4,000 | 8,000 | 1,000 | 1,000 | 1,000 | 500 | 1,000 |
| Anthropic | Human footprint | FM | 250 | 10,000 | 10,000 | 500 | 250 | 250 | 250 | 250 |
| | Human population density | FM | 6,000 | 250 | 10,000 | 250 | 500 | 4,000 | 2,000 | 4,000 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Forest loss | PLAND | 8,000 | 4,000 | 10,000 | 8,000 | 500 | 2,000 | 500 | 500 |
| | Protected areas | PLAND | 8,000 | 4,000 | 8,000 | 10,000 | 250 | 1,000 | 500 | 1,000 |
| Topographic | Elevation | FM | 250 | 250 | 250 | 10,000 | 10,000 | 8,000 | 8,000 | 10,000 |
| | Roughness | FM | 1,000 | 2,000 | 1,000 | 10,000 | 8,000 | 4,000 | 8,000 | 6,000 |
| | Slope position | FM | 250 | 8,000 | 250 | 250 | 250 | 6,000 | 250 | 10,000 |
| | Compound topographic index | FM | 500 | 2,000 | 500 | 10,000 | 1,000 | 2,000 | 1,000 | 6,000 |

* Covariates excluded from the models because occurring in <10% of camera trap stations.

⁑ Covariates excluded because the univariate models did not converge.

**Table S8.** Most representative spatial scales for felids sampled in Sumatra, obtained by applying univariate GLM and univariate RF modelling frameworks.

| | | | GLM best scales (m) | | | | RF best scales (m) | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Class | Covariate | Metric | *Neofelis diardi* | *Catopuma temminckii* | *Pardofelis marmorata* | *Prionailurus javanensis* | *Neofelis diardi* | *Catopuma temminckii* | *Pardofelis marmorata* | *Prionailurus javanensis* |
| Landscape | Lowland forests | PLAND | 4,000 | 6,000 | 500 | 6,000 | 500 | 500 | 500 | 500 |
| | Lowland mosaic areas | PLAND | 8,000 | 10,000 | 8,000 | 4,000 | 2,000 | 2,000 | 4,000 | 2,000 |
| | Lower-montane forests | PLAND | 250 | 6,000 | 2,000 | 500 | 250 | 250 | 500 | 500 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Lowland open areas | PLAND | 6,000 | 10,000 | 6,000 | 4,000 | 4,000 | 4,000 | 6,000 | 6,000 |
| Mangroves | PLAND | NA* | NA* | NA* | NA* | NA* | NA* | NA* | NA* |
| Montane mosaic areas | PLAND | 2,000 | 10,000 | 6,000 | 10,000 | 2,000 | 2,000 | 4,000 | 6,000 |
| Montane open areas | PLAND | 8,000 | 8,000 | 8,000 | 10,000 | 6,000 | 10,000 | 6,000 | 8,000 |
| Palm plantations | PLAND | 8,000 | 10,000 | 8,000 | NA⁂ | 8,000 | 8,000 | 8,000 | NA⁂ |
| Plantation and regrowth | PLAND | 1,000 | 2,000 | 2,000 | 2,000 | 250 | 500 | 500 | 500 |
| Upper-montane forests | PLAND | 6,000 | 10,000 | 10,000 | 10,000 | 4,000 | 4,000 | 4,000 | 4,000 |
| Water | PLAND | NA* | NA* | NA* | NA* | NA* | NA* | NA* | NA* |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Continuous tree canopy cover | FM | 500 | 1,000 | 6,000 | 2,000 | 500 | 2,000 | 6,000 | 6,000 |
| | Non-forested areas | PLAND | 2,000 | 4,000 | 6,000 | 10,000 | 2,000 | 500 | 2,000 | 2,000 |
| | Open forests | PLAND | 6,000 | 8,000 | 8,000 | 10,000 | 8,000 | 2,000 | 4,000 | 4,000 |
| | Closed forests | PLAND | 6,000 | 4,000 | 6,000 | 10,000 | 1,000 | 1,000 | 2,000 | 2,000 |
| Anthropic | Human footprint | FM | 10,000 | 10,000 | 6,000 | 10,000 | 250 | 250 | 250 | 500 |
| | Human population density | FM | 500 | 1,000 | 10,000 | 10,000 | 250 | 250 | 250 | 250 |
| | Forest loss | PLAND | 6,000 | 4,000 | 4,000 | 4,000 | 500 | 1,000 | 1,000 | 1,000 |
| | Protected areas | PLAND | 4,000 | 10,000 | 250 | 250 | 250 | 500 | 250 | 250 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Topographic | Elevation | FM | 250 | 8,000 | 4,000 | 250 | 4,000 | 4,000 | 8,000 | 2,000 |
| | Roughness | FM | 4,000 | 10,000 | 10,000 | 10,000 | 250 | 4,000 | 6,000 | 10,000 |
| | Slope position | FM | 8,000 | 250 | 1,000 | 2,000 | 250 | 250 | 1,000 | 4,000 |
| | Compound topographic index | FM | 500 | 250 | 1,000 | 250 | 250 | 8,000 | 6,000 | 2,000 |

* Covariates excluded from the models because occurring in <10% of camera trap stations.

⁑ Covariates excluded because the univariate models did not converge.

**Table S9.** Best scale and coefficient of the covariates included in the final model for clouded leopard in mainland Southeast Asia.

| Covariate | Best scale (m) | Coefficient |
|---|---|---|
| Intercept | NA | -2.89 |
| Camera trap nights | NA | 0.01 |
| Slope position | 8,000 | 0.14 |
| Compound topographic index | 500 | -0.30 |
| Water | 16,000 | -0.06 |
| Mosaic areas | 8,000 | -0.27 |
| Shrublands and grasslands | 16,000 | 0.40 |
| Protected areas | 32,000 | -0.10 |
| Urban areas | 32,000 | 0.08 |
| Forested areas | 32,000 | -0.13 |
| Croplands | 16,000 | -0.21 |
| Human footprint | 500 | 0.61 |
| Closed forests | 16,000 | 0.82 |

**Table S10.** Best scale and coefficient of the covariates included in the final model for Asiatic golden cat in mainland Southeast Asia.

| Covariate | Best scale (m) | Coefficient |
|---|---|---|
| Intercept | NA | -4.84 |
| Camera trap nights | NA | 0.01 |
| Slope position | 8,000 | 0.08 |
| Protected areas | 500 | -0.23 |
| Compound topographic index | 32,000 | -0.02 |
| Shrublands and grasslands | 32,000 | 1.61 |
| Water | 8,000 | -6.04 |
| Forested areas | 32,000 | 1.43 |
| Forest loss | 500 | 0.49 |
| Urban areas | 32,000 | -0.47 |
| Open forests | 32,000 | -0.69 |
| Closed forests | 32,000 | 0.63 |

**Table S11.** Best scale and coefficient of the covariates included in the final model for marbled cat in mainland Southeast Asia.

| Covariate | Best scale (m) | Coefficient |
|---|---|---|
| Intercept | NA | -3.75 |
| Camera trap nights | NA | 0.00 |
| Protected areas | 250 | -0.04 |
| Slope position | 250 | 0.20 |
| Shrublands and grasslands | 1,000 | -0.20 |
| Compound topographic index | 250 | -0.20 |
| Water | 16,000 | -0.31 |
| Mosaic areas | 250 | -0.50 |
| Forest loss | 16,000 | 0.08 |
| Forested areas | 32,000 | 0.01 |
| Urban areas | 32,000 | -0.31 |
| Open forests | 32,000 | 0.32 |
| Closed forests | 32,000 | 1.56 |

**Table S12.** Best scale and coefficient of the covariates included in the final model for leopard cat in mainland Southeast Asia.

| Covariate | Best scale (m) | Coefficient |
|---|---|---|
| Intercept | NA | -2.43 |
| Camera trap nights | NA | 0.01 |
| Urban areas | 16,000 | -0.20 |
| Protected areas | 8,000 | 0.20 |
| Water | 16,000 | 0.15 |
| Human population density | 2,000 | -0.01 |
| Slope position | 4,000 | 0.14 |
| Bare ground | 32,000 | -0.22 |
| Open forests | 32,000 | 0.57 |
| Shrublands and grasslands | 2,000 | -0.11 |
| Compound topographic index | 32,000 | -0.73 |
| Mosaic areas | 32,000 | 0.02 |
| Forested areas | 32,000 | 0.16 |
| Mean annual precipitation | 8,000 | 0.72 |
| Croplands | 32,000 | 0.55 |

**Table S13.** Best scale and coefficient of the covariates included in the final model for Sunda clouded leopard in Borneo.

| Covariate | Best scale (m) | Coefficient |
|---|---|---|
| Intercept | NA | -5.64 |
| Camera trap nights | NA | 0.01 |
| Human footprint | 250 | 0.68 |
| Mangroves | 8,000 | -4.02 |
| Palm plantations | 2,000 | -6.50 |
| Plantation and regrowth | 6,000 | -0.58 |
| Lowland open areas | 10,000 | -0.24 |
| Slope position | 250 | 0.60 |
| Protected areas | 8,000 | 0.35 |
| Forest loss | 8,000 | -0.70 |
| Elevation | 250 | -0.11 |
| Compound topographic index | 500 | -0.31 |

**Table S14.** Best scale and coefficient of the covariates included in the final model for Borneo bay cat in Borneo.

| Covariate | Best scale (m) | Coefficient |
|---|---|---|
| Intercept | NA | -6.19 |
| Camera trap nights | NA | 0.01 |
| Plantation and regrowth | 250 | 0.23 |
| Human footprint | 10,000 | 0.96 |
| Palm plantations | 4,000 | -1.08 |
| Lowland mosaic areas | 8,000 | 0.35 |
| Mangroves | 8,000 | -2.32 |
| Slope position | 8,000 | -0.25 |
| Continuous tree canopy cover | 2,000 | 0.40 |
| Compound topographic index | 2,000 | 0.25 |
| Forest loss | 4,000 | -2.09 |
| Protected areas | 4,000 | 0.94 |

**Table S15.** Best scale and mean decrease accuracy of the covariates included in the final model for marbled cat in Borneo.

| Covariate | Best scale (m) | Mean decrease accuracy |
|---|---|---|
| Camera trap nights | NA | 5.34 |
| Closed forests | 500 | -0.17 |
| Compound topographic index | 1,000 | 7.53 |
| Elevation | 8,000 | 5.05 |
| Human footprint | 250 | 3.54 |
| Lowland forests | 500 | 0.73 |
| Lowland mosaic areas | 500 | 0.89 |
| Lower-montane forests | 2,000 | 0.21 |
| Lowland open areas | 500 | -0.47 |
| Forest loss | 500 | 0.64 |
| Non-forested areas | 500 | -0.29 |
| Open forests | 4,000 | 2.98 |
| Palm plantations | 2,000 | -0.64 |
| Human population density | 2,000 | 2.14 |
| Plantation and regrowth | 500 | -0.08 |
| Protected areas | 500 | 6.92 |
| Roughness | 8,000 | 4.28 |
| Slope position | 250 | 8.10 |
| Continuous tree canopy cover | 4,000 | 4.39 |
| Water | 4,000 | -0.60 |

**Table S16.** Best scale and coefficient of the covariates included in the final model for Sunda leopard cat in Borneo.

| Covariate | Best scale (m) | Coefficient |
|---|---|---|
| Intercept | NA | -3.22 |
| Camera trap nights | NA | 0.01 |
| Slope position | 250 | 0.13 |
| Human population density | 250 | -0.13 |
| Closed forests | 1,000 | -0.38 |
| Human footprint | 500 | 0.59 |
| Open forests | 10,000 | -0.48 |
| Forest loss | 8,000 | 0.51 |
| Lowland open areas | 10,000 | -2.48 |
| Palm plantations | 2,000 | 1.53 |
| Protected areas | 10,000 | 0.70 |
| Plantation and regrowth | 8,000 | -1.27 |

**Table S17.** Best scale and mean decrease accuracy of the covariates included in the final model for Sunda clouded leopard in Sumatra.

| Covariate | Best scale (m) | Mean decrease accuracy |
|---|---|---|
| Camera trap nights | NA | 2.91 |
| Compound topographic index | 250 | 1.49 |
| Elevation | 4,000 | 12.85 |
| Human footprint | 250 | 3.93 |
| Lowland forests | 500 | 6.03 |
| Lower-montane forests | 250 | 11.09 |
| Lowland open areas | 4,000 | 1.41 |
| Montane mosaic areas | 2,000 | 2.46 |
| Non-forested areas | 2,000 | 4.97 |
| Open forests | 8,000 | 10.45 |
| Palm plantations | 8,000 | 7.48 |
| Human population density | 250 | 3.34 |
| Plantation and regrowth | 250 | 12.35 |
| Roughness | 250 | 4.58 |
| Slope position | 250 | 6.24 |
| Continuous tree canopy cover | 500 | 12.08 |
| Upper-montane forests | 4,000 | 3.77 |

**Table S18.** Best scale and mean decrease accuracy of the covariates included in the final model for Asiatic golden cat in Sumatra.

| Covariate | Best scale (m) | Mean decrease accuracy |
|---|---|---|
| Camera trap nights | NA | 11.04 |
| Compound topographic index | 8,000 | 4.54 |
| Elevation | 4,000 | 15.00 |
| Human footprint | 250 | 3.47 |
| Human population density | 250 | 0.55 |
| Roughness | 4,000 | 3.99 |
| Continuous tree canopy cover | 2,000 | 4.80 |

**Table S19.** Best scale and mean decrease accuracy of the covariates included in the final model for marbled cat in Sumatra.

| Covariate | Best scale (m) | Mean decrease accuracy |
|---|---|---|
| Camera trap nights | NA | 2.07 |
| Closed forests | 2,000 | 0.45 |
| Compound topographic index | 6,000 | 1.22 |
| Elevation | 8,000 | 0.67 |
| Human footprint | 250 | 0.65 |
| Lowland forests | 500 | -0.53 |
| Lowland mosaic areas | 4,000 | -0.08 |
| Lower-montane forests | 500 | 0.76 |
| Lowland open areas | 6,000 | 1.47 |
| Forest loss | 1,000 | 0.61 |
| Montane mosaic areas | 4,000 | 0.14 |
| Montane open areas | 6,000 | 0.08 |
| Non-forested areas | 2,000 | -0.16 |
| Open forests | 4,000 | 1.32 |
| Palm plantations | 8,000 | 0.91 |
| Human population density | 250 | -0.31 |
| Plantation and regrowth | 500 | 1.84 |
| Protected areas | 250 | 1.06 |
| Roughness | 6,000 | 0.34 |
| Slope position | 1,000 | -0.53 |

| | | |
|---|---|---|
| Continuous tree canopy cover | 6,000 | 1.91 |
| Upper-montane forests | 4,000 | 0.30 |

**Table S20.** Best scale and coefficient of the covariates included in the final model for Sunda leopard cat in Sumatra.

| Covariate | Best scale (m) | Coefficient |
|---|---|---|
| Intercept | NA | -8.29 |
| Camera trap nights | NA | 0.03 |
| Lowland forests | 6,000 | -0.66 |
| Lowland open areas | 4,000 | 0.24 |
| Upper-montane forests | 10,000 | -0.22 |
| Montane mosaic areas | 10,000 | 0.29 |
| Protected areas | 250 | 1.43 |
| Plantation and regrowth | 2,000 | 0.37 |
| Non-forested areas | 10,000 | -1.47 |
| Roughness | 10,000 | -0.67 |
| Forest loss | 4,000 | 0.47 |
| Lower-montane forests | 500 | -1.40 |
| Slope position | 2,000 | -0.55 |
| Compound topographic index | 250 | 0.30 |
| Elevation | 250 | -2.60 |

**References**

CIESIN, and CIAT. 2016. Gridded Population of the World version 4 (GPWv4): Population density grids. Socioeconomic Data and Applications Center (SEDAC), Columbia University, Palisades, NY. http://sedac.ciesin.columbia.edu/gpw.

ESA Land Cover CCI. 2017. Product user guide version 2.0. http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf.

Evans, J. S., J. Oakleaf, S. A. Cushman, and D. Theobald. 2014. An ArcGIS toolbox for surface gradient and geomorphometric modeling, version 2.0-0. Available: http://evansmurphy.wix.com/evansspatial.

Fick, S. E., and R. J. Hijmans. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology **37**:4302-4315.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science **342**:850-853.

Jarvis, A., H. I. Reuter, A. Nelson, and E. Guevara. 2008. Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT). http://srtm.csi.cgiar.org.

Miettinen, J., C. H. Shi, W. J. Tan, and S. C. Liew. 2012. 2010 land cover map of insular Southeast Asia in 250-m spatial resolution. Remote Sensing Letters **3**:11-20.

UNEP-WCMC, and IUCN. 2022. Protected Planet: The World Database on Protected Areas (WDPA), Accessed April 2022, Cambridge, UK: UNEP-WCMC and IUCN. Available at: www.protectedplanet.net.

WCS, and CIESIN. 2005. Last of The Wild data version 2, (LTW-2): global human footprint dataset (geographic). http://sedac.ciesin.columbia.edu/wildareas/.

# Chapter 6

# General discussion

In this chapter, I will first summarise the main findings from each of the previous data chapters, which I will then integrate in a final paragraph highlighting the main general conclusions.

**Chapter 2.** Predicting biodiversity richness in rapidly changing landscapes: climate, low human pressure or protection as salvation?

In this chapter, I have first highlighted the most important habitat factors driving species richness in the mainland extents of Southeast Asia, showing that different assemblages of species are distinctly affected by diverse groups of habitat factors. Then, building on the modelled multivariate relationships between the sampled terrestrial vertebrate species and the major habitat factors, I mapped the areas characterised by the highest richness of species. Finally, I evaluated the effectiveness of the existing system of protected areas in preserving the modelled terrestrial vertebrate biodiversity. I highlighted that, for Southeast Asian vertebrate terrestrial biodiversity, environmental factors such as presence of forested areas and of water are the most important drivers of patterns of species richness regionally. However, elevation and latitude were also extremely influential in shaping the distribution of some of the sampled species. This was not surprising, considering the rich literature available on the species richness along altitudinal and latitudinal gradients, driven by combinations of different climatic and ecological conditions that, together, promote habitat heterogeneity and consequent availability of several ecological niches that can be occupied by a higher number of species (Willig and Presley 2016, Cuesta et al. 2017).

By mapping the geographic regions harbouring the highest number of species, I revealed that biodiversity was concentrated in two primary areas, namely the mountainous

region of Southwest China, in the northern extent of the study area, and the Thai-Malay Peninsula, in the southern extent. Embedded within a less suitable matrix, a scattering of smaller biodiverse areas were predicted in the central regions of Southeast Asia. A similar geographic pattern of the distribution of terrestrial vertebrate biodiversity was highlighted, using different methods, by Hughes (2017a), and it is strongly driven by environmental and climatic heterogeneity. In the mountainous region of Southwest China, landscapes are characterised by ridge-valley complexes shaped by rivers and divided by tall mountains, driving pronounced environmental and climatic heterogeneity that, in turn, promote ecological niche diversity, allowing a variety of different species to occur in the area (Wikramanayake 2002). Similarly, in the Thai-Malay Peninsula, the highly productive and humid tropic rainforest climate facilitates high levels of biodiversity (Wikramanayake 2002).

While forested landscapes in areas that had little human disturbance consistently were predicted to support high richness in species, anthropic landscapes characterised by high levels of human footprint were the predictor of low biodiversity. This result was strongly corroborated by the analyses I carried out, which showed that, besides climatic factors, human footprint had the strongest negative effect on modelled number of species, while protected areas had the strongest positive effect, showing overwhelmingly positive influence on the number of predicted species. Biodiversity in Southeast Asia has been threatened for decades by direct and indirect human activities (Sodhi et al. 2004, Hughes 2017b), and it was not surprising to highlight a strong avoidance for anthropic disturbances and a preference for protected areas. By mapping the areas that host the richest terrestrial vertebrate biodiversity, this tool could become beneficial also for current policy discussions on protection of biodiversity and natural habitats, such as the achievement of the 30 by 30 target, aiming at protecting at least 30% of Earth's land and ocean area by 2030 in order to preserve areas of particular importance for ecosystems and biodiversity. Clearly defining the geographic

distribution of terrestrial vertebrate biodiversity allows to precisely focus conservation actions, in order not only to preserve portions of the landscape, but also to prioritise those including the highest number of species.

Worryingly, I also highlighted a relative poor correspondence between existing protected areas and regions of modelled biodiversity. Indeed, less than one-fifth of the modelled species richness occurred within protected areas, raising concerns for the unique and threatened biodiversity of Southeast Asia. On the one hand, this result would suggest that the most straightforward solution is expanding the current system of protected areas, ensuring protection of the areas that I highlighted as the most biodiverse, as well as the corridors connecting them (Cushman et al. 2018, Kaszta et al. 2020b, Penjor et al. 2021).

However, implementing and managing protected areas over as a vast and socioeconomically diverse region as Southeast Asia is a difficult task. In recommending simple expansions of the protected areas network I am mindful of the risk of creating paper parks if protection does not advance hand in hand with enforcement, enhanced willingness to embrace coexistence and adequacy of finances (Di Minin and Toivonen 2015). Therefore, besides the more classical biogeographic thinking of preserving the most biodiverse areas, a parallel approach encouraging tolerance and coexistence with wildlife has to be implemented (Western et al. 2019). I then advocated for a holistic approach between conservationists, stakeholders and governmental institutions to find an equilibrium between preserving biodiversity and ensuring the well-being of local human populations. In this chapter, I provided a tool to try to achieve this result, by identifying the habitat factors that promote richness in species, and highlighting the regions where the highest values of terrestrial vertebrate biodiversity occur.

I acknowledge, however, also limitations in the analyses I have carried out to model species richness in mainland Southeast Asia. First, and most importantly, there is a clear bias and spatial unrepresentativeness of the camera trap surveys, which were clustered and occurred mostly in protected areas, instead of being representative of the full gradients of all the environmental factors in the study area. This limitation might have likely biased the spatial projection toward protected areas, and further research to include also biodiversity data in the less sampled area is strongly recommended. Additionally, I carried out the analyses in a multivariate context, analysing species abundance data altogether in order to highlight the drivers of the sampled biodiversity community as a whole, while still keeping the individual contribution of each sampled species. The choice of this modelling framework has determined the use of suboptimal analyses to model habitat suitability, such as the use of forward selection analysis to assess the most importance habitat variables. Additionally, the use a multivariate modelling framework prevented me from modelling the habitat suitability independently for each species, while I modelled habitat suitability for single species using the habitat variables found important for the whole sampled biodiversity.

**Chapter 3.** Multi-scale, multivariate community models improve designation of biodiversity hotspots in the Sunda Islands.

The objective of the third chapter was to identify the main drivers of terrestrial vertebrate biodiversity assemblages, independently in the islands of Borneo and Sumatra. I have carried out the analyses independently for each of Borneo and Sumatra due to the different environments and biotic communities occurring in the two islands. Then, building on the modelled multivariate relationships between sampled species and habitat factors, I

have produced spatially explicit predictions of the geographic distribution of terrestrial vertebrate biodiversity in Borneo and Sumatra, and used the modelled species richness to assess the effectiveness of the systems of protected areas in preserving regional biodiversity, as well as to identify the most suitable areas currently lacking protection. In Sumatra, the spatially explicit prediction revealed a distribution pattern of the sampled species coherent with previous, more anecdotal knowledge (i.e., IUCN species ranges), with the areas of highest biodiversity occurring mainly along the Barisan Mountains, representing the backbone of the islands along the western coast. This specific distribution of biodiversity has already been highlighted for other species, such as Sunda clouded leopard (Macdonald et al. 2018a), and it seems to be driven by both bio-geographic drivers (i.e., altitudinal and latitudinal gradients, as discussed also for Chapter 2) and anthropic pressure occurring on the rest of the island, especially in the eastern lowlands. The eastern provinces of Sumatra are among the most affected by deforestation for oil palm and wood fibre plantations (Singh and Yan 2021), and I highlighted that the remaining refuges for biodiversity in these extents are restricted to protected areas.

In Borneo, my results provided novel insights into the distribution of local biodiversity. Indeed, while both anecdotal knowledge (i.e., IUCN species ranges) and formal assessments of biodiversity distribution (Mathai et al. 2016) highlighted the central highlands of Borneo as the stronghold of regional biodiversity, I highlighted that the areas of highest species richness occurred mostly in the Malaysian state of Sabah, in the northern extent of the island, and in the Indonesian province of North Kalimantan, in the northwest of the island, displaying a pattern much more similar to what modelled for plant biodiversity (Raes et al. 2009). These results are likely due to the fact that assessments of Bornean biodiversity have usually focused mostly on carnivore species, for which the pristine and intact forests of the central highlands are ideal habitat, while I considered a much broader suit of species,

including more generalist ones. Supporting this conclusion, I highlighted that elevation was an important factor driving the presence of several carnivore species, while other species were less associated with elevation, and more affected by other factors such as presence of protected areas. High altitude areas in Borneo are indeed concentrated towards the centre of the island, which has traditionally been considered the most important area for conservation in Borneo. While strongly supporting the protection of this pristine extent of the island to support the highly specialised carnivores of the region, in this chapter I also highlighted that a broader, likely more generalist suite of species occur also in other regions of the islands, such as the northern extents.

One of the more striking results in Borneo, was the association of endangered species like Bornean orangutan *Pongo pygmaeus*, banteng *Bos javanicus* and Asian elephant *Elephas maximus* with highly disturbed environments characterised by high human footprint and forest loss. I interpreted this result in light of the recent and ongoing deforestation that Borneo has been facing for the last 50 years, and that has caused a disequilibrium between the current distribution of species and the availability of their key habitat resources (i.e., a *lag effect*, sensu Kaszta et al. (2020b)). Namely, deforestation in Borneo has been so rapid and extensive that the current distribution of several endangered species in disturbed lowland forest ecosystems does not reflect their habitat requirements, but an obligate association that has not re-established its natural equilibrium yet.

Comparing the modelled geographic distribution of the terrestrial vertebrate biodiversity in Borneo and Sumatra with the current systems of protected areas, I demonstrated that the effectiveness of the existing systems of protected areas in preserving biodiversity is worryingly poor, especially in Borneo, where less than one-tenth of the modelled species richness occurred within protected areas, while this proportion was less than one-fifth in Sumatra.

Protected areas in Borneo have been established mostly in the central highlands of the island where the largest patches of intact forest occur, representing precious refuge sites for biodiversity. However, while acknowledging the importance of the existing protected areas, in this chapter I have formally highlighted, by using empirical data of species occurrence, that also other regions of Borneo, such as Sabah, are particularly important for biodiversity conservation, and protection in these areas should be increased. Unfortunately, we are witnessing a rather opposite trend, with Sabah's remaining forests being logged in favour of monocrops such as oil palm plantations, and with strongly biodiverse regions in North and East Kalimantan being threatened by human development, including the proposed new Indonesian capital (Cushman et al. 2017, Teo et al. 2020).

In Sumatra, I highlighted that two of the largest protected areas in the island, namely Kerinci Seblat and Bukit Barisan Selatan National Parks in the Barisan Mountains, harbour more than half of the protected biodiversity on the island. Therefore, the geographic pattern of species richness in Sumatra is also negatively affected by anthropogenic factors, with forest loss and consequent conversion to oil palm plantations, occurring mostly in the eastern lowlands of the island, representing the strongest predictors of low species richness.

As discussed for Chapter 2, also this chapter was characterised by similar limitations associated with the multivariate modelling context and with the camera traps sampling design. The camera trap surveys failed to sample the full gradients of habitat factors chosen to model species richness. Ideally, sampling strategies should be designed to identify the full gradients of habitat factors that are most influential to species' habitat selection, rather than being biased towards a fraction of the full spectrum of environmental conditions (Wessels et al. 1998). In Sunda Islands, however, it has been impossible to implement a randomly stratified survey given the extent of the study area and the inaccessibility of several regions. I tried to compensate for this limitation by producing Mahalanobis distance masks to cover the

area that were ecologically too different from the sampling locations. However, I also have to acknowledge that the survey strategy in Sumatra was more geographically biased and less representative than the one implemented in Borneo. Therefore, the model for Sumatra is inevitably less reliable than the model for Borneo, as also suggested by the different variance explained by the habitat factors used in the two islands. Therefore, when consulting the model for Sumatra, it is important to highlight that its reliability in areas further away from the sampling locations might be limited.

**Chapter 4.** Not seeing the forest for the trees: generalised linear model out-performs random forest in species distribution modelling for Southeast Asian felids.

In the fourth chapter, I carried out a formal assessment to evaluate the statistically strongest and most supported methodological approach to model habitat suitability and species distribution, using small felids in both the mainland and offshore extents of Southeast Asia as a case study. There have been several main algorithms adopted in the field of landscape ecology to model the relationships occurring between species and habitat factors. However, two of the mostly applied are the generalised linear model (GLM) and random forest (RF). The former represents an extension of linear model, that can account for non-linearity of the data (Guisan et al. 2002), and that has been the most common algorithm adopted for habitat suitability modelling (Elith and Leathwick 2009). The latter belongs to the Classification and Regression Tree (CART) machine learning algorithms (Breiman 2001), of which it is the most common in the field of landscape ecology (Stupariu et al. 2021). RF overcomes some of the limitations of CARTs such as model over-fitting through the process of bagging (bootstrapping aggregating), which consists in creating multiple subsets of the

training data by randomly sampling with replacement, and then by training separate decision trees with each subset. The predictions resulting from each decision trees are then combined, allowing the model to stabilise and to reduce the impact of over-fitting individual decision trees. Additionally, RF is often the preferred among the CART techniques due to its flexibility in predicting different relationships between predictor and response variables. In addition to assessing the effect of algorithm, I also wanted to evaluate the effect of different resampling techniques on the training datasets and, specifically, the effect of bootstrapping the training datasets on the performances of habitat suitability models.

In total, I evaluated four different modelling frameworks: GLM and RF, with and without bootstrapping the training datasets. I applied these four methodological frameworks to four different small felid species in three study regions across Southeast Asia. In each assessment I produced species distribution models for each of these four combinations, and I evaluated their predictive performances by calculating the Area Under the ROC Curve (AUC), Percent Correctly Classified (PCC), sensitivity, specificity and True Skill Statistic (TSS) on independent hold-out data.

The most striking result that I highlighted was that species distribution models produced with GLM and with RF showed strong differences between each other, while differential bootstrapping techniques did not produce large differences between models. I did not expect strong differences between resulting models based on the algorithm used, as they are both intended to produce unbiased and accurate estimates of the relationship between species occurrence and habitat factors. Finding large and consistent differences in the predictions of GLM and RF suggests caution in selecting and applying species distribution algorithms, and that this sensitivity must be taken into account by future studies when defining the framework to model habitat suitability. Habitat suitability models are used to prioritise areas for protection of endangered and rare species (Penjor et al. 2021), and

differences based on the algorithm applied, and not on the underlying ecological relationships between species and habitat factors, can have profound and serious repercussions on the effectiveness of the outcomes of these conservation actions.

Additionally, another striking result that I highlighted, was that models trained with GLM showed, overall, higher AUC scores than RF. AUC is acknowledged to be a strong metric to evaluate the reliability of habitat suitability models (Shabani et al. 2016), and I adopted it to define the most supported models. A number of past studies suggested that RF represents a more robust modelling technique, producing more strongly supported models (Cushman et al. 2017, Kumar et al. 2021). Therefore, I was surprised that the most supported models trained with GLM were twice as frequent as the most supported models trained with RF. However, there are cases in which RF has been reported to perform poorly when trained with imbalanced presence-absence data, likely due to the sensitivity of this algorithm to the most abundant class (Evans and Cushman 2009, Freeman et al. 2012). Additionally, it has been reported that, while machine learning techniques would only slightly improve the predictive performances compared to simpler algorithms, the main impact of the increased complexity of machine learning was observed in model overfitting (Ramampiandra et al. 2023). In addition, more complex modelling technique, such as RF, while showing better interpolation performances than simpler techniques, perform poorly when extrapolating the models over broader geographic extents (Heikkinen et al. 2012, Stock et al. 2020).

I also found strong evidence supporting the hypothesis that the adoption of different algorithms affected the results of the scales optimisation process, which I carried out to evaluate the spatial scale at which each habitat factor showed its strongest effect on felid species. Together, these results raised a worrying red flag since I demonstrated previously unrecognised uncertainty and variability in the outcomes of habitat suitability models, which are systematically affected by the choice of modelling algorithms, instead of being

objectively driven by the ecological relationships between species and habitat factors. I strongly advocate for landscape ecologists and conservationists to take much more care in the choice of the modelling algorithms, to choose algorithms after careful and thorough comparisons of multiple methods, and to select the one that is more robust and effective for the particular system and species being studied.

As previously discussed for Chapters 2 and 3, one of the main limitation of the study is the lack of spatial representativeness of the camera trap surveys, which were clustered and failed to comprehensively depict the full gradients of the habitat factors used to models habitat suitability for sampled felids. However, by producing Mahalanobis distance masks independently for mainland Southeast Asia and for the Sunda Islands to mask the areas that were ecologically too different from the sampling locations, I have minimised the issues associated with using a clustered camera trap dataset. In addition, when training RF I have not checked the variables for multicollinearity. Although some authors state that RF is not affected by multicollinearity between predictor variables (Evans and Cushman 2009), others have highlighted that when RF are trained with correlated variables, these provide redundant information in the model, and potentially leading to overfitting the models and decreasing their performances. As a result, if correlated variables are include in RF, their variable importance would be lower compared to uncorrelated variables, leading to potentially underestimate the importance of correlated variables in the model (Genuer et al. 2010, Bradter et al. 2013). While acknowledging this possible issue, I decided not to check for multicollinearity between variables.

**Chapter 5.** Identifying gaps in the conservation of small wild cats of Southeast Asia.

In this chapter, building on the results highlighted in Chapter 4 and using the statistically most supported models for small felids in each of the three study regions, I produced species richness models identifying the most suitable areas for Southeast Asian small felids individually and collectively. I then evaluated the most effective protected areas in preserving this group of species, and highlighted the suitable areas still lacking formal protection.

Southeast Asia hosts more felid species than any other region globally, with 13 felids species occurring, of which 11 are endemic to the region (Macdonald and Loveridge 2010). However, conservation efforts and economic investments for preserving species and habitats are strongly biased towards a few charismatic species (Brodie 2009, Macdonald et al. 2015). It has been observed that the species whose conservation is more likely to receive attention from the general public and that can, therefore, play major roles as ambassador species, are mostly big, endangered carnivores (Macdonald et al. 2015, Macdonald et al. 2017). Thus, it is unsurprising that, in Southeast Asia, tiger *Panthera tigris* and leopard *Panthera pardus*, the two largest felids in the region, have catalysed the largest investments for conservation, as well as the awareness of the public (Brodie 2009). Additionally, the two clouded leopard species, mainland clouded leopard *Neofelis nebulosa* and Sunda clouded leopard *Neofelis diardi* represent the next felid species for size in Southeast Asia, and their conservation status has recently become a regional focus for conservation actions, with the near disappearance of tigers and leopards from the region (Macdonald et al. 2018a, Macdonald et al. 2019).

Smaller felid species, however, remain largely understudied, especially at continental scale, and assessments of their conservation status and ecology are limited to local contexts

(Haidir et al. 2018, Hearn et al. 2018, Rasphone et al. 2021, Rostro-Garcia et al. 2021). Nevertheless, small felids are secondary consumers, and occupy important ecological roles in ecosystems, where they exercise top-down control on herbivores which, in turn, affect apex predators (Roemer et al. 2009). In addition, if larger predators are extirpated from the ecosystem, small felids can take over the role of apex predator (de Oliveira et al. 2010). Therefore, here I assessed the most suitable areas for small felids, and highlighted that, in mainland Southeast Asia, the most suitable habitats occur in the northern extents of Myanmar, specifically in the Northern Forest Complex, a vast system of contiguous forested areas extending at the border with China and India, and area already highlighted as fundamental for the conservation of mainland clouded leopard by Kaszta et al. (2020b).

In this area, I found some of the most important protected areas for small felids in Southeast Asia, but I also highlighted extensive regions of highly suitable, but unprotected, habitats. This is worrying, because these areas, despite their suitability, are also at the centre of planned human infrastructure development, including the Belt and Road Initiative, which poses a serious risk for the integrity of the suitable habitat for biodiversity (Kaszta et al. 2020a).

I also modelled a stronghold for the conservation of small felids in the eastern extent of Myanmar, including Laos and Vietnam. While promising, this results is also concerning since it is known that felid species occurring in these parts of Southeast Asia are at serious risk of illegal poaching and snaring (Johnson et al. 2016) and, unless dramatic improvement in species protection occurs, smaller felids risk to be extirpated from these areas, following the fate of tigers and leopards (Rasphone et al. 2019, Rasphone et al. 2021).

In Borneo, I highlighted that the most suitable habitats for small felids occurred in the central highlands of the island, as also found by Mathai et al. (2016) for Bornean carnivores.

This area, closely matching also the proposed outline of the Hearth of Borneo initiative, a transboundary conservation area signed between Brunei Darussalam, Indonesia and Malaysia (Keong and Onuma 2021), is less impacted by habitat loss and deforestation than the surrounding lowlands (Cushman et al. 2017), and stands as a uniquely precious refuge site for felids and other carnivores, which are more specialised than other generalist species and less able to survive in degraded landscapes. In addition, I found that also the northern Malaysian state of Sabah harbours extensive and highly suitable habitats for Bornean small felids. Mostly driven by the protection of the central highlands of Borneo, I found that protected areas were critical to the preservation the most suitable habitats for small felids, with one-fourth of the highly suitable habitats encompassed within the ten most important protected areas.

In Sumatra, I found that the Barisan Mountains were critical in providing suitable habitats for small felids. Although the entire mountain range represented a stronghold, protected areas showed overwhelming importance. For example, two of the largest protected areas on the island, namely Kerinci Seblat and Bukit Barisan Selatan National Parks, hosted more than one-third of the highly suitable habitat for small felids of the entire island. Protected areas were critical also in the eastern portions of Sumatra, where I found that they preserve the last small remnants of highly suitable habitats for small felids in this highly impacted part of the island (Singh and Yan 2021).

Chapter 5 was built on the results obtained in Chapter 4 and, consequently, it suffers from the same limitations as the previous chapter, notably the lack of geographic representativeness of camera traps, and the fact that I did not check for multicollinearity between variables when training RF.

**Implications for conservation in Southeast Asia**

It is widely recognised that global biodiversity is facing a period of threat and decline comparable with the five previous mass extinctions of Earth's history in both rate and magnitude (Barnosky et al. 2011, Dirzo et al. 2014). The drivers of this sixth mass extinction are widely recognised to be the effects of human population and economic growth, which drive direct exploitation of habitats and species, land use change and climate change (Barnosky et al. 2011). As the human population is expected to keep growing for at least the remainder of the 21st century (Lee 2011), and resource consumption and economic growth are accelerating rapidly in the developing world, the anthropic pressure on biodiversity is unlikely to diminish (Newbold 2018). Therefore, it is essential to focus conservation efforts in order to optimise conservation actions to where it is urgent and effective. In other words, it is critical to optimise and prioritise the limited resources available for conservation.

Myers et al. (2000)'s work, and subsequent updates by Mittermeier et al. (2004) and Mittermeier et al. (2011), led the way towards prioritisation of areas important for biodiversity, by highlighting the global regions richest in endemism and facing high rates of habitat loss, namely, the biodiversity hotspots. The idea of highlighting the areas where conservation actions are most urgent and are likely to achieve the greatest impacts has been noteworthy, and its significance is evidenced by the fact that, since its introduction, the concept of biodiversity hotspots has catalysed more than $1 billion in conservation investments (Sloan et al. 2014).

Biodiversity hotspots, however, have also raised criticisms, mostly concerning the coarse and excessively broad scale at which these regions have been identified, which diminishes their effectiveness in focusing spatially optimal conservation actions.

Among the different possible solutions to remedy this limitation of biodiversity hotspots, I found the concept of "micro-hotspots" or "hotspots within hotspots" (Murray-Smith et al. 2009, Canadas et al. 2014, Noroozi et al. 2018) particularly insightful and promising. Spatial prioritisation of smaller areas within the biodiversity hotspots, where most of the species richness is concentrated offers a robust and informative approach not only to overcome some of the limitations of the biodiversity hotspots, such as the difficulty of implementing effective conservation actions in excessively vast areas, but also to allow meta-population dynamics within broader natural populations (Dallas et al. 2020). In the context of micro-hotspots, connectivity becomes important for maintaining genetic diversity and promoting species resilience. By enabling the movement of individuals between micro-hotspots and their surrounding environments, populations can avoid inbreeding and can increase their ability to adapt to changing conditions. However, in order to ensure the long term survival of species, it is also crucial that micro-hotspots can maintain minimum viable populations, and that a sufficient area is available to support the requirements of the diverse species occurring. I then applied this method in the Southeast Asian context. Additionally, acknowledging the ambassadorial role that charismatic species have for raising awareness and funds for conservation, and knowing that felid species have great potentialities as ambassador species (Brodie 2009, Macdonald et al. 2015, Macdonald et al. 2017), I decided to focus my thesis on this family. Specifically, since there is paucity of range-wide studies on small felids, while larger felid species such as tigers and leopards have been the main focus of conservation actions in Southeast Asia, I decided to focus on this still largely understudied portion of regional biodiversity.

I found that, overall, the most suitable habitats both for Southeast Asian terrestrial vertebrate biodiversity and for smaller felids occurred in association with large patches of forested areas, still relatively untouched by anthropic pressure, often in existing protected

areas, while the areas with lowest biodiversity and smaller felid diversity were consistently areas with high human impact.

When I mapped the geographic distribution of the suitable habitats for biodiversity and for small felids, I highlighted differences that provided interesting insights into the requirements of different subsets of regional biodiversity. In mainland Southeast Asia, I highlighted two main strongholds for biodiversity or, to follow Canadas et al. (2014)'s definition, two "hotspots within hotspots", namely the mountainous region of Southwest China and the Thai-Malay Peninsula, similarly to the projections of Hughes (2017a). The area of high biodiversity that I predicted to occur in the mountainous region of Southwest China closely matched the officially recognised Mountains of Southwest China biodiversity hotspot. However, illustrating the often excessively coarse nature of official biodiversity hotspots, Myers et al. (2000) highlighted the Indo-Burma region as a whole as a biodiversity hotspot, while I predicted that, in this vast area, it is mostly the Thai-Malay Peninsula that harbours the highest richness in terrestrial vertebrate species.

Besides the importance of forested areas, I found that additional critical factors predicting species richness in Southeast Asia, in line with classic biogeographic theory, were altitude and latitude (Hillebrand 2004, McCain 2005). In the Thai-Malay Peninsula, the highly productive, wet and stable tropical rainforest climate, coupled with topographic heterogeneity along altitudinal gradients promotes environmental diversity, where several species distribute their ecological niches. Similarly, the landscapes of the mountainous region of Southwest China are characterised by ridge-valley complexes, shaped by large rivers and divided by tall mountain ranges, which drives environmental and climatic heterogeneity that, in turn, promotes niche diversity and supports high levels of biodiversity.

The spatial projections that I produced for small felids show a somewhat slightly different geographic pattern of habitat suitability than I predicted for biodiversity as a whole. While the southern extent of the mountainous region of Southwest China, in correspondence with the Northern Forest Complex of Myanmar, matched the model for overall terrestrial biodiversity, I predicted that the suitable habitat for small felids extended westwards and southwards, including areas that harbour relatively less biodiversity. The strongest difference, however, was seen at the southern extents of the study areas, where the model for small felids highlighted highly suitable habitat in eastern Myanmar, Laos and, partially, in Vietnam, while these areas showed less, and much more scattered, terrestrial biodiversity. These differences are likely due to the fact that smaller felids are much more specialised than the rest of terrestrial biodiversity that, in contrast, included also several generalist species.

Among smaller felids, I also assessed the suitable habitat for leopard cat, known to be a rather generalist species that can survive also in disturbed environments (Mohamed et al. 2013). However, the results that I highlighted suggest that smaller felids, generally, are strongly associated with more pristine, untouched and large extents of forested areas, as also suggested by previous studies (Macdonald et al. 2018a, Macdonald et al. 2019). This conclusion was particularly evident in Borneo, where the model for terrestrial vertebrate biodiversity indicated that the northern parts of the island were the most biodiverse region, while the central highlands, usually considered the core of Bornean biodiversity, harboured relatively less species richness. However, the central highlands of Borneo were revealed as highly suitable for smaller felids.

From these results, I can first conclude that, consistent with Brodie (2009) and Macdonald et al. (2015), there is a bias towards carnivores in the general knowledge on species' conservation. Additionally, similar to mainland Southeast Asia, Bornean small felids also show higher specialisation than the rest of the terrestrial vertebrate biodiversity, and

while the latter can occur in more disturbed areas, the former require more vast and more pristine forested landscapes.

Similar conclusions were apparent for Sumatra as well, although in this case I highlighted much stronger similarities between the modelled distribution for terrestrial vertebrate biodiversity and for smaller felids. In the eastern regions of the Sumatra, where deforestation and habitat conversion have been more intense (Singh and Yan 2021), the only areas of high predicted biodiversity were concentrated around protected areas, while in the much less degraded Barisan Mountains, although protected areas still represented strongholds for conservation, suitable habitat occurred throughout the mountain range.

Even though I highlighted the most suitable areas for terrestrial vertebrate biodiversity and for small felids, as well as the areas where protection should be increased to gain the maximum outcomes, for these solutions to be effective, conservation planning must go hand in hand with the assessment and the integration of social issues within conservation actions (Sodhi and Brook 2006). Sadly relevant examples of the poor performance of protected areas that are not accompanied by an increase of the cultural tolerance and of an ethic of wildlife protection is given by the recent extirpation of tigers and leopards from one of the most important protected area in Laos (Rasphone et al. 2019, Rasphone et al. 2021). I believe that the results that I have provided in my thesis represent a reliable and robust tool to prioritise conservation areas in Southeast Asia. However, there is a risk that even if such prioritisation is acted on, it might result in the creation of so-called “paper parks” if conservation does not proceed with an integrated approach that embraces the holism of the human-wildlife coexistence: ensuring not only the conservation of species and habitats, but also the economic benefits, equity, rights and active participation of local stakeholders (Chazdon et al. 2009).

**Scope and limitation**

This thesis built on one of the richest camera trap dataset ever collected in Southeast Asia, spanning a survey period of more than a decade, and including abundance data for a suite of regional terrestrial vertebrate biodiversity. However, it was not without weaknesses, and I have acknowledged them in the various chapters and taken them into account to the best of my abilities.

Although the dataset used to model the biodiversity and the small-sized felids habitat suitability included almost 3,000 camera trap locations, they suffered from spatial bias. Camera trap surveys were not designed, in their original intent, to sample biodiversity throughout Southeast Asia, but to monitor felid species and, specifically, clouded leopards, at finer scales. Therefore, the camera trap surveys were not optimal to sample the full gradients of all the habitat factors that I evaluated. Ideally, sampling strategies should be designed to identify the full gradients of the factors expected to affect sampled species, rather than being biased towards a fraction of the full spectrum of environmental conditions (Wessels et al. 1998). Failing to survey the full gradients of the habitat conditions might result in models with erroneous parameter estimates and weak predictive performances (Chiaverini et al. 2021). Given that models risk unreliability when projected outside their domain (Barbosa et al. 2009), I acknowledged this limitation by masking out, from the spatially explicit predictions of biodiversity and of smaller felids' distributions, the areas whose habitat conditions were ecologically very different from the sampling locations.

The fourth chapter of this thesis brought up another possible fundamental limitation of my work. The evaluation of the different modelling frameworks to model habitat suitability for smaller felids in Southeast Asia revealed that the algorithm I thought would have provided the most reliable results did not, in fact, perform best. This result was worrying

considering the importance that habitat suitability models have in the decision-making process for conservation plans, and I tried to minimise the drawbacks by testing two of the more broadly adopted algorithms. However, I tested only two of the multitude of algorithms available for modelling habitat suitability, and a much more comprehensive assessment is needed to draw a more solid conclusion on the effects of algorithms on model performance (Valavi et al. 2022). Additionally, as mentioned above, the dataset used to model habitat suitability resulted from spatially biased camera trap surveys; this is known to affect the outcomes and the predictive performances of such models (Chiaverini et al. 2021), masking the real effects of the different algorithms. Therefore, I recommend future research to comprehensively evaluate the effects of geographic bias, algorithms and resampling techniques by carrying out formal simulation experiments, in which the confounding effect of each factor is controlled, and their effects identified (Atzeni et al. 2020, Chiaverini et al. 2021, Kumar et al. 2022).

Finally, although I adopted cutting-edge and robust modelling techniques to investigate species-habitat relationships and to highlight the most suitable habitats for biodiversity and for smaller felids, Southeast Asia is affected by rates of habitat loss and deforestation that are the highest globally. Therefore, in order for habitat suitability models to be a useful tool for implementing effective conservation actions within such a dynamic region, they should model not only the current distribution of biodiversity, but also its distribution within future scenarios (Macdonald et al. 2018b, Kaszta et al. 2019, Kaszta et al. 2020a). Additionally, the majority of the habitat suitability assessments that evaluated future scenarios, focus solely on climate change projections (Titeux et al. 2016). Although the climatic projections are critically important, I believe that future scenarios for Southeast Asian biodiversity should consider also landscape changes, as well as socio-economic dynamics (Macdonald et al. 2018b, Kaszta et al. 2019, Kaszta et al. 2020a). These models

would require synergy between different disciplines and complex frameworks that account for intertwined relationships between economics, policy and demography, but I believe that only in this way can effective and proactive conservation interventions for Southeast Asia be designed and implemented.

## Conclusions

In this thesis I have provided additional insights into the distribution of terrestrial vertebrate biodiversity in Southeast Asia, a region known for its uniquely rich biodiversity, but that is also facing extremely high rates of habitat loss and deforestation. Therefore, although Southeast Asia as a whole has already been defined as a biodiversity hotspot, I explicitly focused on spatially prioritising the areas of highest species richness in order to optimise conservation actions. I also evaluated the habitat suitability for smaller felids, of which Southeast Asia represents the region of globally highest sympatry, providing additional information on an understudied group of carnivores, which is often overshadowed by more charismatic species. The assessment of the most statistically supported models for small felids revealed interesting and unexpected characteristic of the modelling algorithms, which suggest caution when choosing the modelling framework to assess species habitat suitability.

## References

Atzeni, L., S. A. Cushman, D. F. Bai, J. Wang, P. J. Chen, K. Shi, and P. Riordan. 2020. Meta-replication, sampling bias, and multi-scale model selection: A case study on snow leopard (Panthera uncia) in western China. Ecology and Evolution **10**:7686-7712.

Barbosa, A. M., R. Real, and J. M. Vargas. 2009. Transferability of environmental favourability models in geographic space: The case of the Iberian desman (Galemys pyrenaicus) in Portugal and Spain. Ecological Modelling **220**:747-754.

Barnosky, A. D., N. Matzke, S. Tomiya, G. O. U. Wogan, B. Swartz, T. B. Quental, C. Marshall, J. L. McGuire, E. L. Lindsey, K. C. Maguire, B. Mersey, and E. A. Ferrer. 2011. Has the Earth's sixth mass extinction already arrived? Nature **471**:51-57.

Bradter, U., W. E. Kunin, J. D. Altringham, T. J. Thom, and T. G. Benton. 2013. Identifying appropriate spatial scales of predictors in species distribution models with the random forest algorithm. Methods in Ecology and Evolution **4**:167-174.

Breiman, L. 2001. Random forests. Machine Learning **45**:5-32.

Brodie, J. F. 2009. Is research effort allocated efficiently for conservation? Felidae as a global case study. Biodiversity and Conservation **18**:2927-2939.

Canadas, E. M., G. Fenu, J. Penas, J. Lorite, E. Mattana, and G. Bacchetta. 2014. Hotspots within hotspots: Endemic plant richness, environmental drivers, and implications for conservation. Biological Conservation **170**:282-291.

Chazdon, R. L., C. A. Harvey, O. Komar, D. M. Griffith, B. G. Ferguson, M. Martinez-Ramos, H. Morales, R. Nigh, L. Soto-Pinto, M. van Breugel, and S. M. Philpott. 2009. Beyond Reserves: A Research Agenda for Conserving Biodiversity in Human-modified Tropical Landscapes. Biotropica **41**:142-153.

Chiaverini, L., H. Y. Wan, B. Hahn, A. Cilimburg, T. N. Wasserman, and S. A. Cushman. 2021. Effects of non-representative sampling design on multi-scale habitat models: flammulated owls in the Rocky Mountains. Ecological Modelling **450**.

Cuesta, F., P. Muriel, L. D. Llambí, S. Halloy, N. Aguirre, S. Beck, J. Carilla, R. I. Meneses, S. Cuello, A. Grau, L. E. Gámez, J. Irazábal, J. Jácome, R. Jaramillo, L. Ramírez, N. Samaniego, D. Suárez-Duque, N. Thompson, A. Tupayachi, P. Viñas, K. Yager, M.

T. Becerra, H. Pauli, and W. D. Gosling. 2017. Latitudinal and altitudinal patterns of plant community diversity on mountain summits across the tropical Andes. Ecography **40**:1381-1394.

Cushman, S. A., N. B. Elliott, D. Bauer, K. Kesch, L. Bahaa-el-din, H. Bothwell, M. Flyman, G. Mtare, D. W. Macdonald, and A. J. Loveridge. 2018. Prioritizing core areas, corridors and conflict hotspots for lion conservation in southern Africa. Plos One **13**.

Cushman, S. A., E. Macdonald, E. Landguth, Y. Malhi, and D. Macdonald. 2017. Multiple-scale prediction of forest loss risk across Borneo. Landscape Ecology **32**:1581-1598.

Dallas, T. A., M. Saastamoinen, T. Schulz, and O. Ovaskainen. 2020. The relative importance of local and regional processes to metapopulation dynamics. Journal of Animal Ecology **89**:884-896.

de Oliveira, T. G., M. A. Tortato, L. Silveira, C. B. Kasper, F. D. Mazim, M. Lucherini, A. T. Jácomo, J. B. G. Soares, R. V. Marques, and M. Sunquist. 2010. Ocelot ecology and its effect on the small-felid guild in the lowland Neotropics. Pages 559-580 *in* D. W. Macdonald and A. J. Loveridge, editors. The Biology and Conservation of Wild Felids. Oxford University Press, Oxford, UK.

Di Minin, E., and T. Toivonen. 2015. Global Protected Area Expansion: Creating More than Paper Parks. Bioscience **65**:637-638.

Dirzo, R., H. S. Young, M. Galetti, G. Ceballos, N. J. B. Isaac, and B. Collen. 2014. Defaunation in the Anthropocene. Science **345**:401-406.

Elith, J., and J. R. Leathwick. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annual Review of Ecology Evolution and Systematics **40**:677-697.

Evans, J. S., and S. A. Cushman. 2009. Gradient modeling of conifer species using random forests. Landscape Ecology **24**:673-683.

Freeman, E. A., G. G. Moisen, and T. S. Frescino. 2012. Evaluating effectiveness of down-sampling for stratified designs and unbalanced prevalence in Random Forest models of tree species distributions in Nevada. Ecological Modelling **233**:1-10.

Genuer, R., J. M. Poggi, and C. Tuleau-Malot. 2010. Variable selection using random forests. Pattern Recognition Letters **31**:2225-2236.

Guisan, A., T. C. Edwards, and T. Hastie. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling **157**:89-100.

Haidir, I. A., D. W. Macdonald, and M. Linkie. 2018. Assessing the spatiotemporal interactions of mesopredators in Sumatra's tropical rainforest. Plos One **13**.

Hearn, A. J., S. A. Cushman, J. Ross, B. Goossens, L. T. B. Hunter, and D. W. Macdonald. 2018. Spatio-temporal ecology of sympatric felids on Borneo. Evidence for resource partitioning? Plos One **13**.

Heikkinen, R. K., M. Marmion, and M. Luoto. 2012. Does the interpolation accuracy of species distribution models come at the expense of transferability? Ecography **35**:276-288.

Hillebrand, H. 2004. On the generality of the latitudinal diversity gradient. American Naturalist **163**:192-211.

Hughes, A. C. 2017a. Mapping priorities for conservation in Southeast Asia. Biological Conservation **209**:395-405.

Hughes, A. C. 2017b. Understanding the drivers of Southeast Asian biodiversity loss. Ecosphere **8**.

Johnson, A., J. Goodrich, T. Hansel, A. Rasphone, S. Saypanya, C. Vongkhamheng, Venevongphet, and S. Strindberg. 2016. To protect or neglect? Design, monitoring,

and evaluation of a law enforcement strategy to recover small populations of wild tigers and their prey. Biological Conservation **202**:99-109.

Kaszta, Z., S. A. Cushman, A. J. Hearn, D. Burnham, E. A. Macdonald, B. Goossens, S. K. S. S. Nathan, and D. W. Macdonald. 2019. Integrating Sunda clouded leopard (Neofelis diardi) conservation into development and restoration planning in Sabah (Borneo). Biological Conservation **235**:63-76.

Kaszta, Z., S. A. Cushman, S. Htun, H. Naing, D. Burnham, and D. W. Macdonald. 2020a. Simulating the impact of Belt and Road initiative and other major developments in Myanmar on an ambassador felid, the clouded leopard, Neofelis nebulosa. Landscape Ecology **35**:727-746.

Kaszta, Z., S. A. Cushman, and D. W. Macdonald. 2020b. Prioritizing habitat core areas and corridors for a large carnivore across its range. Animal Conservation.

Keong, C. Y., and A. Onuma. 2021. Transboundary Ecological Conservation, Environmental Value, and Environmental Sustainability: Lessons from the Heart of Borneo. Sustainability **13**.

Kumar, S. U., P. K. Maini, L. Chiaverini, A. J. Hearn, D. W. Macdonald, Z. Kaszta, and S. A. Cushman. 2021. Smoothing and the environmental manifold. Ecological Informatics **66**.

Kumar, S. U., J. Turnbull, O. H. Davies, T. Hodgetts, and S. A. Cushman. 2022. Moving beyond landscape resistance: considerations for the future of connectivity modelling and conservation science. Landscape Ecology **37**:2465-2480.

Lee, R. 2011. The Outlook for Population Growth. Science **333**:569-573.

Macdonald, D. W., H. M. Bothwell, A. J. Hearn, S. M. Cheyne, I. Haidir, L. T. B. Hunter, A. Kaszta, M. Linkie, E. A. Macdonald, J. Ross, and S. A. Cushman. 2018a. Multi-scale

habitat selection modeling identifies threats and conservation opportunities for the Sunda clouded leopard (Neofelis diardi). Biological Conservation **227**:92-103.

Macdonald, D. W., H. M. Bothwell, Ż. Kaszta, E. Ash, G. Bolongon, D. Burnham, Ö. E. Can, A. Campos-Arceiz, P. Channa, G. R. Clements, A. J. Hearn, L. Hedges, S. Htun, J. F. Kamler, K. Kawanishi, E. A. Macdonald, S. W. Mohamad, J. Moore, H. Naing, M. Onuma, U. Penjor, A. Rasphone, D. Mark Rayan, J. Ross, P. Singh, C. K. W. Tan, J. Wadey, B. P. Yadav, and S. A. Cushman. 2019. Multi-scale habitat modelling identifies spatial conservation priorities for mainland clouded leopards (Neofelis nebulosa). Diversity and Distributions **0**:1-16.

Macdonald, D. W., and A. J. Loveridge. 2010. Biology and conservation of wild felids. Oxford University Press, Oxford ; New York.

Macdonald, E. A., D. Burnham, A. E. Hinks, A. J. Dickman, Y. Malhi, and D. W. Macdonald. 2015. Conservation inequality and the charismatic cat: Felis felicis. Global Ecology and Conservation **3**:851-866.

Macdonald, E. A., S. A. Cushman, E. L. Landguth, A. J. Hearn, Y. Malhi, and D. W. Macdonald. 2018b. Simulating impacts of rapid forest loss on population size, connectivity and genetic diversity of Sunda clouded leopards (Neofelis diardi) in Borneo. Plos One **13**.

Macdonald, E. A., A. Hinks, D. J. Weiss, A. Dickman, D. Burnham, C. J. Sandom, Y. Malhi, and D. W. Macdonald. 2017. Identifying ambassador species for conservation marketing. Global Ecology and Conservation **12**:204-214.

Mathai, J., J. W. Duckworth, E. Meijaard, G. Fredriksson, J. Hon, A. Sebastian, M. Ancrenaz, A. J. Hearn, J. Ross, S. Cheyne, A. Wilting, and B. C. Consortium. 2016. Carnivore conservation planning on Borneo: identifying key carnivore landscapes, research priorities and conservation interventions. Raffles Bulletin of Zoology:186-217.

McCain, C. M. 2005. Elevational gradients in diversity of small mammals. Ecology **86**:366-372.

Mittermeier, R. A., P. R. Gil, M. Hoffman, J. Pilgrim, T. Brooks, C. G. Mittermeier, J. Lamoreux, and G. A. B. da Fonseconda. 2004. Hotspots revisited: Earth's biologically richest and most endangered terrestrial ecoregions. Cemex, Mexico City.

Mittermeier, R. A., W. R. Turner, F. W. Larsen, T. M. Brooks, and C. Gascon. 2011. Global Biodiversity Conservation: The Critical Role of Hotspots. Pages 3-22 *in* F. E. Zachos and J. C. Habel, editors. Biodiversity Hotspots. Distribution and Protection of Conservation Priority Areas. Springer, Berlin, Heidelberg.

Mohamed, A., R. Sollmann, H. Bernard, L. N. Ambu, P. Lagan, S. Mannan, H. Hofer, and A. Wilting. 2013. Density and habitat use of the leopard cat (Prionailurus bengalensis) in three commercial forest reserves in Sabah, Malaysian Borneo. Journal of Mammalogy **94**:82-89.

Murray-Smith, C., N. A. Brummitt, A. T. Oliveira-Filho, S. Bachman, J. Moat, E. M. N. Lughadha, and E. J. Lucas. 2009. Plant Diversity Hotspots in the Atlantic Coastal Forests of Brazil. Conservation Biology **23**:151-163.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. Nature **403**:853-858.

Newbold, T. 2018. Future effects of climate and land-use change on terrestrial vertebrate community diversity under different scenarios. Proceedings of the Royal Society B-Biological Sciences **285**.

Noroozi, J., A. Talebi, M. Doostmohammadi, S. B. Rumpf, H. P. Linder, and G. M. Schneeweiss. 2018. Hotspots within a global biodiversity hotspot - areas of endemism are associated with high mountain ranges. Scientific Reports **8**.

Penjor, U., Z. Kaszta, D. W. Macdonald, and S. A. Cushman. 2021. Prioritizing areas for conservation outside the existing protected area network in Bhutan: the use of multi-species, multi-scale habitat suitability models. Landscape Ecology **36**:1281-1309.

Raes, N., M. C. Roos, J. W. F. Slik, E. E. van Loon, and H. ter Steege. 2009. Botanical richness and endemicity patterns of Borneo derived from species distribution models. Ecography **32**:180-192.

Ramampiandra, E. C., A. Scheidegger, J. Wydler, and N. Schuwirth. 2023. A comparison of machine learning and statistical species distribution models: Quantifying overfitting supports model interpretation. Ecological Modelling **481**.

Rasphone, A., J. F. Kamler, M. Tobler, and D. W. Macdonald. 2021. Density trends of wild felids in northern Laos. Biodiversity and Conservation **30**:1881-1897.

Rasphone, A., M. Kery, J. F. Kamler, and D. W. Macdonald. 2019. Documenting the demise of tiger and leopard, and the status of other carnivores and prey, in Lao PDR's most prized protected area: Nam Et - Phou Louey. Global Ecology and Conservation **20**.

Roemer, G. W., M. E. Gompper, and B. Van Valkenburgh. 2009. The Ecological Role of the Mammalian Mesocarnivore. Bioscience **59**:165-173.

Rostro-Garcia, S., J. F. Kamler, C. Minge, A. Caragiulo, R. Crouthers, M. Groenenberg, T. N. E. Gray, V. In, C. Pin, P. Sovanna, M. Kery, and D. W. Macdonald. 2021. Small cats in big trouble? Diet, activity, and habitat use of jungle cats and leopard cats in threatened dry deciduous forests, Cambodia. Ecology and Evolution **11**:4205-4217.

Shabani, F., L. Kumar, and M. Ahmadi. 2016. A comparison of absolute performance of different correlative and mechanistic species distribution models in an independent area. Ecology and Evolution **6**:5973-5986.

Singh, M., and S. H. Yan. 2021. Spatial-temporal variations in deforestation hotspots in Sumatra and Kalimantan from 2001-2018. Ecology and Evolution **11**:7302-7314.

Sloan, S., C. N. Jenkins, L. N. Joppa, D. L. A. Gaveau, and W. F. Laurance. 2014. Remaining natural vegetation in the global biodiversity hotspots. Biological Conservation **177**:12-24.

Sodhi, N. S., and B. W. Brook. 2006. Southeast Asian biodiversity in crisis. Cambridge University Press, Cambridge.

Sodhi, N. S., L. P. Koh, B. W. Brook, and P. K. L. Ng. 2004. Southeast Asian biodiversity: an impending disaster. Trends in Ecology & Evolution **19**:654-660.

Stock, B. C., E. J. Ward, T. Eguchi, J. Jannot, J. T. Thorson, B. E. Feist, and B. X. Semmens. 2020. Comparing predictions of fisheries bycatch using multiple spatiotemporal species distribution model frameworks. Canadian Journal of Fisheries and Aquatic Sciences **77**:146-163.

Stupariu, M. S., S. A. Cushman, A. I. Plesoianu, I. Patru-Stupariu, and C. Furst. 2021. Machine learning in landscape ecological analysis: a review of recent approaches. Landscape Ecology.

Teo, H. C., A. M. Lechner, S. Sagala, and A. Campos-Arceiz. 2020. Environmental Impacts of Planned Capitals and Lessons for Indonesia's New Capital. Land **9**.

Titeux, N., K. Henle, J. B. Mihoub, A. Regos, I. R. Geijzendorffer, W. Cramer, P. H. Verburg, and L. Brotons. 2016. Biodiversity scenarios neglect future land-use changes. Global Change Biology **22**:2505-2515.

Valavi, R., G. Guillera-Arroita, J. J. Lahoz-Monfort, and J. Elith. 2022. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. Ecological Monographs **92**.

Wessels, K. J., A. S. Van Jaarsveld, J. D. Grimbeek, and M. J. Van der Linde. 1998. An evaluation of the gradsect biological survey method. Biodiversity and Conservation **7**:1093-1121.

Western, G., D. W. Macdonald, A. J. Loveridge, and A. J. Dickman. 2019. Creating Landscapes of Coexistence: Do Conservation Interventions Promote Tolerance of Lions in Human-dominated Landscapes? Conservation & Society **17**:204-217.

Wikramanayake, E. D. 2002. Terrestrial ecoregions of the Indo-Pacific : a conservation assessment. Island Press, Washington D.C.; London.

Willig, M. R., and S. J. Presley. 2016. Biodiversity and metacommunity structure of animals along altitudinal gradients in tropical montane forests. Journal of Tropical Ecology **32**:421-436.