ASSESSING SPECIES DISTRIBUTIONS AND THE EFFECTS OF HABITAT FRAGMENTATION: THE CASE OF THE GIANT PANDA

By

Thomas Connor

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Fisheries and Wildlife – Doctor of Philosophy
Ecology, Evolutionary Biology, and Behavior – Dual Major

2019

ABSTRACT

ASSESSING SPECIES DISTRIBUTIONS AND THE EFFECTS OF HABITAT FRAGMENTATION: THE CASE OF THE GIANT PANDA

By

Thomas Connor

Environmental degradation has become a ubiquitous feature in the modern world. This degradation is resulting in widespread loss and fragmentation of wildlife habitat, leading to increased extinction risks and population declines. In order to stem these threats, it is imperative to accurately predict species' habitats, how to optimize the restoration and protection of these habitats, and better understand how their ecology interacts with habitat requirements. I used both simulated and empirical study systems to investigate these topics and focused heavily on giant panda (*Ailuropoda melanoleuca*) populations across Wolong Nature Reserve and Southwest China.

Given the uncertainty and debate surrounding the relative effects of habitat amount and habitat fragmentation on ecological responses (Chapter 1), I set out to accurately define habitat before further investigating these effects. I found that the grain size of environmental predictor variables had important effects on modeling the distributions of virtual species simulated on real landscapes, and that modeling with grain sizes farther away from the "true" grain size used to simulate the species resulted in lower predictive accuracy and incorrect ecological inferences about the importance of environmental variables to habitat (Chapter 2). I then went a step further and investigated interactive spatial scale effects on species distribution modeling by varying the total extent of the study areas and grain size of the environmental variables used to predict panda habitat and distributions across Southwest China (Chapter 3). I found that increasing total extent offset the negative effects of increasing grain size on model accuracy and that total extent can be

optimized as both larger (at smaller spatial scales) and smaller (at the geographic range scale) than the study area of interest. I then further improved the accuracy of our species habitat and distribution modeling by leveraging empirical movement distributions derived from GPS-collar data to transform the environmental predictor variables, and used the resulting habitat map to investigate the effects of habitat amount and fragmentation on functional connectivity in the panda population in Wolong (Chapter 4). I found that the standard deviation of the core area index, a measure of habitat configuration, was the best predictor of functional connectivity. Habitat amount was the second-best predictor and we found that in our study system it could optimized to cover about 80% of a local landscape to maximize functional connectivity. Habitat fragmentation also showed a nonlinear and threshold-dependent relationship with functional connectivity – important findings to consider in the spatial planning of protected areas. Finally, I used the noninvasive genetics data to "capture" and "recapture" unique individuals across a core habitat area in Wolong and conduct the first social network analysis of pandas (Chapter 5). I found strong evidence of two to three social clusters in the population, defined as groups of pandas that associated with each other at a significantly higher rate than individuals outside the cluster. These clusters may represent cryptic family structuring, as genetic relatedness was a significant positive predictor of associations between individuals. My detailed approaches to investigating the habitat and ecology of giant pandas used throughout this manuscript resulted in unique insights into this threatened habitat specialist species, and we recommend they be applied widely to other species. Optimizing the way in which we predict and conserve habitat in each landscape and system, as opposed to relying on expert opinion or competing theories, will be increasingly important as environmental degradation continues in the Anthropocene.

This dissertation is dedicated to my parents, James Connor and Sharon Kahkonen, as well as my partner, Erin Tracy. Thank you for the constant support through the good and bad times

ACKNOWLEDGEMENTS

This dissertation would not have been completed without the help of many others throughout my life and particularly the last four-and-a-half years. I thank my parents for their loving support and valuing my education from a young age. I thank my brother for the lifelong camaraderie and friendship, as well as my old friends that I have known nearly as long. Our explorations of the woods, creeks, and gorges the Fingerlakes region of New York State left a long and lasting impression on me that heavily influenced my choice to study and conserve the natural world. I would also like to thank the many Professors and classmates that I met and learned from at Tompkins Cortland Community College and then Cornell University where I took undergraduate courses that prepared me well for pursuing my PhD. I must also acknowledge the many colleagues I met and worked with in the three years that I jumped from field job to field job across the country (and not forgetting those I met in the six months I volunteered to monitor greater bamboo lemur behavior in Madagascar!).

All scholarly preparations and field experiences aside, starting a PhD felt like a baptism by fire. I would not have stayed afloat without the support of my adviser, Dr. Jianguo (Jack) Liu. He has always encouraged me in generating ideas and supported the research that followed, as well as helped me extract and best express the broader meaning of that research. I am also indebted to the members of my guidance committee: Dr. Ashton Shortridge, Dr. Kim Scribner, and Dr. Ken Frank, who have each helped me in the research and writing involved with at least one of the chapters contained herein. They also taught courses that I count among my favorite that I have taken in my entire academic career. I thank my colleagues and friends in China, where I spent nearly two years of my life conducting the research for this dissertation, namely

Dr. Jindong Zhang and his dedicated students. Without the help of Yang Wenbin and Yang Hong in the rugged mountains of Wolong, I would not have completed the fieldwork needed for my dissertation and may have come to great harm. I thank them immensely.

I must also thank the institutions and agencies that have made this research possible with their generous funding, including Michigan State University, the National Science Foundation, the US Department of State through the Fulbright Program, and the US Department of Education through the Foreign Language and Area Studies Program. I would also like to thank my many lab mates and classmates who have helped me with their friendship and advice. My good friend, housemate, and fellow PhD student Sean Sultaire is deserving of special thanks for the guidance and camaraderie through the years. Finally, I would like to thank my partner Erin Tracy, who has been a source of immense love and support even when slaving over her own research and even when we were separated by thousands of miles.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Background:

Environmental degradation has become one of the defining issues of our time (Lewis and Maslin 2015). In an increasingly industrialized and globalized world, the scale of anthropogenic global change has reached unprecedented levels (Dirzo et al. 2014). A major component of this change has been the widespread loss and fragmentation of habitat for flora and fauna, which has contributed to species declines, endangerment, and extinctions worldwide (Krauss et al. 2010). Because anthropogenic effects on wildlife and their habitat have become ubiquitous, an adequate understanding of ecological responses to these effects in order to inform conservation action has become increasingly important (Scott et al. 2010, Caro et al. 2012)

Although the negative effects of habitat loss on wildlife populations and biodiversity are widely acknowledged, the ecological effects of habitat fragmentation debated. Some researchers have argued that there are clear, overall negative effects of fragmentation on biodiversity/ecological metrics (Haddad et al. 2015), while others have argued that there are varied and perhaps even more positive than negative effects of fragmentation (Fahrig 2017). There have even been conflicting results about the effects of habitat fragmentation of specific ecological responses, such as the functional connectivity (measured as spatial genetic structure) of wildlife across a landscape (Keyghobadi 2007). This disagreement is present even in purely simulation studies. Specifically, Jackson and Fahrig (2016) found habitat amount to be the driving force for spatial genetic structure and that fragmentation had a minor effect that reduced genetic structure across space, while Cushman et al. (2012) found that habitat fragmentation was main cause of increased genetic structure across space compared to habitat amount. The relative effects of habitat loss and habitat fragmentation on this ecological response are thus especially uncertain. Maintaining functional connectivity among wildlife populations distributed across a

landscape is essential to promote their long-term survival through the avoidance of deleterious inbreeding effects (Reed et al. 2003), the maintenance of the genetic diversity necessary to effectively adapt to changing conditions (Allendorf et al. 2013), and the ability to recolonize habitat areas after localized extirpations (Hanski 1998).

Because many of the effects commonly associated with habitat fragmentation, such as edge effects, can have substantially different effects on different species and landscapes (Pfeifer et al. 2017), it follows that habitat fragmentation in general will likewise have varied effects. For example, habitat specialists have been found to be more sensitive to fragmentation compared to generalists (Devictor et al. 2008). If we consider functional connectivity across a habitat-matrix system, one can imagine that habitat specialists are likely less able to disperse through non-habitat areas compared to generalists that may be able to better utilize resources and survive in the intervening matrix. The conservation of habitat specialists may thus be especially challenging and important to consider as habitat loss and fragmentation continue to occur. I hypothesized that in general, increasing habitat fragmentation from a non-fragmented landscape would increase positive ecological responses (like functional connectivity) until a threshold, at which point further habitat fragmentation would result in decreases in the positive ecological response being studied (Fig. 1.1). Considering functional connectivity, I hypothesize that more fragmented habitat will "force" individual to disperse further on the landscape and thus increase its functional connectivity until a threshold at which further fragmentation eliminates or severely reduces the ability of individuals to successfully disperse between habitat patches on the landscape.

In order to better understand the relative importance and effects of habitat loss and fragmentation and effectively take conservation action in the face of these threats, it is first

necessary to accurately quantify habitat. A common approach to do this is to statistically model the effect of environmental variables on species presence and absence and predict the probability of presence across the landscape as defined by those environmental variables (Elith and Leathwick 2009). This process involves important spatial-scale considerations, such as the grain size of environmental variables/presence data and total extent of analysis to use in training the model, that are frequently not optimized and have been under-addressed in the literature (McGarigal et al. 2016). Once habitat is defined, mapped, and summarized at a spatially relevant scale, habitat fragmentation metrics must be selected that are both uncorrelated with habitat amount and are biologically meaningful to the species of interest and its successful dispersal across the landscape. It is then necessary to measure this functional connectivity in some manner as a response variable and model the effect habitat amount and habitat fragmentation on it in a quantitative manner. Better understanding these effects, particularly on the functional connectivity of habitat specialist species, will be vital for their effective management through conservation planning and protected area-design. Aspects of a species' ecology that are understudied, such as the level of sociality in solitary species and if there are certain areas that are important to associations within a given population, may also be important to prioritize conservation efforts. This research is particularly timely given the ambitious goal to protect half of the earth's current land surface by 2050 that is gaining widespread attention (Ellis and Mehrabi 2019).

1.2 Research goals

In order to measure the effects of both habitat amount and fragmentation on the functional connectivity of a habitat specialist, I conducted a case study of the giant panda and followed a series of research steps. With the goal of accurately determining giant panda habitat

and distributions, I also aimed to fill a knowledge gap regarding spatial scale and investigated the effects of grain size and total extent on species distribution model results. These research steps were completed sequentially in the chapters of this dissertation which targeted the following, explicit goals.

1. Determine the effects of grain size and niche breadth on species distribution and habitat modeling in order to produce the most accurate and biologically meaningful classifications of habitat and non-habitat.
2. Further explore the effects of spatial scale, defined here as the total extent of analysis and grain size of the presence response and environmental predictor variables, on species distribution and habitat modeling.
3. Use output from optimized habitat classifications developed in Chapter 2 to measure habitat amount and habitat fragmentation across a landscape and determine their effects on the functional connectivity of a habitat specialist species across that landscape.
4. Better understand panda behavioral ecology in contiguous habitat and how this may be affected by habitat fragmentation.

Chapter 2 is focused on the first goal and involved simulating virtual species distributions across real landscapes in Scandinavia through controlling their habitat suitability responses to three environmental variables. Chapter 3 is focused on the second goal and involved an empirical analysis of giant panda habitat at multiple total spatial extents and grain sizes, with extents ranging from 50 km^2 to the entire current range for the species and grain sizes varied from 30 m to 1920 m. Chapter 4 employs the lessons gleaned from the first two chapters to derive habitat classifications across a key giant panda nature reserve in southwest China in order to relate the effects of habitat amount and habitat fragmentation, as well as other environmental variables, on

the functional connectivity of giant pandas in a landscape genetics analysis. Chapter 5 takes a step back and uses the noninvasive genetics information from Chapter 4 to investigate evidence of social networks in a study area containing contiguous habitat within Wolong.

1.3 Study system:

1.3.1 Giant panda ecology:

Except for Chapter 1, the simulated study system of which is justified within that chapter, my dissertation focuses on wild giant panda (*Ailuropoda melanoleuca*) habitat and populations across southcentral China. Giant pandas made an ideal study species for my research for the following reasons: 1. They exclusively rely on understory bamboo forests for habitat and are thus a habitat specialist species; 2. This habitat can be readily measured via remotely sensed and other available variables; 3. My research team had access to a wealth of range-wide occurrence data seldom seen for elusive mammals; 4. Individuals can be relatively easily and noninvasively genotyped across the landscape via the collection of feces and extraction/amplification of their DNA in the laboratory.

Because the bulk of this dissertation uses giant pandas as a case study species, I will briefly introduce important aspects of their ecology. Giant pandas have several unique biological and life history adaptations. The most ancient lineage in the family *Ursidae*, they are the only carnivore adapted to an almost entirely herbivorous diet. Specifically, 99% of wild panda diets are comprised of a suite of understory bamboo species (Schaller et al. 1985). Although they only digest an average of less than 20% of consumed bamboo (Dierenfield et al. 1982), these bamboo species are widely distributed and fast-growing (Vina et al. 2010). Although Nie et al. (2015) concluded that pandas have exceptionally low energy expenditures due to this inefficient digestion of their primary food source, Fei et al. (2015) refuted those results and found that

resting metabolic rates are only slightly below what would be expected for a panda-sized mammal and that active metabolic rates are in a normal range. They cited several issues with the methodology of the Nie et al. (2015), and their conclusions support the relatively high activity levels and mobility inherent to giant panda ecology and manifested in several activity peaks throughout the day and night, large mating season movements, and movements between habitat patches during bamboo flowering events (Reid et al. 1989, Connor et al. 2016, Zhang et al. 2016).

Pandas also have life-history traits that are likely driven by the low-energy value of bamboo, however. For example, female giant pandas enter estrus for only a few days sometime in April/May and gestate one to two offspring, which are born in an underdeveloped state and only one of which are nurtured (Schaller et al. 1985, Pan et al. 2014). At 1.5 to 2 years of age juvenile pandas reach subadult status and leave their mothers to establish independent home ranges (Pan et al. 2014). Both genetic and relocation data indicate that subadult females disperse from their natal ranges while subadult males stay within or nearby their natal ranges (Zhan et al. 2007, Connor et al. 2016). Telemetry evidence suggests that females then establish discrete home ranges spatially segregated from other females, while males hold larger home ranges that overlap other the ranges of other males and females (Connor et al. 2016). Pandas reach sexual maturity at about 5 years of age and can breed until their natural lifespan in the wild is reached at about age 15-20 (Pan et al. 2014). Wild populations feature approximately even sex ratios (see Chapter 5), and since females can produce a cub roughly every two years, the maximum population growth due to fecundity is about (Total Population Size/2)/2 individuals per year. Although survival rates were likely historically affected by predation due to the presence of tigers and higher

densities of leopards throughout their range, there is little evidence for predation having a large effect on current giant panda populations.

1.3.2. Study areas:

Chapter 3 includes study areas across the entire giant panda geographic range, while Chapters 4 and 5 focus the giant panda population within Wolong Nature Reserve (see Chapter appendices for maps of specific study areas). Although fossil evidence indicate that giant pandas historically occurred across much of China as well as parts of Myanmar, Vietnam, Laos, and Thailand (Hu 2001), a severe range contraction in recent centuries has resulted in an extant distribution across six isolated mountain ranges in Southwest China. Though this most recent range retraction has been linked most strongly to anthropogenic encroachment, changes in climate are also a factor in shifts in the panda's range (Li et al. 2015). There is genetic evidence for a particularly severe prehistoric population bottleneck in the last ice age, at which time the panda geographic range was likely even more restricted then it is today (Chen et al. 2013).

The current panda range in Southwest China is a global biodiversity hotspot, with thousands of plant and animal species occupying the diverse habitats spanning the large elevation gradients that occur as the land rises from the Sichuan basin to the Tibetan plateau (Myers 2000). The panda populations currently residing in the six isolated mountain ranges mentioned above are thought to be further segregated into a total of 33 subpopulations due to habitat fragmentation and both natural and anthropogenic barriers (State Council 2015), though there is debate concerning the level of connectivity featured across these barriers. For example, Qiao et al. (2019) recently found evidence of relatively rates of gene flow across a national road in Wolong Nature Reserve, which was previously used to subdivide the panda population found in the Qionglai Mountains.

Wolong itself is a national-level nature reserve covering about 2000 km^2 in the central Qionglai. A large range in elevations spanning 1700 to over 6000 meters make for diverse habitats from subtropical rainforests to alpine meadows and screes. Three main understory bamboo species - *F. spacathea*, *F. robustus,* and *Y. bravipaniculata* - occur in large aggregations in the understory of deciduous, evergreen, and mixed deciduous-evergreen forests from around 1800 to nearly 4000 meters. These three species are the staple food for pandas in Wolong, and their distribution make up the main panda habitat in the reserve (Reid et al. 1989). The entire panda population and habitat within Wolong is the focal system for Chapter 3. Chapter 4 takes a more detailed look at an approximately 50 km^2 area within Wolong known as Hetaoping which features particularly good panda habitat and population density (Hull et al. 2015).

APPENDIX

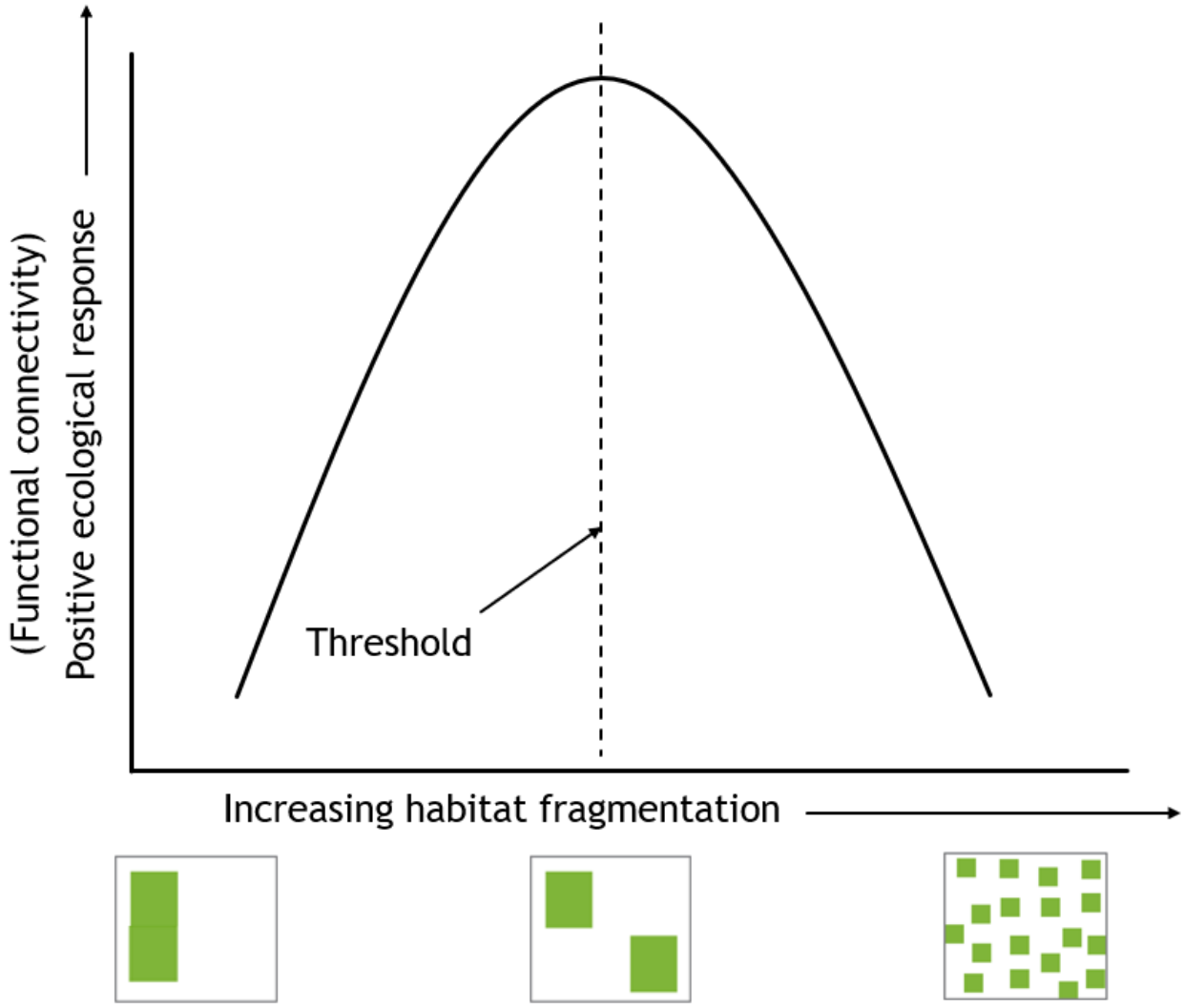

Figure 1.1. The central hypothesis of the dissertation, which states that as habitat fragmentation increases from 0, there will be increases in some positive ecological response until a threshold at which further fragmentation will decrease the positive ecological response.

REFERENCES

REFERENCES

Allendorf, F. W., Luikhart, G. H., and S. Aitken. 2013. Conservation and the Genetics of Populations. Wiley-Blackwell, West Sussex, UK.

Caro, T., J. Darwin, T. Forrester, C. Ledoux-Bloom, and C. Wells. 2012. Conservation in the Anthropocene. Conservation Biology **26**:185-188.

Chen, Y. Y., Y. Zhu, Q. H. Wan, J. K. Lou, W. J. Li, Y. F. Ge, and S. G. Fang. 2013. Patterns of Adaptive and Neutral Diversity Identify the Xiaoxiangling Mountains as a Refuge for the Giant Panda. Plos One **8**.

Connor, T., V. Hull, and J. G. Liu. 2016. Telemetry research on elusive wildlife: A synthesis of studies on giant pandas. Integrative Zoology **11**:295-307.

Cushman, S. A., A. Shirk, and E. L. Landguth. 2012. Separating the effects of habitat area, fragmentation and matrix resistance on genetic differentiation in complex landscapes. Landscape Ecology **27**:369-380.

Devictor, V., R. Julliard, and F. Jiguet. 2008. Distribution of specialist and generalist species along spatial gradients of habitat disturbance and fragmentation. Oikos **117**:507-514.

Dierenfeld, E. S., H. F. Hintz, J. B. Robertson, P. J. Vansoest, and O. T. Oftedal. 1982. UTILIZATION OF BAMBOO BY THE GIANT PANDA. Journal of Nutrition **112**:636-641.

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.

Ellis, E. C., and Z. Mehrabi. 2019. Half Earth: promises, pitfalls, and prospects of dedicating Half of Earth's land to conservation. Current Opinion in Environmental Sustainability **38**:22-30.

Fahrig, L. 2017. Ecological Responses to Habitat Fragmentation Per Se. Pages 1-23 *in* D. J. Futuyma, editor. Annual Review of Ecology, Evolution, and Systematics, Vol 48. Annual Reviews, Palo Alto.

Fei, Y. X., R. Hou, J. R. Spotila, F. V. Paladino, D. W. Qi, and Z. H. Zhang. 2016. Metabolic rates of giant pandas inform conservation strategies. Scientific Reports **6**.

Haddad, N. M., L. A. Brudvig, J. Clobert, K. F. Davies, A. Gonzalez, R. D. Holt, T. E. Lovejoy, J. O. Sexton, M. P. Austin, C. D. Collins, W. M. Cook, E. I. Damschen, R. M. Ewers, B. L. Foster, C. N. Jenkins, A. J. King, W. F. Laurance, D. J. Levey, C. R. Margules, B. A. Melbourne, A. O. Nicholls, J. L. Orrock, D. X. Song, and J. R. Townshend. 2015. Habitat fragmentation and its lasting impact on Earth's ecosystems. Science Advances **1**:9.

Hanski, I. 1998. Metapopulation dynamics. Nature **396**:41-49.

Hu, J.C. 2001. Research on the giant panda. Shanghai Publishing House of Science and Technology, Shanghai:

Hull, V., J. D. Zhang, S. Q. Zhou, J. Y. Huang, R. G. Li, D. Liu, W. H. Xu, Y. Huang, Z. Y. Ouyang, H. M. Zhang, and J. G. Liu. 2015. Space use by endangered giant pandas. Journal of Mammalogy **96**:230-236.

Jackson, N. D., and L. Fahrig. 2016. Habitat amount, not habitat configuration, best predicts population genetic structure in fragmented landscapes. Landscape Ecology **31**:951-968.

Keyghobadi, N. 2007. The genetic implications of habitat fragmentation for animals. Canadian Journal of Zoology **85**:1049-1064.

Krauss, J., R. Bommarco, M. Guardiola, R. K. Heikkinen, A. Helm, M. Kuussaari, R. Lindborg, E. Ockinger, M. Partel, J. Pino, J. Poyry, K. M. Raatikainen, A. Sang, C. Stefanescu, T. Teder, M. Zobel, and I. Steffan-Dewenter. 2010. Habitat fragmentation causes immediate and time-delayed biodiversity loss at different trophic levels. Ecology Letters **13**:597-605.

Lewis, S. L., and M. A. Maslin. 2015. Defining the Anthropocene. Nature **519**:171-180.

Li, X. H., G. S. Jiang, H. D. Tian, L. Xu, C. Yan, Z. W. Wang, F. W. Wei, and Z. B. Zhang. 2015. Human impact and climate cooling caused range contraction of large mammals in China over the past two millennia. Ecography **38**:74-82.

McGarigal, K., K. A. Zeller, and S. A. Cushman. 2016. Multi-scale habitat selection modeling: introduction to the special issue. Landscape Ecology **31**:1157-1160.

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.

Nie, Y. G., J. R. Speakman, Q. Wu, C. L. Zhang, Y. B. Hu, M. H. Xia, L. Yan, C. Hambly, L. Wang, W. Wei, J. G. Zhang, and F. W. Wei. 2015. Exceptionally low daily energy expenditure in the bamboo-eating giant panda. Science **349**:171-174.

Pan, W. 2014. A chance for lasting survival: Ecology and behavior of wild giant pandas. Smithsonian Institution.

Pfeifer, M., V. Lefebvre, C. A. Peres, C. Banks-Leite, O. R. Wearn, C. J. Marsh, S. H. M. Butchart, V. Arroyo-Rodriguez, J. Barlow, A. Cerezo, L. Cisneros, N. D'Cruze, D. Faria, A. Hadley, S. M. Harris, B. T. Klingbeil, U. Kormann, L. Lens, G. F. Medina-Rangel, J. C. Morante-Filho, P. Olivier, S. L. Peters, A. Pidgeon, D. B. Ribeiro, C. Scherber, L. Schneider-Maunoury, M. Struebig, N. Urbina-Cardona, J. I. Watling, M. R. Willig, E. M. Wood, and R. M. Ewers. 2017. Creation of forest edges has a global impact on forest vertebrates. Nature **551**:187-+.

Qiao, M. J., T. Connor, X. G. Shi, J. Huang, Y. Huang, H. M. Zhang, and J. H. Ran. 2019. Population genetics reveals high connectivity of giant panda populations across human disturbance features in key nature reserve. Ecology and Evolution **9**:1809-1819.

Reed, D. H., E. H. Lowe, D. A. Briscoe, and R. Frankham. 2003. Inbreeding and extinction: Effects of rate of inbreeding. Conservation Genetics **4**:405-410.

Reid, D. G., J. C. Hu, D. Sai, W. Wei, and Y. Huang. 1989. GIANT PANDA AILUROPODA-MELANOLEUCA BEHAVIOR AND CARRYING-CAPACITY FOLLOWING A BAMBOO DIE-OFF. Biological Conservation **49**:85-104.

Schaller, George B. 1985. Giant pandas of Wolong. University of Chicago press, Chicago.

Scott, J. M., D. D. Goble, A. M. Haines, J. A. Wiens, and M. C. Neel. 2010. Conservation-reliant species and the future of conservation. Conservation Letters **3**:91-97.

State Council Information Office. 2015. The State Forestry Administration held the press conference for the fourth national giant panda survey results (In Chinese).

Vina, A., M. N. Tuanmu, W. H. Xu, Y. Li, Z. Y. Ouyang, R. DeFries, and J. G. Liu. 2010. Range-wide analysis of wildlife habitat: Implications for conservation. Biological Conservation **143**:1960-1969.

Zhan, X. J., Z. J. Zhang, H. Wu, B. Goossens, M. Li, S. W. Jiang, M. W. Bruford, and F. W. Wei. 2007. Molecular analysis of dispersal in giant pandas. Molecular Ecology **16**:3792-3800.

Zhang, J. D., V. Hull, Z. Y. Ouyang, L. He, T. Connor, H. B. Yang, J. Y. Huang, S. Q. Zhou, Z. J. Zhang, C. Q. Zhou, H. M. Zhang, and J. G. Liu. 2017. Modeling activity patterns of wildlife using time-series analysis. Ecology and Evolution **7**:2575-2584.

CHAPTER 2: EFFECTS OF GRAIN SIZE AND NICHE BREADTH ON SPECIES DISTRIBUTION MODELING

In collaboration with

Vanessa Hull, Andrés Viña**,** Ashton Shortridge, Ying Tang, Jindong Zhang, Fang Wang, Jianguo Liu

Abstract

Scale is a vital component to consider in ecological research, and spatial resolution or grain size is one of its key facets. Species distribution models (SDMs) are prime examples of ecological research in which grain size is an important component. Despite this, SDMs rarely explicitly examine the effects of varying the grain size of the predictors for species with different niche breadths. To investigate the effect of grain size and niche breadth on SDMs, we simulated four virtual species with different grain sizes/niche breadths using three environmental predictors (elevation, aspect, and percent forest) across two real landscapes of differing heterogeneity in predictor values. We aggregated these predictors to seven different grain sizes and modeled the distribution of each of our simulated species using MaxEnt and GLM techniques at each grain size. We examined model accuracy using the AUC statistic, Pearson's correlations of predicted suitability with the true suitability, and the binary area of presence determined from suitability above the maximum True Skill Statistic (TSS) threshold. Habitat specialists were more accurately modeled than generalist species, and constructed at the grain size from which a species was derived generally performed the best. The accuracy of models in the homogenous landscape deteriorated with increasing grain size to a greater degree than models in the heterogenous landscape. Variable effects on the model varied with grain size, with elevation increasing in importance as grain size increased while aspect lost importance. The area of predicted presence was drastically affected by grain size, with larger grain sizes over predicting this value by up to a factor of 14. Our results have implications for species distribution modeling and conservation planning, and we suggest more studies include analysis of grain size as part of their protocol.

2.1 Introduction

Spatial scale has long been recognized as a vital component of ecological research, with complex effects that vary by species and system (Wiens 1989). Specifically, the grain of the sample and the extent of the study area are the components of spatial scale that must be defined in the research and management of an ecological phenomenon (Levin 1992). Spatial scale is important to consider in species distribution models (SDMs), an increasingly widely used suite of methods for both ecological research and conservation management. SDMs have been applied to a large range of taxa to estimate occurrence across landscapes and seascapes (see review in Guisan and Thuillier 2005, Leipold et al. 2017). They can aid in understanding of species-habitat relationships and delineating species distributions. They also have direct application to prioritizing areas for protection and can model future effects of land-use and climate change on species habitat and distributions (Porfirio et al. 2009). SDMs use environmental predictors combined with species presence and absence or background points to make these estimations (Elith and Leathwick 2009). Although building SDMs only requires predictor values at species presence and absence/background locations, making predictions about species occurrence across a landscape requires spatially explicit environmental predictors. The continuing improvement of remote sensing technology, such as Light Detection and Ranging (LiDAR, Hudak et al. 2009) methods, has resulted in higher resolution environmental variables available to be included as predictors in SDMs.

Despite this increase in the variety of available data resolutions, research on the effect of grain size on SDMs is scant and usually limited to comparing three or fewer grain sizes (Guisan et al. 2007, Revermann et al. 2012, Bean et al. 2014). There is also some uncertainty within the existing body of literature on the effects of increasing grain size of environmental predictors on

SDMs. Guisan et al. (2007) found that a 10-fold increase in grain size degraded the accuracy (as measured by AUC) of SDMs of some species, but did not result in large changes and in fact improved AUC in other species. Additional studies have found small effects of increasing grain size on SDMs of birds (Seoane et al. 2004), generally concluding that coarser grains capture the necessary habitat variation. A study of bird habitat occupancy in Germany found that grain size was important in model performance, however, with several species modeled with the highest AUC at a fine resolution of 1 m (Gottschalk et al. 2011). A recent investigation of marine predator habitat at six grain sizes ranging from 3 to 111 km found that finer resolution variables improved models unless data were missing due to cloud cover (Scales et al. 2017).

Effects of grain size on SDMs at a large spatial scale have also been investigated on plant species, with applications to climate change projections. Seo et al. (2009) analyzed the effect of increases in grain size (seven grains from 1 km to 64 km) on SDMs of several tree species in California and found that AUC declined while the predicted area of occurrence increased. A similar relationship was identified by Trivedi et al. (2008) in a comparison of similar SDMs on plants under climate change. They concluded that coarse models in rugged terrain likely overestimate the ability of species to persist under climate change. Another recent study likewise found that total predicted area of presence was greater at coarser grain sizes (4 km vs. 800 m data), but important potential climate refugia were detected at a 90-m grain that were lost at the 4-km grain size (Franklin et al. 2013).

This wide range in findings indicates a significant lack of consensus regarding the importance of grain size to SDMs. There is also a lack of research into the effects of grain size on models of species that utilize a narrow range of habitat conditions (specialists) vs. species that utilize a wider range of habitat (generalists), and the effects of grain size on models of species in

landscapes of varying heterogeneity. These questions are difficult to answer because the true habitat responses of any given species, and consequently the true suitability of an area to that species, cannot be known. Because of this, the simulation of virtual species across a landscape, using predefined responses to habitat variables, is increasingly being used to answer questions related to ecology and the modeling of species distributions. Examples of previous research using this technique include simulation of species to compare different SDM modeling techniques (Elith et al. 2009), effects of prevalence and sampling bias on SDMs (Jimenez-Valverde et al. 2009), and effects of different methods for selecting pseudo-absences on SDMs (Barbet-Massin et al. 2012). To investigate the effects of grain size and niche breadth on SDMs, we adapted this simulation approach to model habitat suitability for a set of four virtual species across two landscapes of contrasting heterogeneity. We hypothesized that the best models would be built at the grain size at which the species was simulated (defined in methods, hereafter referred to as "correct" grain size). We predicted that small changes in grain size (1 aggregation, or doubling of grain size, i.e. 25 m to 50 m) would not have significant effects on model performance, but larger changes (2 aggregations or greater, i.e. 25 m to 100 m or more) would significantly deteriorate model accuracy. We also hypothesized that changes in grain size would more greatly affect models of the narrow-niched (specialist) species as compared to the wide-niched (generalist) species.

2.2 Methods

We selected two landscapes of approximately 40,000 km^2 upon which to simulate our species. One encompasses the southeastern range of the Scandinavian Mountains in Norway and Sweden, while the other lies in southern Finland. We chose these areas because of the availability of environmental predictors with small grain size and large differences in variance of

the included predictor variables between the two (one “homogenous” landscape and one “heterogeneous” landscape). These predictor variables were elevation, aspect, and percent forest. In the heterogeneous landscape a random draw of 1,000,000 cells resulted in predictor values with standard deviations of 368.48, 109, and 33.96 for elevation, aspect, and percent forest, respectively, while the homogenous landscape had corresponding values of 34.03, 113.54, and 27.35. Additionally, all pair-wise combinations of these three variables had Pearson’s correlation coefficient values of less than 0.3, suggesting that multi-collinearity was not an issue. (Dormann et al. 2012). The terrain model and a tree cover density raster, which we labeled percent forest for the simulation of our species, were obtained from the European Environment Agency’s Copernicus Land Monitoring Service (DEM resolution = 25 m, tree cover density = 20 m, EU 2016). Elevation and aspect were derived from the EUDEM dataset, a digital elevation model produced by hybridizing data from the SRTM and GDEM missions through a weighted averaging approach. The tree cover density raster was resampled to 25-m resolution. We then aggregated the environmental variables to larger grain sizes by calculating the mean values within moving windows with sizes of 50, 100, 200, 400, 800, and 1600 meters.

We created two virtual species whose environmental suitability and ultimately presence on the landscape depended on these three variables (Fig. 2.1). Suitability responses were kept to normal distributions over the range of each variable. These were then combined in a simple multiplicative formula to come to a final suitability response for the species: percent forest * elevation * aspect. We then rescaled suitability values for each pixel between 0 and 1 for a final reference probabilistic suitability map on which to later test models against. We created one species with a narrower range of suitability across each of the variables (hereafter referred to as specialist), and one species with a wider range (standard deviation of response distribution

doubled) of suitability across each of these variables (hereafter referred to as generalist, Fig. 2.1). We built two species with each of these suitability responses, one from the environmental variables at a fine scale (grain size of 25 meters) and the other from the environmental variables at a coarser scale (200 meters). This was performed to investigate the effect of using grain sizes that are both smaller and larger than the "correct" grain size at which a species responds. In the homogenous landscape, the suitable elevation values had to be decreased due to the large differences in elevation between the two landscapes, but otherwise the suitability responses were kept the same.

After calculating environmental suitability values for each species, we used a logistic curve of probability pivoting around a suitability threshold of 0.5 to generate presence and absence values throughout the study area. In this process, a cell with a suitability value of 0.7 is more likely to become a presence cell than a cell with a suitability value of 0.4, but the latter cell does have a chance to become a presence cell according to the probability derived from the logistic curve. We then randomly selected 2,000 points across each landscape for both the generalist and specialist species to use as presence/absence data input for the SDMs, the results of which are summarized in Table 2.1.

We modeled species' distributions using two different methods across all grain sizes. These were Maximum Entropy (MaxEnt) (Phillips et al. 2006) and Generalized Linear Modelling (GLM). Although MaxEnt has been argued to be equivalent to a Poisson point-process model and thus can be viewed as a type of regression (Renner and Warton 2013), these two techniques have been used extensively and produce distinct results (Merow et al. 2013, Elith et al. 2009), warranting their inclusion and evaluation. We divided the presence data into training

(80%) and validation (20%) sets using a k-fold partitioning design, with k=5, and used the same method on the absence points for the GLMs.

MaxEnt – MaxEnt is a machine learning method that has gained wide popularity in species distribution modeling. It uses presence-only data and environmental predictor variables to come up with a distribution of probability that minimizes the relative entropy in the predicted suitability values at the presence vs. background points (Elith et al. 2011). We used the k-fold partitioned sets of presence points described above as input into the models of the species they corresponded to. We randomly selected an additional 10,000 random background points to serve as pseudo-absences. All other model parameters were left at their default values. In the MaxEnt models we calculated variable permutation importance at each grain size. This was determined by changing the values of each variable among the training presence and background points and measuring the loss in AUC. We did this for each variable separately, and the final values were normalized to percentages.

GLM – We used the k-fold partitioned presence and absence points from the generated presence absence map as input for our GLMs, which we built for each virtual species at each grain size. We included cubic, quadratic, and linear terms for each environmental variable in multiplicative and additive relationships and used the 'step' function to remove variables until the optimum model was reached based on AIC. The GLMs were fit with a "binomial" family and the "identity" link function.

We calculated the Area Under Receiving Operating Characteristic Curve (AUC), which measures the ability to discriminate between observed presence and absence, for each model. We then randomly chose 10,000 points across the study area to extract model suitability predictions and ran Pearson's correlation tests between model predictions and the "true" suitability across

the same points. To analyze the effects of grain size on predicted area of presence, we made presence/absence range maps based on threshold at which the value of the True Skill Statistic (TSS) was maximized (Liu et al. 2013). The TSS measures model sensitivity and specificity and unlike kappa is independent of prevalence (Allouche et al. 2006). We calculated the area of presence predicted by each model based on this threshold and compared it to the true simulated presence area. Our reported results across all analyses are the average of all five training/evaluation runs for each model.

In order to test for significant differences in model accuracy (measured by AUC) of both the MaxEnt models and GLMs separately at different grain sizes, we used a Van der Waerden normal scores test on the AUC outputs of each training/testing run. There were thus 5 AUC scores for each of 7 different models across 4 model/species combinations. The Van der Waerden test is an effective alternative to Tukey's or Analysis of Variance (ANOVA) when data is not normally distributed (Conover 1999). If significance was found in this test, we then performed pairwise comparisons using Van der Waerden normal score tests to look at individual comparisons of the different grain sizes in the virtual species being analyzed.

All analyses were conducted using the R software platform (R Core team 2016). We used the package 'virtualspecies' (Leroy 2015) to create the virtual species, the 'dismo' package (Hijmans et al. 2011) for MaxEnt and model comparison analyses, the 'glm' function to build the GLMs, and the 'PMCMR' package to conduct the Van der Waerden tests.

2.3 Results

The performance of our models varied with grain size, with generally decreasing AUC values as grain size was varied further from the correct grain size for the given species (Table 2.2, Fig. 2.2). These decreases in AUC were often significant within 1 aggregation (i.e. 25 m vs. 50 m), and almost always significant at 2 aggregations or more (i.e. 25 m vs. 100 m, Table 2.2). The major exception to this was the MaxEnt models of the coarse-scale habitat specialist, in which AUC values were significantly higher at 50 m and 100 m, compared to the correct grain size of 200 m. AUC values were also higher in the heterogeneous landscape compared to the homogeneous landscape, higher in the habitat specialist compared to the habitat generalist, and generally higher in the MaxEnt models compared to the GLMs (Fig. 2.2). Increasing grain size deteriorated model accuracy in the homogeneous landscape to a greater degree than in the heterogeneous landscape (Fig. 2.2).

The Pearson's correlation coefficients between predicted suitability and true suitability in all the derived species were highest in the models built at the correct grain sizes and decreased drastically as grain size increased and/or decreased from these optimum models (Table 2.2, Fig. 2.3). In the fine-scale habitat specialist Pearson's r decreased from a high of 0.95 in the MaxEnt model built at the "correct" grain size of 25 m to a low of 0.47 in the MaxEnt model built at the 1600-m grain (Fig. 2.3). These decreases with changes in grain size were much less pronounced in the GLMs, but predictions from the GLMs of all four species had lower correlations with the true suitability values than the MaxEnt models (Table 2.2). Similar to the AUC values, model predictions of the habitat specialist generally had higher correlations with the true suitability than model predictions of the habitat generalist. Models of the generalist improved relative to the specialist as the distance from the optimum 200-m grain increased, however, and often surpassed

them in Pearson's correlation coefficient at just one aggregation from the correct grain size (Table 2.2).

The permutation importance of aspect, elevation, and percent forest to the MaxEnt models varied with grain size – in the fine-grain versions of both the generalist and specialist in the heterogeneous landscape the variable percent importance was fairly even in the model at the correct (25-m) grain size. As grain size increased, the discrepancy between variables' importance also increased, with elevation gaining importance and aspect losing importance (Fig. 2.4a,b). Although the variables were further spread apart in their permutation importance values in the coarse-grain generalist species, a similar pattern emerged with the model at the correct (200-m) grain size having a more even distribution. In the homogeneous landscape, the permutation importance of the predictor variables were substantially different, with elevation dropping in importance while aspect and percent forest had higher importance (Fig. 2.4c,d). However, the trend in which the importance of elevation went up at the expense of that of aspect with increasing grain size was consistent in both landscapes.

The predicted area of presence based on the TSS was greater than the true presence area across all models of each species in each landscape (Fig. 2.5). The closest models to the truth were those at or close to the correct grain size, with increasingly severe over-predictions in models of increasing grain size. For example, the true presence area of the fine-scale generalist in the heterogeneous landscape was 3691 km^2, and the 25-m MaxEnt model this species predicted 4591 km^2 (1.24 times more area) while the 1600-m Maxent model predicted 14907 km^2 (4.03 times more area, Fig. 2.5a, Fig. 2.6). The over prediction was even more pronounced in the specialist species, for which the true area of presence in the fine-scale species was 527 km^2. The closest model to the truth was the MaxEnt model at 50 m, which predicted 1116 km^2

(2.94 times more area), while the 1600-m MaxEnt model predicted 7260 km^2 (13.77 times more area). These over predictions were even larger in most of the GLMs, though they improved relative to the MaxEnt models at very large grain sizes (Fig. 2.5).

2.4 Discussion

Our simulation of virtual species and subsequent testing of models of those species revealed complex effects of niche breadth and grain size on SDMs. Without considering grain size, our results show species are more accurately modeled across heterogeneous landscapes compared to homogeneous landscapes. This is likely due to the larger differences in the predictor variables that drive suitability in heterogeneous landscapes, meaning that it is easier for the models to delineate areas of high vs. low suitability (and thus presence and absence on a landscape). Along similar lines, specialist species are more accurately modeled than more generalist species. This may be due to the fact that modeling algorithms should more easily differentiate between areas of higher suitability and areas of lower suitability when suitability is more restricted. Higher accuracy in models of the specialist species was seen even though there were a seventh or less of the presence points in the specialist compared to the generalist species. Greater accuracy in modeling specialist compared to generalist species has also been found in empirical studies (Hernandez et al. 2006, McPherson and Jetz 2007, Tsoar et al. 2007, Evangelista et al. 2008), and results of our simulation corroborate this.

Although our results have implications for the modeling of species with different niche breadths, they do not have direct applications to niche theory and modeling habitat specialization. This is because true habitat specialists will have evolved traits that make them more effective in certain environments compared to generalists, which is not captured in our virtual species. Examples of species that outperform others in a wide variety of environments are

on the rise, however, particularly with the species invasion phenomenon. Invasive species often have much broader success across landscapes and habitat conditions than native species (Richards et al. 2006). Predicting the spread of invasive species is an increasingly important endeavor for ecological research and conservation, and our results indicate that this will be challenging for successful invaders with broad niches. Our study kept suitability responses for each predictor to normal distributions. We acknowledge that this is a simplification, and therefore additional research to investigate different suitability response curves is necessary. This would more accurately simulate niche specialization and show its influence on the grain-size effects of the environmental variables.

Considering grain size, our results suggest that using predictors at the scale at which a species responds (in our case the scale at which it was simulated) will generally maximize model accuracy. However, the extent of these differences in accuracy varied by landscape, model, and virtual species. The habitat specialist species built at a coarser grain (200 m) in the heterogeneous landscape, for example, was in fact better modeled at smaller grain resolutions (50 and 100 m) than the one at which it was simulated. This suggests that in species with a narrower niche, selecting a fine enough grain to capture more details in the variation of the environmental variables is important. On the other hand, the deterioration in model accuracy at larger grain sizes was greater in the habitat generalist than in the habitat specialist. The deterioration of model accuracy with increasing grain size was also higher in the homogeneous landscape than the heterogeneous landscape. This makes sense considering there is a smaller suitability "signal" embedded in the background landscape for both habitat generalists and species in homogeneous landscapes, and our results indicate that this signal becomes harder to capture when increasing the grain size of the environmental predictors.

It is interesting to consider how variable importance changed with increasing grain size. The overwhelming importance of elevation in the heterogeneous landscape at the expense of aspect at larger grain sizes is likely because the average elevation value across a larger area is more meaningful than the average aspect value: consider the largest grain size of 1600 m. The average elevation of the area around a mountain top will produce a decent approximation with meaning to the model, while the average aspect will be a poor representation of the range of suitable aspect values throughout that cell. In the homogeneous landscape, elevation was far less important due to the small variance in its values across the landscape but followed a similar trend with increasing grain size. This has implications for any study system in which important variables lose meaning at coarser scales.

An additional advantage of modeling a species at the correct grain size was that it generally resulted in the most accurate measure of total area of presence based on the TSS. Our results suggest that if producing binary distribution or habitat maps for presentation or downstream applications, modeling at an inappropriate grain size will vastly overestimate a species' presence area. Models just 1 aggregation apart resulted in up to 3 times more predicted presence area while models at the largest grain size of 1600 m predicted up to 14 times more area than the true presence area. Similar relationships in landscape modeling have been described regarding increasing erosion area (Schoorl et al. 2000) and increasing predicted areas of landslide vulnerability (Claessens et al. 2005) with increasing grain size. Other ecological studies have likewise found an effect of larger predicted areas of presence from larger grain sizes (Seo et al. 2009), but ours is the first to measure the effect against the "truth" known from a derived virtual species. This has large implications for conservation planning when vital decisions about areas to protect and manage must be made with as accurate information as possible. The fact that

all models overestimated each species' presence area is also an important point to keep in mind when making conservation decisions on a landscape - species are likely less widely distributed then even models at empirically accurate scales are predicting.

Unfortunately, many of the climate data used as environmental predictors in SDMs are only available at grain sizes of 1 km or greater, which are often interpolations from even coarser data (Hijmans et al. 2005). Although generally avoided in ecological studies, there is potential in using geostatistical interpolation methods to further disaggregate predictor variables at larger grain sizes to match ecologically important variables for the species in question at smaller ones. Methods of obtaining finer resolution land cover information have been advanced in remote sensing sciences using super-resolution mapping (Atkinson et al. 2008), and textural analysis of remotely sensed imagery has recently been used to characterize sub-pixel habitat heterogeneity at global scales (Tuanmu and Jetz 2016). Further research should investigate the best methods of disaggregation for SDM application and the effects of disaggregating predictor variables on SDMs.

A lion most likely perceives habitat at a much larger scale than a shrew, and thus it makes more sense to model its habitat at larger grain sizes than those used to model a shrew's habitat. In addition to this point, however, is the fact that any given species is likely respond to their environment at a variety of scales. It is important to consider this scale-mismatch within a species distribution model. For example, the scale at which a species responds to water availability may be finer than that at which it responds to air temperature. The variation of response-scales across species is likely one reason that there has been a wide variety of findings with regards to the effects of grain size across studies. There is precedent in modeling species distributions with multiple scales of predictors in a single model (Bradter et al. 2013, Bellamy et

al. 2013), but research has also shown that including multiple empirically-selected grain sizes for predictors does not necessarily improve model performance over single-grain sized models (Martin and Fahrig 2012). The task becomes more complicated when considering model evaluation and selection criteria, as even commonly used statistics like AUC have been criticized as inappropriate in some cases (Lobo et al. 2008). In evaluating models of real species in which the true environmental suitability of an area is unknown, additional metrics such as explained deviance, AIC, BIC (for Bayesian models), and point biserial correlation should be employed, depending on the modeling method and research goals.

In the face of these complications, our results suggest that model accuracy and presence area predictions can be improved by evaluating models at multiple grain sizes and selecting the most accurate one. We increased grain size using mean values from the smaller cells, as our simulated species responded simply to values of the variables within a cell. Different summary statistics (such as variance, median, minimum value, etc.) may better explain a species' response to a given environmental variable when increasing grain size, depending on the study system. We recommend that every study modeling a species' distribution or habitat suitability across a landscape explicitly perform an analysis of grain size to decide what scale to model their species, as even well-founded expert opinion can be inaccurate (Peterman et al. 2014). Considering our findings that grain sizes larger than those at which species respond distort model accuracy and predictions to a greater degree than smaller ones, if a full analysis of grain-size effects is not feasible for a given project it is likely better to use smaller grain sizes.

APPENDIX

Table 2.1. The results of a random sampling of N = 2000 points across the 8 landscape/species combinations modeled.

Landscape / Species	Presence points	Absence points	Prevalence
Heterogeneous / Fine-scale generalist	172	1828	0.09
Heterogeneous / Coarse-scale generalist	220	1780	0.11
Heterogeneous / Fine-scale specialist	31	1964	0.02
Heterogeneous / Coarse-scale specialist	17	1983	0.01
Homogeneous / Fine-scale generalist	503	1497	0.25
Homogeneous / Coarse-scale generalist	646	1354	0.32
Homogeneous / Fine-scale specialist	176	1824	0.09
Homogeneous / Coarse-scale specialist	234	1766	0.12

Table 2.2. Area under receiver operating curve (AUC) and Pearson's correlation coefficients between predicted and true suitabilities of models at increasing grain size of the simulated species. Different letters in the AUC columns indicate values that are significantly different from others in that species/model group (b is significantly different from a and c, but not another b).

a. Heterogeneous landscape, habitat generalist

Grain size (m)	MaxEnt				GLM			
	Fine-Scale Species		Coarse-Scale Species		Fine Scale-Species		Coarse-Scale Species	
	AUC	Pearson's r	AUC	Pearson's r	AUC	Pearson's r	AUC	Pearson's r
25	0.95_{a}	0.90	0.90_{a}	0.77	0.93_{a}	0.67	0.90_{a}	0.71
50	$0.94_{a,b}$	0.83	0.90_{a}	0.79	$0.92_{a,}$	0.63	0.90_{a}	0.73
100	0.93_{b}	0.78	0.91_{a}	0.84	$0.91_{a,b}$	0.61	0.91_{a}	0.77
200	$0.91_{b,c}$	0.72	0.93_{b}	0.90	$0.90_{b,c}$	0.59	0.93_{b}	0.81
400	0.89_{c}	0.66	0.90_{a}	0.82	$0.89_{b,c}$	0.57	$0.91_{b,c}$	0.78
800	0.86_{d}	0.62	0.87_{c}	0.76	0.87_{c}	0.56	$0.90_{a,c}$	0.74
1600	0.82_{e}	0.58	0.84_{d}	0.70	0.85_{c}	0.56	0.89_{a}	0.70

b. Heterogeneous landscape, habitat specialist

Grain size (m)	MaxEnt				GLM			
	Fine-Scale Species		Coarse-Scale Species		Fine Scale-Species		Coarse-Scale Species	
	AUC	Pearson's r	AUC	Pearson's r	AUC	Pearson's r	AUC	Pearson's r
25	0.99_{a}	0.97	$0.96_{a,c}$	0.66	0.95_{a}	0.55	0.91_{a}	0.52
50	$0.98_{a,b}$	0.87	0.99_{b}	0.71	$0.95_{a,b}$	0.53	0.91_{a}	0.53
100	$0.96_{b,c}$	0.78	0.99_{a}	0.77	$0.94_{b,c}$	0.51	0.94_{a}	0.58
200	0.96_{c}	0.68	0.96_{a}	0.93	$0.92_{b,c}$	0.47	0.96_{a}	0.62
400	0.95_{c}	0.58	$0.96_{a,c}$	0.75	$0.92_{c,d}$	0.45	0.94_{a}	0.59
800	0.91_{d}	0.52	$0.95_{c,d}$	0.64	$0.91_{c,d}$	0.43	0.93_{a}	0.55
1600	0.90_{e}	0.45	0.91_{d}	0.54	0.91_{d}	0.42	0.93_{a}	0.52

Table 2.2 (cont'd).

c. Homogeneous landscape, habitat generalist

Grain size (m)	MaxEnt				GLM			
	Fine-Scale Species		Coarse-Scale Species		Fine Scale-Species		Coarse-Scale Species	
	AUC	Pearson's r	AUC	Pearson's r	AUC	Pearson's r	AUC	Pearson's r
25	0.86_{a}	0.94	$0.74_{a,b}$	0.75	0.81_{a}	0.65	0.79_{a}	0.65
50	0.83_{b}	0.85	0.76_{a}	0.77	0.81_{a}	0.63	0.79_{a}	0.67
100	0.80_{c}	0.76	0.75_{a}	0.83	$0.79_{a,b}$	0.59	0.79_{a}	0.71
200	0.76_{d}	0.64	0.81_{c}	0.94	$0.77_{b,c}$	0.54	0.84_{b}	0.76
400	0.69_{e}	0.53	0.73_{b}	0.78	$0.73_{c,d}$	0.49	0.79_{a}	0.70
800	0.64_{f}	0.45	0.68_{d}	0.65	$0.70_{d,r}$	0.44	0.75_{a}	0.62
1600	0.61_{g}	0.38	0.64_{e}	0.53	0.66_{e}	0.37	0.71_{c}	0.53

d. Homogeneous landscape, habitat specialist

Grain size (m)	MaxEnt				GLM			
	Fine-Scale Species		Coarse-Scale Species		Fine Scale-Species		Coarse-Scale Species	
	AUC	Pearson's r	AUC	Pearson's r	AUC	Pearson's r	AUC	Pearson's r
25	0.94a	0.92	0.84a	0.63	0.81a	0.46	0.84a	0.47
50	0.92b	0.84	0.85a	0.65	0.79a	0.45	0.85a	0.50
100	0.89b,c	0.73	0.86a	0.71	0.76b	0.44	0.86a	0.55
200	0.84c	0.59	0.93b	0.92	0.72b,c	0.39	0.90b	0.63
400	0.76d	0.45	0.84a	0.67	0.70c,d	0.36	0.85a	0.55
800	0.68e	0.33	0.77c	0.52	0.67d,e	0.31	0.80c	0.47
1600	0.65e	0.27	0.73d	0.42	0.65e	0.27	0.75c	0.40

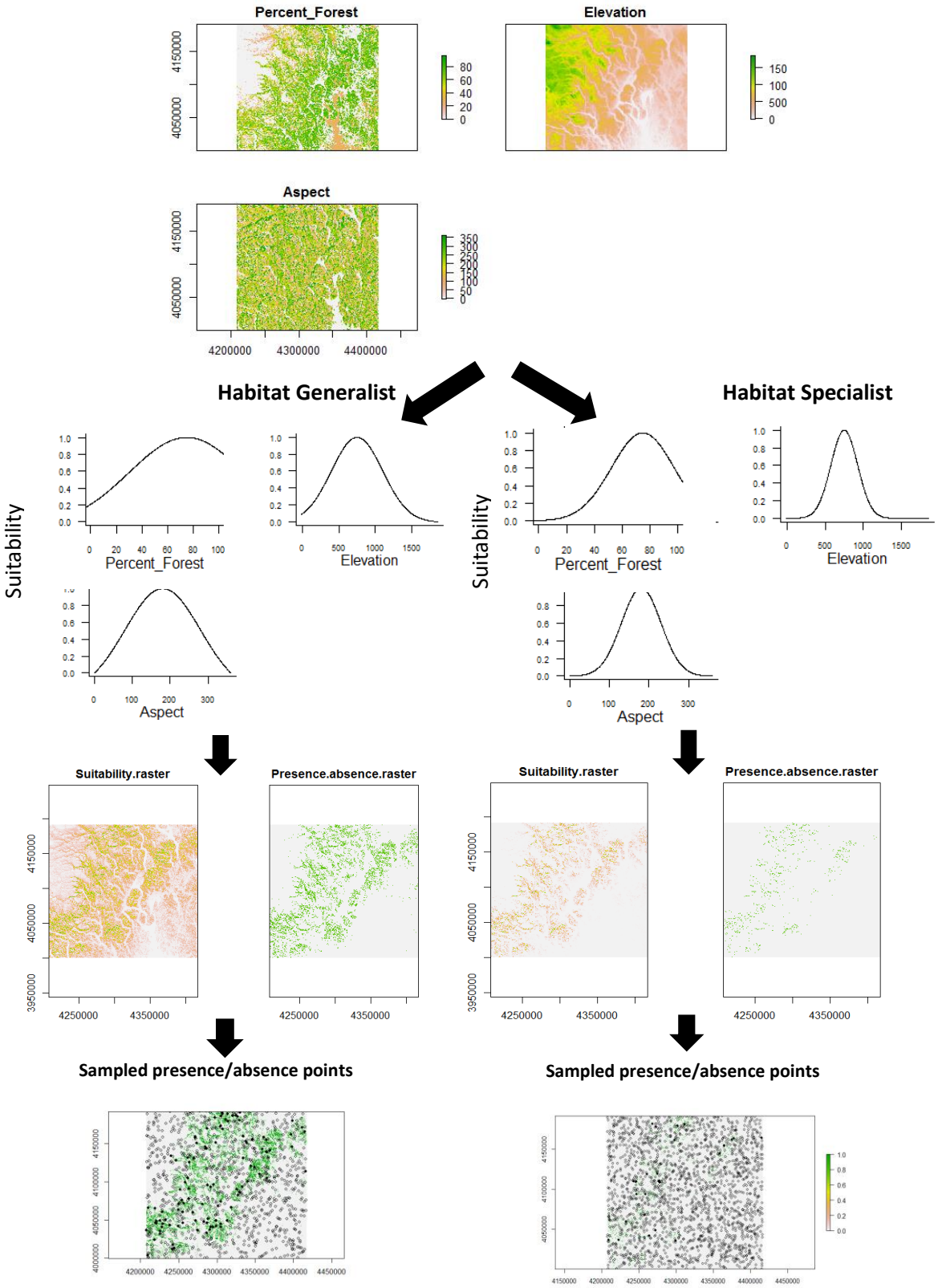

Figure 2.1. Flowchart of steps to simulate virtual species presence on a landscape.

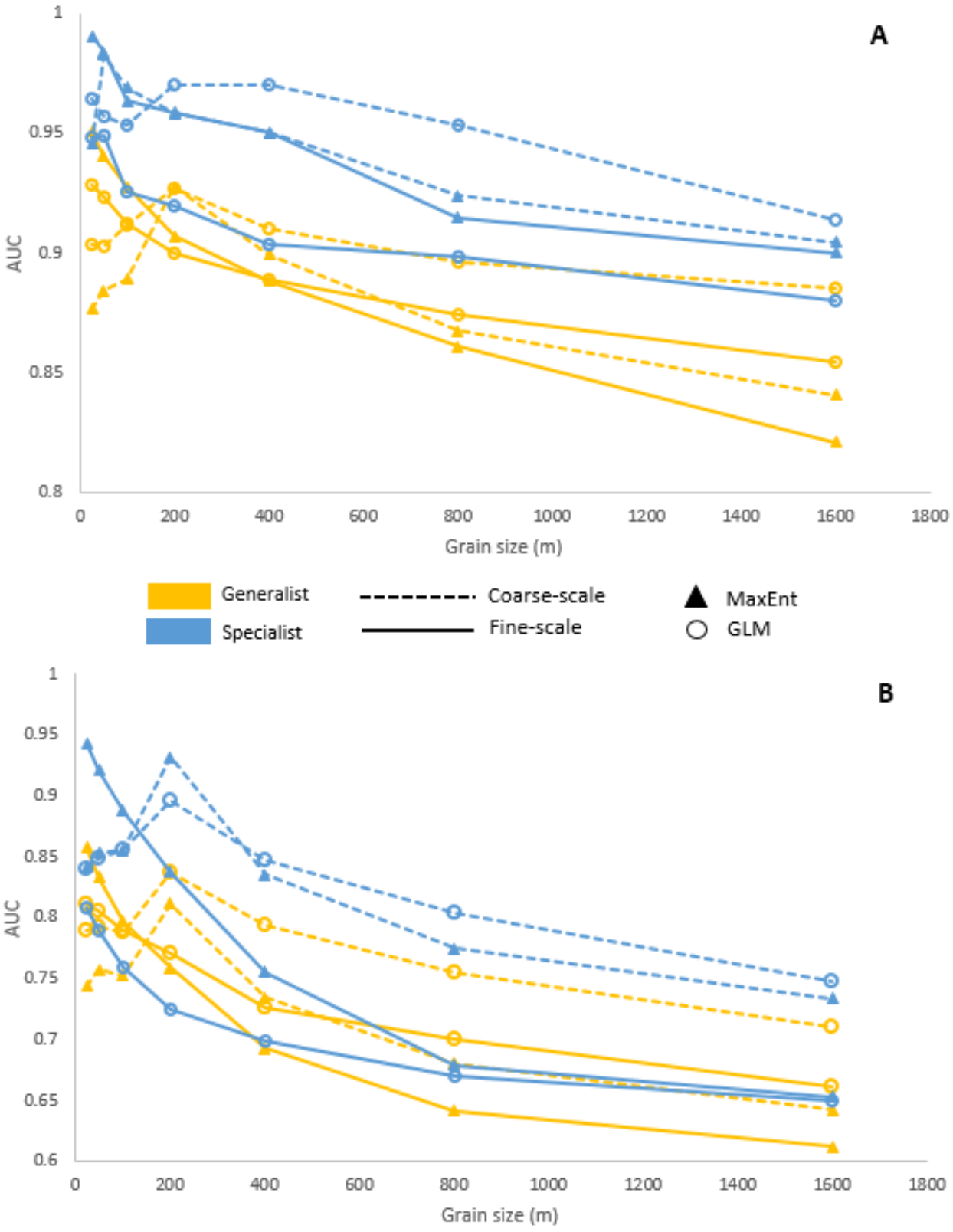

Figure 2.2. AUC scores of MaxEnt models and GLMs of virtual species across 7 grain sizes in (A) heterogenous landscape and (B) homogeneous landscape.

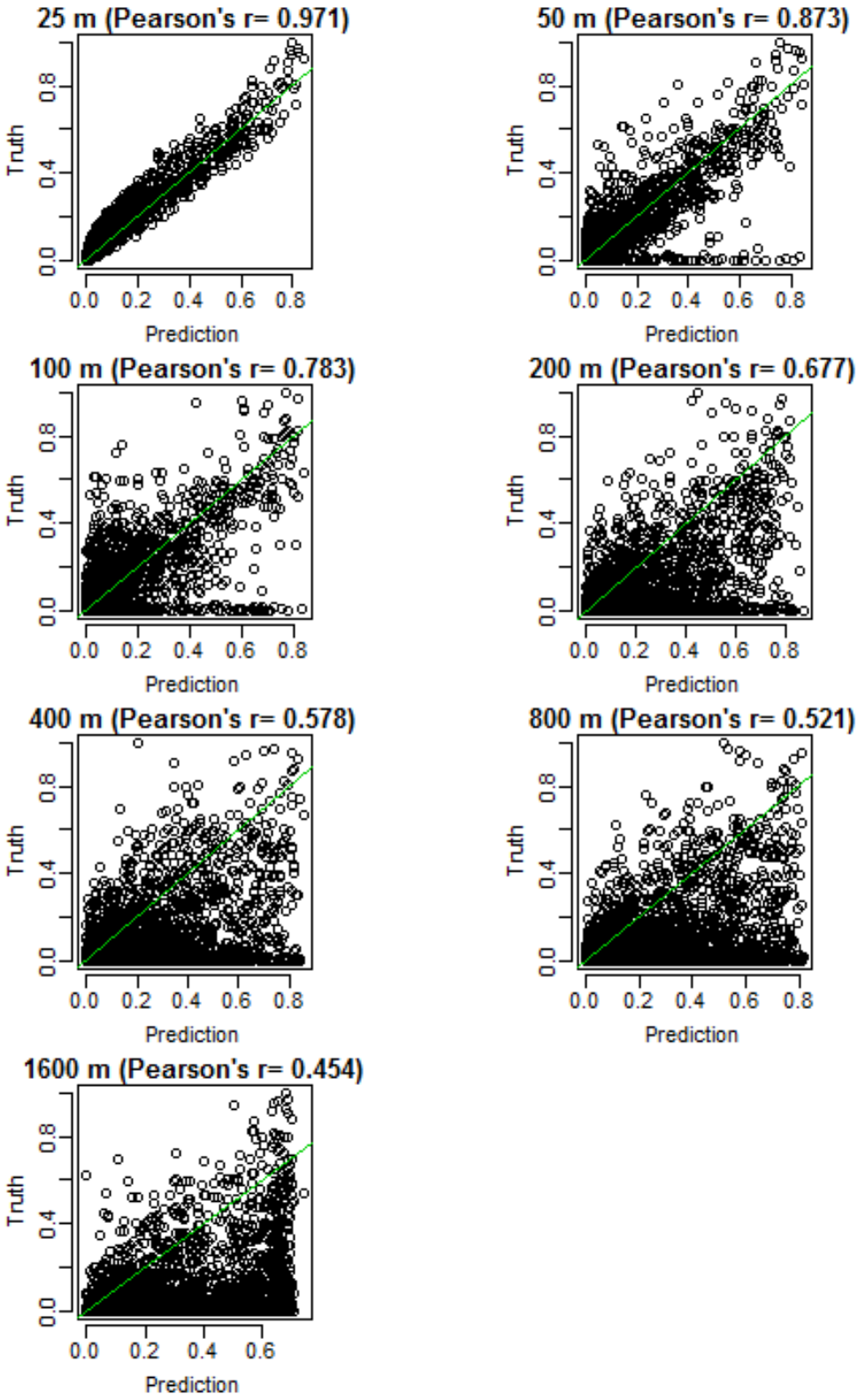

Figure 2.3. Scatterplots and Pearson's correlation coefficients of suitability predictions from MaxEnt models of and the true suitability values of the fine-scale habitat generalist species simulated in the heterogeneous landscape.

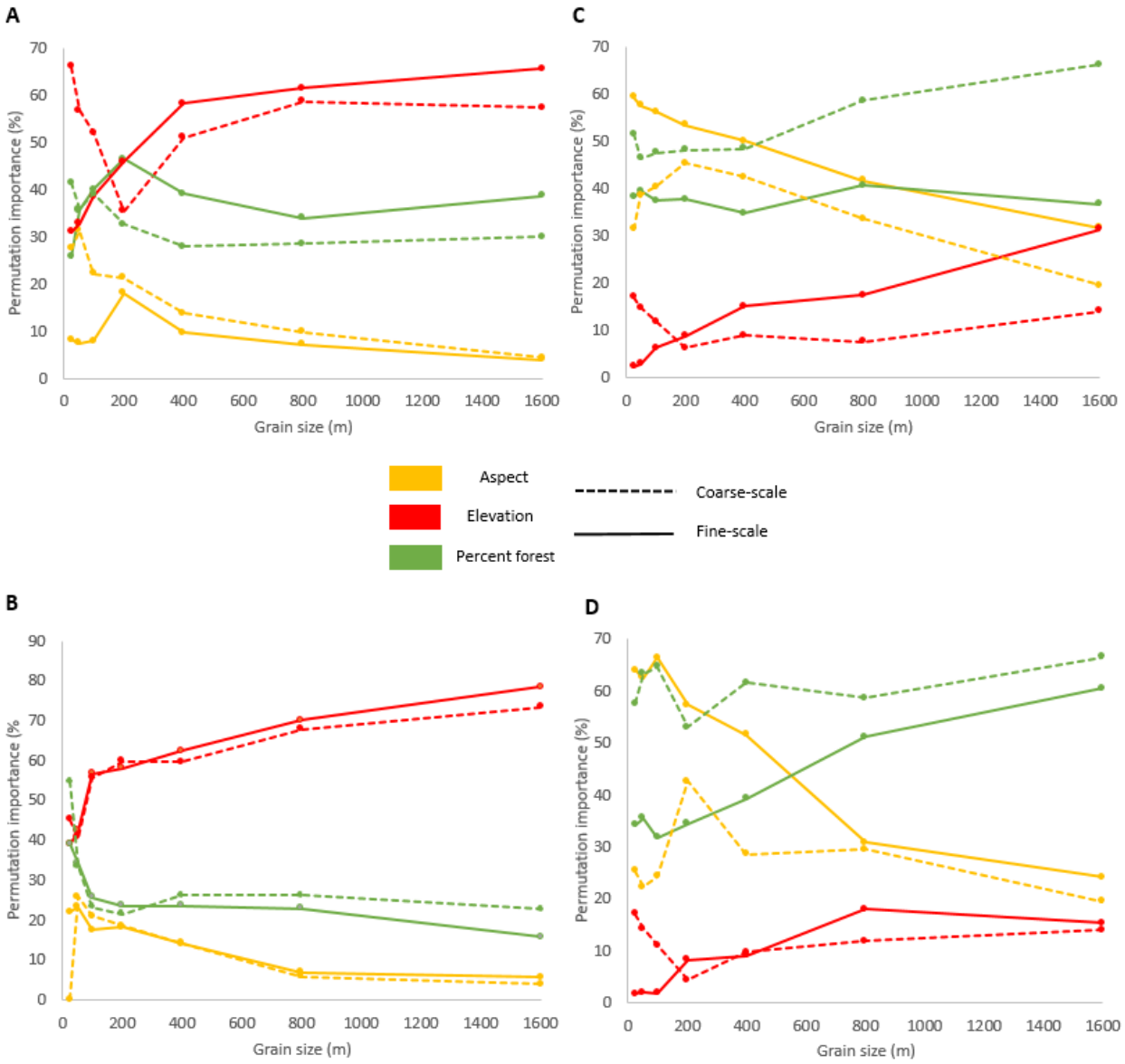

Figure 2.4. Permutation importance of variables to the MaxEnt models of (A) habitat generalist in heterogeneous landscape, (B) habitat specialist in heterogeneous landscape, (C) habitat generalist in homogeneous landscape, and (D) habitat specialist in homogeneous landscape.

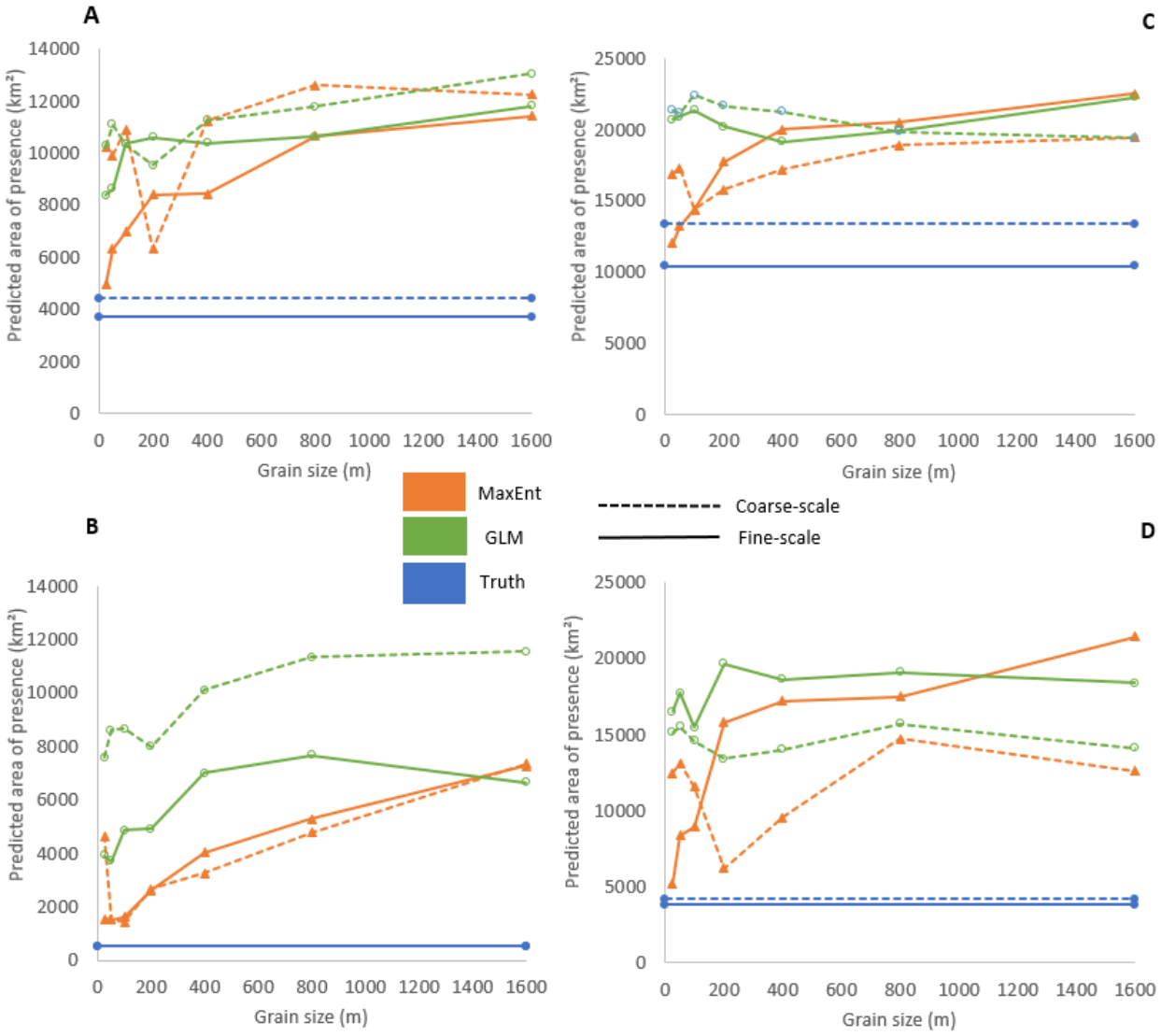

Figure 2.5. Total predicted area of presence above maximum TSS threshold in (A) habitat generalist in heterogeneous landscape, (B) habitat specialist in heterogeneous landscape, (C) habitat generalist in homogeneous landscape, and (D) habitat specialist in homogeneous landscape.

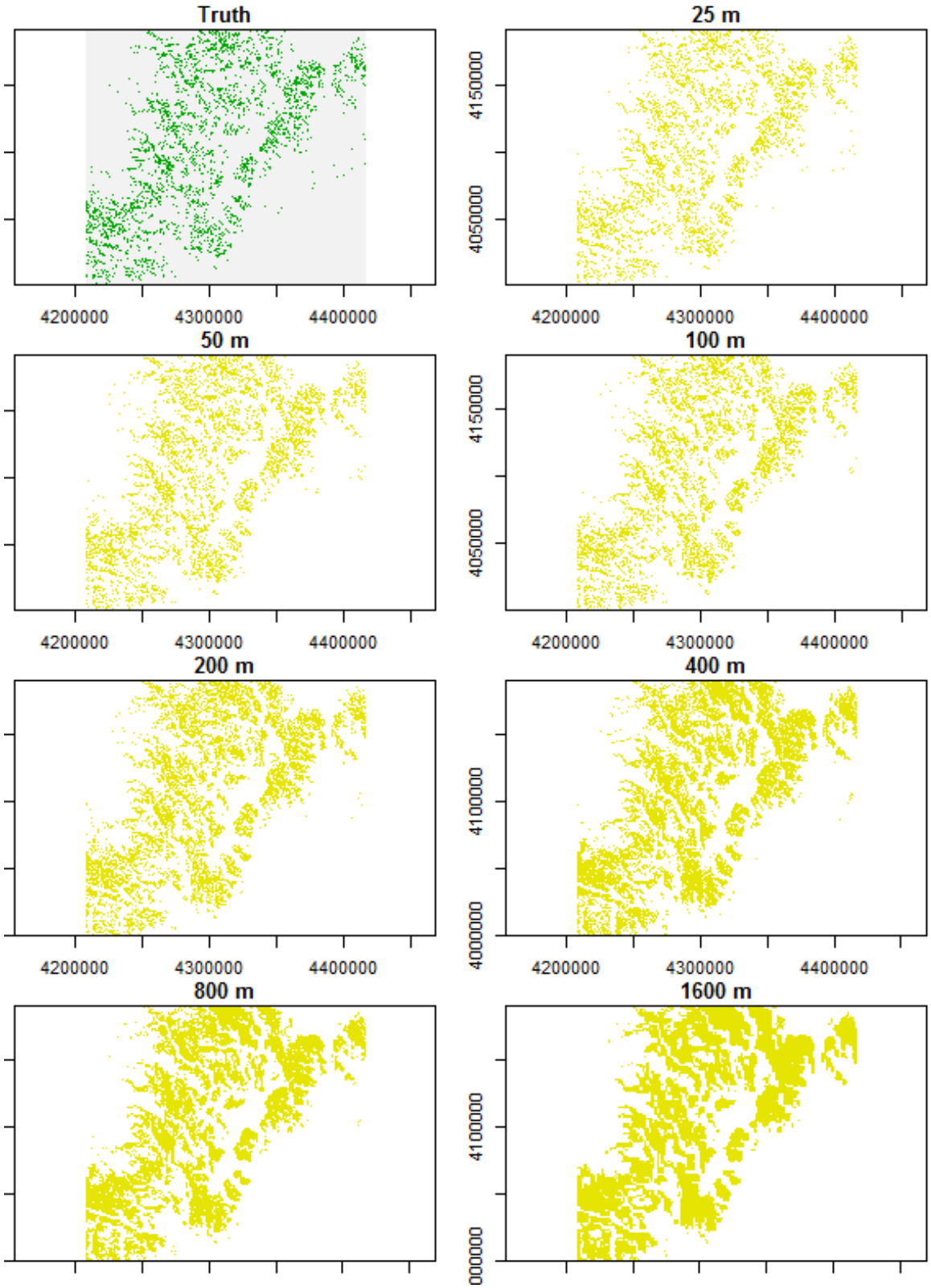

Figure 2.6. Presence maps of the fine-scale habitat generalist species simulated on the heterogeneous landscape produced by MaxEnt models of increasing grain size.

REFERENCES

REFERENCES

Allouche, O., A. Tsoar, and R. Kadmon. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology **43**:1223-1232.

Atkinson, P. M., E. Pardo-Iguzquiza, and M. Chica-Olmo. 2008. Downscaling cokriging for super-re solution mapping of continua in remotely sensed images. Ieee Transactions on Geoscience and Remote Sensing **46**:573-580.

Barbet-Massin, M., F. Jiguet, C. H. Albert, and W. Thuiller. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? Methods in Ecology and Evolution **3**:327-338.

Bean, W. T., L. R. Prugh, R. Stafford, H. S. Butterfield, M. Westphal, and J. S. Brashares. 2014. Species distribution models of an endangered rodent offer conflicting measures of habitat quality at multiple scales. Journal of Applied Ecology **51**:1116-1125.

Bellamy, C., C. Scott, and J. Altringham. 2013. Multiscale, presence-only habitat suitability models: fine-resolution maps for eight bat species. Journal of Applied Ecology **50**:892-901.

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.

Brotons, L. 2014. Species Distribution Models and Impact Factor Growth in Environmental Journals: Methodological Fashion or the Attraction of Global Change Science. Plos One **9**:5.

Claessens, L., G. B. M. Heuvelink, J. M. Schoorl, and A. Veldkamp. 2005. DEM resolution effects on shallow landslide hazard and soil redistribution modelling. Earth Surface Processes and Landforms **30**:461-477.

Conover, W. J. 1999. Practical Nonparametric Statistics. 3rd edition. Wiley.

Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carre, J. R. G. Marquez, B. Gruber, B. Lafourcade, P. J. Leitao, T. Munkemuller, C. McClean, P. E. Osborne, B. Reineking, B. Schroder, A. K. Skidmore, D. Zurell, and S. Lautenbach. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography **36**:27-46.

Elith, J., and C. H. Graham. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography **32**:66-77.

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.

Elith, J., S. J. Phillips, T. Hastie, M. Dudik, Y. E. Chee, and C. J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions **17**:43-57.

Evangelista, P. H., S. Kumar, T. J. Stohlgren, C. S. Jarnevich, A. W. Crall, J. B. Norman, and D. T. Barnett. 2008. Modelling invasion for a habitat generalist and a specialist plant species. Diversity and Distributions **14**:808-817.

Franklin, J., F. W. Davis, M. Ikegami, A. D. Syphard, L. E. Flint, A. L. Flint, and L. Hannah. 2013. Modeling plant species distributions under future climates: how fine scale do climate projections need to be? Global Change Biology **19**:473-483.

Gottschalk, T. K., B. Aue, S. Hotes, and K. Ekschmitt. 2011. Influence of grain size on species-habitat models. Ecological Modelling **222**:3403-3412.

Graham, C. H., S. Ferrier, F. Huettman, C. Moritz, and A. T. Peterson. 2004. New developments in museum-based informatics and applications in biodiversity analysis. Trends in Ecology & Evolution **19**:497-503.

Guisan, A., C. H. Graham, J. Elith, F. Huettmann, and N. S. Distri. 2007a. Sensitivity of predictive species distribution models to change in grain size. Diversity and Distributions **13**:332-340.

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

Guisan, A., N. E. Zimmermann, J. Elith, C. H. Graham, S. Phillips, and A. T. Peterson. 2007b. What matters for predicting the occurrences of trees: Techniques, data, or species' characteristics? Ecological Monographs **77**:615-630.

Hernandez, P. A., C. H. Graham, L. L. Master, and D. L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography **29**:773-785.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology **25**:1965-1978.

Hijmans, R. J., S. Phillips, J. Leathwick, and Ja Elith dismo: Species Distribution Modeling. R package version 1.1-4. https://CRAN.R-project.org/package=dismo (2017).

Hudak, A. T., J. S. Evans, and A. M. S. Smith. 2009. LiDAR Utility for Natural Resource Managers. Remote Sensing **1**:934-951.

Jimenez-Valverde, A., J. M. Lobo, and J. Hortal. 2009. The effect of prevalence and its interaction with sample size on the reliability of species distribution models. Community Ecology **10**:196-205.

Leipold, M., S. Tausch, P. Poschlod, and C. Reisch. 2017. Species distribution modeling and molecular markers suggest longitudinal range shifts and cryptic northern refugia of the typical calcareous grassland species Hippocrepis comosa (horseshoe vetch). Ecology and Evolution **7**:1919-1935.

Leroy, B., and C. N. B. Meynard, CélineCourchamp, Franck. 2015. virtualspecies, an R package to generate virtual species distributions. Ecography **39**:8.

Levin, S. A. 1992. THE PROBLEM OF PATTERN AND SCALE IN ECOLOGY. Ecology **73**:1943-1967.

Liu, C. R., M. White, and G. Newell. 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. Journal of Biogeography **40**:778-789.

Lobo, J. M., A. Jimenez-Valverde, and R. Real. 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography **17**:145-151.

Martin, A. E., and L. Fahrig. 2012. Measuring and selecting scales of effect for landscape predictors in species-habitat models. Ecological Applications **22**:2277-2292.

McPherson, J. M., and W. Jetz. 2007. Effects of species' ecology on the accuracy of distribution models. Ecography **30**:135-151.

Merow, C., M. J. Smith, and J. A. Silander. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography **36**:1058-1069.

Peterman, W. E., G. M. Connette, R. D. Semlitsch, and L. S. Eggert. 2014. Ecological resistance surfaces predict fine-scale genetic differentiation in a terrestrial woodland salamander. Molecular Ecology **23**:2402-2413.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling **190**:231-259.

Porfirio, L. L., R. M. B. Harris, E. C. Lefroy, S. Hugh, S. F. Gould, G. Lee, N. L. Bindoff, and B. Mackey. 2014. Improving the Use of Species Distribution Models in Conservation Planning and Management under Climate Change. Plos One **9**.

Renner, I. W., and D. I. Warton. 2013. Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology. Biometrics **69**:274-281.

Revermann, R., H. Schmid, N. Zbinden, R. Spaar, and B. Schroeder. 2012. Habitat at the mountain tops: how long can Rock Ptarmigan (Lagopus muta helvetica) survive rapid climate change in the Swiss Alps? A multi-scale approach. Journal of Ornithology **153**:891-905.

Richards, C. L., O. Bossdorf, N. Z. Muth, J. Gurevitch, and M. Pigliucci. 2006. Jack of all trades, master of some? On the role of phenotypic plasticity in plant invasions. Ecology Letters **9**:981-993.

Scales, K. L., E. L. Hazen, M. G. Jacox, C. A. Edwards, A. M. Boustany, M. J. Oliver, and S. J. Bograd. 2017. Scale of inference: on the sensitivity of habitat models for wide-ranging marine predators to the resolution of environmental data. Ecography **40**:210-220.

Schoorl, J. M., M. P. W. Sonneveld, and A. Veldkamp. 2000. Three-dimensional landscape process modelling: The effect of DEM resolution. Earth Surface Processes and Landforms **25**:1025-1034.

Seo, C., J. H. Thorne, L. Hannah, and W. Thuiller. 2009. Scale effects in species distribution models: implications for conservation planning under climate change. Biology Letters **5**:39-43.

Seoane, J., J. Bustamante, and R. Diaz-Delgado. 2004. Are existing vegetation maps adequate to predict bird distributions? Ecological Modelling **175**:137-149.

Trivedi, M. R., P. M. Berry, M. D. Morecroft, and T. P. Dawson. 2008. Spatial scale affects bioclimate model projections of climate change impacts on mountain plants. Global Change Biology **14**:1089-1103.

Tsoar, A., O. Allouche, O. Steinitz, D. Rotem, and R. Kadmon. 2007. A comparative evaluation of presence-only methods for modelling species distribution. Diversity and Distributions **13**:397-405.

Tuanmu, M. N., and W. Jetz. 2015. A global, remote sensing-based characterization of terrestrial habitat heterogeneity for biodiversity and ecosystem modelling. Global Ecology and Biogeography **24**:1329-1339.

Wiens, J. A. 1989. SPATIAL SCALING IN ECOLOGY. Functional Ecology **3**:385-397.

Wisz, M. S., R. J. Hijmans, J. Li, A. T. Peterson, C. H. Graham, A. Guisan, and N. P. S. Distribut. 2008. Effects of sample size on the performance of species distribution models. Diversity and Distributions **14**:763-773.

CHAPTER 3: INTERACTIVE SPATIAL SCALE EFFECTS ON SPECIES DISTRIBUTION MODELING: THE CASE OF THE GIANT PANDA

In collaboration with

Andrés Viña, Julie Winkler, Vanessa Hull, Ying Tang, Ashton Shortridge, Hongbo Yang, Zhiqiang Zhao, Fang Wang, Jindong Zhang, Zejun Zhang, Caiquan Zhou, Wenke Bai, Jianguo Liu

Abstract

Research has shown that varying spatial scale through the selection of the total extent of investigation and the grain size of environmental predictor variables has effects on species distribution model (SDM) results and accuracy, but there has been minimal investigation into the interactive effects of extent and grain. To do this, we used a consistently sampled range-wide dataset of giant panda occurrence across southwest China and modeled their habitat and distribution at 4 extents and 7 grain sizes. We found that increasing grain size reduced model accuracy at the smallest extent, but that increasing extent negated this effect. Increasing extent also generally increased model accuracy, but the models built at the second-largest (mountain range) extent were more accurate than those built at the largest, geographic range-wide extent. When predicting habitat suitability in the smallest nested extents (50 km^2), we found that the models built at the next-largest extent (500 km^2) were more accurate than the smallest-extent models but that further increases in extent resulted in large decreases in accuracy. Overall, this study highlights the impacts of the selection of spatial scale when evaluating species' habitat and distributions, and we suggest more explicit investigations of scale effects in future modeling efforts.

3.1 Introduction

Spatial scale has long been recognized as important to consider in ecological research (Wiens 1989). It varies along two main dimensions - grain, the spatial size of individual samples, and extent, the total area under study (Levin 1992). Because changes in the spatial scale of a study may result in complex effects that vary by species and system (Turner et al. 1989, Zeller et al. 2017), it is an important factor to consider when investigating habitat and species distributions (Saab 1999). Empirical investigations of the effects of spatial scale are not only rare, but also mainly involve varying the "local extent" surrounding a species or individual location within which environmental data are averaged (Thompson and McGarigal 2002, Zeller et al. 2017). While this may constitute an effective way to study the effects of spatial scale on species distributions and habitat selection, the effect of varying the local extent itself is constrained by both the grain size of the included environmental variables and the total extent of the study area. It has also been commonplace to select the grain size of a study a *priori* and without any kind of optimization, usually based on expert opinion and/or the format of the available data needed for modeling (Austin and Van Niel 2011, Manzoor et al. 2018).

One of the main reasons behind the usage of varying spatial scales between studies is that the goals of habitat modeling for a species can vary widely, ranging from investigations on the within-home range habitat selection of individual animals (Hull et al. 2016) to the delineation of a species' entire distribution (Yang et al. 2017). These goals directly relate with Johnson's ((Johnson 1980) four orders of habitat selection, which in descending order include geographic range, home range, differential space use within the home range, and the final procurement of resources. Investigations into these different orders of habitat selection largely drive the spatial scale used in a given study. These are most apparent through changes to the total extent of

analysis (a home range compared to a species' geographic range), but they also influence the grain through the common use of coarser climate variables (Manzoor et al. 2018) and lack of computational power for fine grains in studies of fourth-order (geographic range) processes (Renner and Warton 2013) . Recent work to integrate multiple orders of habitat selection simultaneously through varying the local extent of included environmental predictor data and hierarchically modeling selection has shown that species may respond to environmental variables differently between orders and/or between scales within orders (DeCesare et al. 2012, Zeller et al. 2017).

Another key driver of the scale of analysis in habitat and species distribution modeling is data availability. The grain size influences which environmental predictor variables can be included in modeling efforts. In general, data availability, such as the availability of climate observations and simulations of future climate (Hijmans et al. 2005, Tang et al. 2018) is greater at larger grain sizes that become relevant at larger extents. The varying resolution of available datasets often results in the spatial aggregation of variables with smaller grain sizes to match the coarser scales of the other variables, or the interpolation of variables of larger grain sizes to match the finer resolution of the other variables (Austin and Van Niel 2011), with the latter approach increasing their spatial autocorrelation (Vina et al. 2016). In addition to impacts on environmental variables, spatial aggregation to coarse scales can also affect the presence dataset used in the habitat modeling of the species in question (Fig. 3.1). For instance, data collected at a fine scale with points falling within 30 meters of one another may fall in different 30 m cells but the same aggregated 60 m cell. Potential impacts are exacerbated when cells are only defined as presence or absence, ignoring the number of points within a cell, the default option for the popular MaxEnt software (Phillips et al. 2006) and any species distribution model for which the

species input is binary (presence or absence for a given location). The potential tradeoffs in model performance that occur with the aggregation of some variables to allow for the addition of others, and how this interacts with changes in the extent of analysis, is another scale-related topic that has been understudied.

While there have been empirical studies evaluating the effects of grain size (Thomas et al. 2002, Guisan et al. 2007a, Guisan et al. 2007b, Connor et al. 2018) and total study extent (Anderson and Raza 2010, Wheatley 2010) on habitat and species distribution modeling, little research has looked at the interactive effects of these components. Seo et al. (Seo et al. 2009) examined the effects of grain size on distribution models of species with different range sizes (i.e., total extents), although they did not vary the total extent of analysis within a given species. They found that model results of species with intermediate range sizes were more sensitive to changes in grain size compared to species with large or small ranges. Although Trivedi et al. (Trivedi et al. 2008) built models at multiple grains and total extents, the sampling designs of the datasets used in the analysis varied for the different scales making direct comparisons difficult. Khosravi et al. (Khosravi et al. 2016) varied the grain size of their environmental predictors in addition to the local extent around each presence and background cell within which they averaged those predictors in modeling goitered gazelle distributions and found intermediately-sized grains and local extents to be most accurate, but they did not vary the total extent of the area modeled. To the best of our knowledge, only one previous study varied both grain size and total extent, in which only two coarse grains (1 km and 10 km) and two broad extents (Africa and West Africa) were used. The smaller 1-km grain size at the regional (as opposed to continental) extent was found to best capture the distributions of three widespread African species (Patas monkey, bull frog, and rock hyrax) at the edge of their range (Vale et al. 2014). A

deeper analysis of the interactive effects of the total extent and grain size of environmental variables on habitat and species distribution modeling is crucial, since the range of scales used to model habitat or species distributions can vary considerably (e.g. total extents of 40 km^2 to 66,456 km^2 and grain sizes of 30 m to 1,000 m for giant pandas (Guan et al. 2016, Hull et al. 2016, Yang et al. 2017), with potentially important effects on model accuracy and sensitivity.

To address these knowledge gaps, here we evaluate the effects of simultaneously changing the grain size and total extent, as well as the use of different environmental predictor variables, on habitat suitability and species distribution modeling of a demonstration species, the giant panda (*Ailuropoda melanoleuca*). Due to their elusive nature and protected status, there are still many uncertainties about giant panda behavior in the wild (Schaller et al. 1985). Although there have been numerous studies on giant panda habitat selection and distribution, there are still many discrepancies concerning the relative importance of and responses to different environmental variables used to model their habitat (Hull et al. 2014). Like most species, there has also been little investigation into spatial scale effects, which may account for some discrepancies seen in the literature. Giant pandas also serve as an effective case study to test for interactive spatial scale effects on habitat and distribution modeling, particularly of habitat specialist species, due to their dependence on understory bamboo and sensitivity to human disturbances (Schaller et al. 1985, Pan et al. 2014). Any evidence of interactive scale effects on model performance, as well as scale interactions with environmental variable selection, will have important implications for both giant panda research and species distribution modeling in general and will imply that spatial scale needs to be explicitly considered for effective research and conservation.

3.2 Methods

3.2.1 Study area and extents

We chose four different total extent sizes to model panda distributions consisting of 50 km^2, 500 km^2, the mountain ranges containing these extents (mean area of 18,264 km^2) and the entire geographic panda range of approximately 109,585 km^2 (Table 3.1, Fig. 3.2). The nested structure of these extents allowed for replication of the smaller three within the largest. Specifically, the total geographic range was gridded into 50 km^2 blocks, and those with at least 20 panda presence locations were selected for modeling (n = 68). Several of these were then randomly selected to be buffered so that there was n = 9 non-overlapping total extents of 500 km^2. Finally, distinct mountain ranges that contained one or more of these 500 km^2 total extents were selected as replicates of the third total extent (n = 4).

3.2.2 Environmental variables and grain sizes

We chose environmental predictors based on their biological significance for pandas, as well as data availability. Our base model included elevation, slope, percent forest, distance to a main road, and distance to a stream or river (Table 3.2). Main road and stream/river data were obtained at a scale of 1:250000 from the Chinese National Fundamental Geographic Information Database, maintained by the National Geomatics Center of China (National Geomatics Center of China 2018). The smallest available grain size for elevation was a digital elevation model (DEM) of 30 m, obtained through NASA's SRTM mission and void-filled by the USGS (USGS 2000). Slope was derived from this DEM in ArcGIS 10.4. Percent forest was obtained from the global dataset developed by Hansen et al. (Hansen et al. 2013) at a grain size of 27 m. These data were resampled to 30 m using the bilinear interpolation, while the distance variables were calculated using a grain size of 30 m. We then successively spatially aggregated these variables by a factor

of 2 using the mean of the smaller cells to develop 6 additional sets of predictors of increasing grain size. We also did the analyses using the median value of the smaller cells, which had nearly identical results. For each of the 4 extents, we thus included the same environmental variables at 7 different grain sizes ranging from 30 to 1920 m (Table 3.2).

At the grain size of 240 m, we also built models that included 11 MODIS-derived land surface phenology metrics that have been used to accurately predict bamboo (Tuanmu et al. 2010) and giant panda (Vina et al. 2010) distributions. These 11 metrics describe the timing and magnitude of phenophases of the land surface (Tuanmu et al. 2010). Many of these metrics are highly correlated ($r > 0.9$). To avoid multi-collinearity issues (Dormann et al. 2013), we conducted a principal component analysis (PCA) on the 11 metrics and used the first five components (explaining over 99% of the variation) as environmental variables in our models. We resampled these first five components from native MODIS resolution of 250 m to the grain size of the coarsened base variables at 240 m using bilinear interpolation. We then aggregated this new set of environmental variables in the same manner as before to derive 3 additional sets of variables at 480, 960, and 1920 m (Table 3.2). The MODIS-derived phenology variables were averaged from satellite data acquired in 2000-2002, the main sampling years of the 3rd National Giant Panda Survey (State Forestry Administration 2006).

Finally, for the habitat modeling at the 960 and 1920 m grain sizes, we trained models that additionally incorporated remotely sensed climate variables. The source of the climate information was the Deblauwe et al. (Deblauwe et al. 2016) dataset of bioclimatic variables derived from remotely-sensed measurements with a ~6 km resolution. Tang et al. (Tang et al. 2018) later interpolated these measurements to a 1 km resolution using an inverse weighted distance approach, which we also used. Five bioclimatic variables were selected based on their

ecological meaning for giant pandas and lack of collinearity (Tang et al. 2018). These included annual mean temperature, temperature seasonality, temperature annual range, annual precipitation, and precipitation seasonality. We resampled the climate data using bilinear interpolation to match the 960 m grain size, and once again aggregated by a factor of 2 using a mean function to produce the final set of environmental variables at 1920 m (Table 3.2). Like the phenology variables, the satellite-derived bioclimatic variables were averaged between the years of 2000 and 2002 to capture the conditions found during the 3rd National Giant Panda Survey.

3.2.3 Giant panda presence data

To model giant panda habitat at different scales, we used presence data from the 3rd National Giant Panda Survey (State Forestry Administration 2006). These data were collected between 2000 and 2003 across the current giant panda geographic range and consist of georeferenced locations that had signs of giant panda presence (e.g., feces, hair; State Forestry Administration 2006). The sampling technique involved placing line transects within 2 km^2 grid cells across the entire known panda range and searching for panda signs along these transects. For the 30 m grain size, the smallest extent replicates had a mean of 30.5 (SD=11.3) presence cells, replicates for the second extent level had a mean of 154.4 (SD=77.8) presence cells, replicates for the third level had a mean of 1137.6 (SD=845.1) presence cells, and the total range (extent 4) had 4707 presence cells (Table 3.1).

3.2.4 Habitat Suitability Modeling

Because only giant panda presence data were available for this analysis, we ran MaxEnt models for the various sets of predictor variables at different grain sizes and total extents (Phillips et al. 2006). MaxEnt makes use of machine learning techniques to derive a probability distribution of environmental suitability across a landscape from the environmental predictor

variables. The algorithm attempts to minimize the relative entropy in the predicted suitability between presence locations and background locations (Elith et al. 2011). We used 10,000 randomly sampled points from within the precise polygon for a given total extent to serve as background locations to the models. At the smaller total extents, limited cells in the larger grain sizes resulted in all the raster cells of the environmental predictors serving as background data. For example, the models built at the smallest total extent and with the 1920 m grain size only had a total of 12 cells. We selected background points only from areas in which a panda could feasibly occur (Phillips et al. 2009), which we defined as below 3600 m in elevation due to the lack of bamboo in alpine areas. We used a k-fold partitioning design to develop five different random combinations of training (80%) and testing (20%) data for model input at every grain size and extent.

To evaluate each model, we calculated the Area under the Receiving Operating Characteristic Curve (AUC), which measures the ability to discriminate between observed presence and background grid cells. Due to criticism of AUC as an accuracy metric that may result in inflated and spurious values (Lobo et al. 2008, Fourcade et al. 2018), we also calculated True Skill Statistics (TSS) and correlation coefficients between the predicted suitability values at test presences vs. background locations obtained from each model (hereafter referred to as 'correlation coefficient'). TSS is advantageous in our case due to its independence from prevalence (Allouche et al. 2006), which changes across extent and grain (Table 3.1). We summarized this statistic at the threshold for conversion of the predicted probability surface into a binary presence and absence surface that maximized its value (i.e., maximum TSS).

Replication of the smaller total extents allowed us to examine the effect of a given study area's placement on the environmental gradients (e.g., latitude, percent forest) found across the

panda range, and if this modified the effect of total extent and/or grain. We calculated the mean value of every environmental predictor within every total extent replicate to determine that extent's placement on the environmental gradient for each respective predictor. In a similar manner, we were able to determine if the amount of environmental heterogeneity (e.g., elevation, percent forest) in a given study area had an impact on the effect of total extent/grain size on model accuracy. We calculated the standard deviation of every environmental predictor within every study area to represent the environmental heterogeneity for a given predictor in each respective total extent replicate.

We also evaluated the performance of models built at larger total extents in predicting habitat suitability in the smallest total extent nested within those larger total extents. This was done for the nine smallest extent replicates that were buffered to produce the next total extent. This allowed us to determine the relationship between increasing total extent and predictive accuracy using the same test area and data in each case. We also produced predicted habitat suitability maps to visually compare the results of each model. This analysis was done using the base models at the 30-m grain size.

To further evaluate the tradeoffs associated with the addition of the MODIS and climate variables at the larger grain sizes, we compared model parsimony using Akaike's Information Criterion (Akaike 1974) corrected for small sample sizes (AICc). This measure is particularly important when evaluating models with a different number of variables, as it incorporates a penalty for additional explanatory variables and overfitting the model (Warren and Seifert 2011). AICc cannot effectively evaluate between models of different total extent or grain size, however, because changing either of these components of spatial scale changes the sample locations

included in the model, thus making the information criterion not comparable (Burnham and Anderson 2004).

To evaluate the effects of environmental predictors on the various models at different scales, we conducted permutation importance tests (Phillips et al. 2006). The permutation test slightly changes the values of each environmental variable among the training presence and background points, and measures the loss in AUC through this process. This was done for each variable separately, and the final values were normalized to percentages. We also produced environmental response plots for each variable to visualize their effects on habitat suitability and how these effects changed across scales, suites of variables, and study areas.

In summary, base models were built at seven different grain sizes, base + phenology models were built at four different grain sizes, and base + phenology + climate models were built at two different grain sizes. These models were built within all replicates of the four total extents. All reported statistics (e.g., AUC, TSS, AICc) are a result of averaging the five training/testing model runs. All these analyses utilized the 'dismo' package in the R environment (Hijmans et al. 2016, R Core Team 2018).

3.3 Results

The accuracy of the models generally increased with increasing total extent and the addition of environmental variables at coarser grains. Increasing grain size in the smallest total extent decreased accuracy, as measured by the max TSS, but this effect was lost in the three larger total extents (Fig. 3.3). The AUC and correlation coefficient results mirrored the max TSS except that increasing grain size reduced accuracy in total extent 2 in addition to the smallest extent (Fig S1). The placement of a given study area along some environmental gradients had a strong effect on accuracy, and this placement also moderated the effect of total extent on

accuracy. For example, the accuracy of models using the base set of variables generally decreased with increasing forest cover, but this decrease occurred at different rates across the different total extents so that their accuracies converged at about 75% forest cover (Fig. 3.4). A similar pattern emerged with increasing latitude in extents 2 and 3, but little to no effect was seen across the elevation and mean annual temperature gradients (Fig. S2). Clear effects of grain size on model accuracy across environmental gradients were not evident, as changes in accuracy showed nearly the same trend across all grain sizes (Fig. S3). Increasing environmental heterogeneity within a given study area resulted in increased model accuracy, but did not appear to moderate the effects of either total extent or grain size (except for the 1920-m grain) (Fig. S4, S5).

The models built at different total extents resulted in different suitability predictions of the smallest (total extent 1) study areas nested within them (Fig .5). The second-smallest total extent models had the highest accuracy when predicted to smallest total extent area (mean max TSS=0.58 (SD=0.17), mean AUC=0.78 (SD=0.13), followed by the smallest extent models (mean max TSS=0.57 (SD=0.20), mean AUC=0.76 (SD=0.13), the second-largest total extent models (mean max TSS=0.48 (SD=0.15), mean AUC=0.71 (SD=0.11), and finally the largest total extent models (mean max TSS=0.39 (SD=0.18), mean AUC=0.64 (SD=0.15)). The predicted suitability maps of these models were qualitatively very different, with the most accurate models producing predictions that had a more pronounced difference between high and low suitability areas while the largest total extent models produced predictions with a "smoothed" appearance (Fig. 3.5).

Increasing grain size in order to incorporate the additional phenology and climate variable sets at the 240 m and 960 m grains, respectively, resulted in improved model accuracy

in the larger two total extent sizes but had no significant effect at a 95% confidence interval in the smaller two total extent sizes (Fig. 3.3). The AICc rankings generally agreed with the accuracy metrics, with the addition of phenology and climate variables resulting in higher ranked models at the two largest total extents, but generally lower ranked models across the replicates of the two smallest total extents (Table 3.3).

The contribution of individual environmental variables to the models, as measured by the permutation importance tests, varied with total extent and grain. For example, although elevation was important to the models incorporating the base variables at the smallest total extent, with an average of about 40% permutation importance, this increased to about 60% permutation importance for the base models at three larger total extents (Fig. 3.6). Conversely, the distance to road variable lost importance as total extent increased with average values of 22, 17, 10, and 7% importance in base models trained at smallest to largest total extents, respectively. Similar losses occurred in slope and distance to river variables with increasing total extent. Though the permutation importance of most variables did not change much with increasing grain size, elevation tended to lose importance and percent forest tended to gain importance (Fig. 3.6). Elevation also lost importance with the incorporation of phenology and climate variable sets at the 240 and 960 m grain sizes, respectively (Fig. S6). Mean annual temperature proved particularly important to the models built at the second-largest (mean permutation importance of 17%) and largest (permutation importance of 33%) total extents.

Predicted suitability responses to the environmental variables changed considerably across the replicated study areas, and to a lesser extent due to changes in scale. For example, although there was generally a positive response to forest cover across the replicates of the smallest total extent, the magnitude of the effect varied and in some areas there was zero effect

(Fig. S7). Many of the models built in the replicates of the smallest total extent featured different environmental response curves than the models built in the larger total extents in which they were nested, while those models built at the larger total extents featured similar environmental responses to each other. The effect of the distance to road variable on environmental suitability is a good example of this, with increasing distance to roads resulting in clear positive effects in the models built at larger total extents but varied effects in the models at the smallest total extent (Fig. 3.7, S8). Increasing grain size tended to result in losses in detail in the environmental responses, with some variables having little to no effect on suitability at the 1920-m grain size (Fig. 3.7).

3.4 Discussion

Our results highlight the interactive and potentially large effects of spatial scale on species distribution modeling. Unlike previous studies, ours is the first to use a single, comprehensive presence dataset to investigate these effects along both main dimensions of spatial scale – total extent and grain size. Models built and validated at larger total extents generally had higher accuracies, likely because as total extent increases, so does the amount of presence data and the range of the environmental variable values used to model habitat. More presence data allows the models to better approximate the response of a species to the environment, though MaxEnt has been shown to produce accurate predictions at low sample sizes (Stockwell and Peterson 2002, Hernandez et al. 2006). Perhaps more important is the fact that, at larger total extents, a larger range of environmental predictor values allows the models to better differentiate between areas of low and high suitability at that given total extent (Anderson and Raza 2010).

A potential exception to these relationships is a situation in which a species' response to its environment changes significantly across its geographic range (Osborne et al. 2007). In this case, a single model with fixed parameters may not be appropriate to describe the responses of individuals to their environment in different populations across their geographic distribution. Our results confirm this phenomenon for the giant panda, as the models trained in the replicates of the second-largest total extent (representing individual mountain ranges) had higher accuracies than models trained at the entire geographic range. Because there is little to no dispersal among giant panda populations living in several of the different mountain ranges due to natural (e.g., large rivers) and anthropogenic (e.g., highways) barriers (Zhao et al. 2013), these populations have likely adapted to specific environmental conditions found in their respective mountain ranges. This suggests that the benefits associated with modeling at larger total extents (more presence locations, increased range of environmental variables) do not necessarily outweigh the negative effects associated with including populations with different environmental responses within the same model. Therefore, this effect should be tested in any geographic range-wide evaluation and depending on the results, predictions from models trained at a mosaic of total extents should be used as opposed to a single, geographic range-wide model. Alternatively, modeling techniques that allow for non-stationarity in environmental responses could be considered (Osborne et al. 2007).

Although grain size had a strong effect on model accuracy in models built at the smallest total extent with smaller grain sizes producing more accurate predictions, this effect was negated by increasing the total extent. Models built at the smallest total extent of 50 km^2 were likely negatively affected by increasing grain because of the severe loss of presence and background cells with each aggregation in grain. By the largest grain size, there was only an average of 8

presence cells and 12 total background cells per 50 km^2 study area. This resulted in large overlaps between presence and background environmental conditions and a weak signal between suitable and unsuitable habitat (Merow et al. 2013). Our results suggest giant panda distributions at total extents of 500 km^2 and higher can be accurately modeled using grain sizes up to 1920 m, however.

The method of aggregation could also impact model results – we repeated the analysis using the mean and median value of the smaller cells for the aggregated cell with nearly identical results, but within-cell variance is lost through these methods. Some variables may aggregate better using variance measures instead. For example, slope values at 30 meters can rapidly lose meaning when aggregated due to variable terrain. Replacing a mean slope aggregate with a standard deviation of elevation among the smaller cells (Tuanmu et al. 2016) would better represent terrain ruggedness. It is also important to note that even though accuracy of models built at large grain sizes may remain high in most cases, there are many situations in which they will not be useful. The decision to model at a certain total extent and grain size thus should consider research objectives (e.g., an estimation of a species range allowing a larger grain vs. a detailed understanding of environmental drivers of habitat suitability a finer grain), in addition to model accuracy.

Our analyses of predictive accuracy of models trained on the first three total extents replicated across the panda range suggest that the placement of a given study area on some range-wide environmental gradients drives both model accuracy and modifies the effect of total extent on model accuracy. Using percent forest cover as an example, the decrease in model accuracy with increasing forest cover makes sense due to the loss of the clear habitat suitability differences found between cells with more forest and those with less forest. This likely partially

explains the decrease in accuracy we found along an increasing latitudinal gradient (Fig. S2), with models trained in the larger study extents within the Qinling mountains (Northeastern most part of the panda range) featuring more forest and lower predictive accuracies. Our finding that increasing environmental heterogeneity resulted in more accurate models follows expectations – higher variation in these variables should allow the models to better differentiate between suitable and unsuitable habitat (Merow et al. 2013). The fact that we did not find a relationship between these increases in environmental heterogeneity and the effect of grain size on model accuracy was unexpected, however. This suggests that even grain sizes up to 1920 m capture enough of the environmental heterogeneity within study area replicates across the panda range to model habitat suitability with accuracy comparable to the finest grain sizes, regardless of the total level of environmental heterogeneity found within a given study area.

Recent developments in species distribution modeling beg more scale-related questions. The demonstration that MaxEnt is equivalent to an inhomogeneous Poisson process (IPP) has resulted in the potential to incorporate the number of presence points within a cell in the modeling framework (Renner and Warton 2013). The new open-source release of MaxEnt takes the IPP formulation and allows for the estimation of the relative abundance or probability of presence (Phillips et al. 2017). This has implications for the grain size used to represent the environment, as it replaces the binary presence/absence response with a count response which retains more information per cell as environmental data is aggregated to larger grain sizes. Specifically, as opposed to multiple occurrence points resulting in one presence cell, those points will be modeled as multiple observations of the species under the "same" environmental conditions. The relationship between the spatial autocorrelation of presence points (Renner et al. 2015) and the grain size of environmental variables in both the original and IPP formulation of

MaxEnt is another important scale-related question that future research should address. The same question would apply to any species distribution modeling framework in which the response variable could be defined as binary or as counts (generalized linear models and other regression frameworks).

The performance of models trained at increasingly larger extents when predicting panda distribution in the smallest extent yielded interesting results. The models trained at the smallest total extent were slightly outperformed by the models trained at the next-largest total extent, suggesting that there may have been environmental responses detected by the latter models that proved important when predicting at the smaller, nested total extent. The models trained at two largest total extent models performed much worse than smaller total extent models, however, indicating that there is a tradeoff between more fully capturing environmental responses and over-generalizing these responses when increasing total extent. Our visual analysis of the resulting suitability predictions of the smallest total extent highlighted the differences in their spatial patterns. Because predictions became more aggregated with increasing total extent and lost the detailed differences between high and low suitability areas, modeling at too large a total extent may have consequences for conclusions made about landscape connectivity. For example, the negative effects of roads are clear in the predictions of the models trained at the smaller total extents but become increasingly smoothed with increases to total extent beyond 500 km^2. This effect could result in researchers and managers to overestimate habitat connectivity across the landscape. The loss of fine-scale details in model predictions when using models built at a large total extent has been documented before (Anderson and Raza 2010, Vale et al. 2014), but our study is the first to quantify the effect of increasing total extent on model accuracy across multiple total extents and present evidence that total extent can be optimized as larger than the

area under investigation. Although previous research has indicated that total extent does not exhibit large effects on modeling landscape effects on connectivity (Cushman and Landguth 2010), our results have implications for modeling at landscape vs. local scales when attempting to model connectivity across high-disturbance areas. More investigation into these effects should be done with further testing of habitat models and movement data in the future.

A discussion of the effects of scale also informs knowledge on the ecology of giant pandas. In any ecological analysis of spatial scale, it is important to consider the relevance of the scales being analyzed to the species in question. As giant pandas generally use home ranges of around 5 km^2, all of our total extents represented areas larger than a single panda would respond to, while our grain sizes varied from a maximum of about a single panda home range to a minimum of an area that would represent foraging habitat selection (Connor et al. 2016). In terms of importance to the models, forest cover was relatively high at all total extents evaluated. Our finding that there was a generally positive effect of forest cover on habitat suitability at all extents and grain sizes (Fig. S9a) corroborates the widely accepted positive effect of forest on panda habitat in the literature (Liu et al. 2001, Zhang et al. 2011). That said, there has been recent evidence that pandas are able to utilize areas with bamboo but no forest cover (Hull et al. 2016), and there were either neutral or even slightly negative suitability response curves to forest cover in a subset of replicates from all three smaller total extents that suggested this (Fig. S7).

The effects of roads on panda habitat suitability also varied across scale, and the decrease in importance of the distance to roads with increasing total extent suggests that the anthropogenic effects which accompany roads most strongly effect pandas at small scales. There was a high amount of variation in the suitability response to roads across the replicates of the smallest total extent, however. (Fig. S8). Although most replicates featured models in which increasing

distance to roads increased suitability, some featured models in which there was no response to roads and others saw a decreasing response. This likely reflects the large variation in traffic and roadside anthropogenic development that occurs across the panda range, and any positive road responses are likely due to more habitat availability in areas closer to roads. The expected negative responses to roads became more consistent in the models trained at increasing total extent sizes, confirming that positive responses are the exception.

The dominant variable in the models across most total extent/grain size combinations was elevation, which has long been accepted as important in modeling panda habitat (Liu et al. 1999, Liu et al. 2001, Bearer et al. 2008, Qi et al. 2012). Elevation itself is not directly important to panda habitat by itself, however, and is more of a proxy variable that relates with other factors such as climate variability (Daly 2006)and in turn bamboo growth patterns) and anthropogenic disturbance. This is apparent in our study with the use of additional predictors like the five bioclimatic variables, which at the largest total extent reduced the importance of elevation from 60.7 to 15.3%. At the range-wide scale it was thus climatic factors that predicted habitat suitability better than elevation alone. At smaller total extents elevation likely represents the importance of local variations in climate due to topography and the importance it has on understory bamboo, while at larger total extents it is the regional variation in climate and both the physiological constraints and vegetation patterns (Zhang et al. 2018) that determine panda distribution. This is an example where although elevation is an effective variable for habitat prediction, its usefulness for explaining ecological phenomenon by itself is limited.

It is also important to consider the relevance and application to management when selecting different spatial scales. In the case of the giant panda, management occurs at total extents as small as parcels patrolled on foot, to entire nature reserves that approximate and

sometimes exceed our 500 km^2 extent (Liu et al. 2016). Management also occurs at the provincial and national levels, approximating our third and largest total extents, respectively. This wide range in management hierarchies makes it important to consider the total extent at which panda distributions are modeled for a given question, which may not necessarily be better analyzed at the total extent of interest (e.g., our second-smallest total extent models best predicted the smallest total extent distributions while our second-largest total extent models were overall the most accurate when predicting to the total extent at which they were trained). We thus recommend, in addition to a *priori* consideration of scale, for researchers and managers to test for the effects of scale together with the use of additional variables on model performance. Optimizing the selection of spatial scale and variables in this manner will improve the use of species distribution modeling in both ecological research and wildlife management.

APPENDIX

Table 3.1. The number of presence cells and their prevalence (presence cells divided by the total cells used) at each extent and grain size.

Total extent	Total extent 1 (50 km^2)		Total extent 2 (500 km^2)		Total extent 3 (Mountain ranges, mean 18,264 $km^{2)}$		Total extent 4 (Geographic range, 109,585 km^2)	
Grain size (m)	Mean presence cells (SD)	Mean Prevalence (SD)	Mean presence cells (SD)	Mean Prevalence (SD)	Mean presence cells (SD)	Mean Prevalence (SD)	Mean presence cells (SD)	Mean Prevalence (SD)
30	30.5 (11.3)	5.5E-04 (2.0E-03)	154.4 (77.8)	2.5E-04 (1.2E-04)	1137.6 (845.1)	3.3E-05 (2.4E-05)	4707 (129.0)	9.3E-06 (2.5E-07)
60	25.4 (6.1)	1.8E-03 (4.0E-04)	153.7 (77.3)	9.7E-04 (4.9E-04)	1130.9 (837.9)	1.3E-04 (9.6E-05)	4679 (128.1)	3.7E-05 (1.0E-06)
120	24.8 (5.9)	7.1E-03 (1.7E-03)	151.1 (75.8)	3.9E-03 (2.0E-03)	1113.1 (822.9)	5.1E-04 (3.8E-04)	4613.8 (126.2)	1.5E-04 (4.0E-06)
240	22.9 (5.7)	0.03 (6.5E-03)	144.3 (71.6)	0.01 (7.3E-03)	1065.7 (776.5)	2.0E-03 (1.4E-03)	4423.3 (109.7)	5.6E-04 (1.4E-05)
480	19.8 (5.1)	0.09 (0.02)	127.9 (60.5)	0.05 (0.02)	952.1 (676.0)	7.0E-03 (5.0E-03)	3975.3 (85.0)	2.0E-03 (4.3E-05)
960	14.5 (4.0)	0.26 (0.07)	98.7 (43.9)	0.16 (0.07)	753.3 (509.8)	0.02 (0.01)	3159.3 (57.9)	6.4E-03 (1.2E-04)
1920	8.1 (2.9)	0.67 (0.24)	62.3 (23.8)	0.40 (0.15)	516.7 (323.1)	0.06 (0.04)	2172 (29.1)	0.02 (2.4E-04)

Table 3.2. Summary of environmental variables and grain sizes used in the study.

Variable name (unit)	Grain sizes (m)
Elevation (m)	30, 60, 120, 240, 480, 960, 1920
Slope (degrees)	30, 60, 120, 240, 480, 960, 1920
Percent forest (%)	30, 60, 120, 240, 480, 960, 1920
Distance to main road (m)	30, 60, 120, 240, 480, 960, 1920
Distance to stream (m)	30, 60, 120, 240, 480, 960, 1920
MODIS phenology metrics*	240, 480, 960, 1920
Annual mean temperature (°C * 10)	960, 1920
Temperature seasonality (standard deviation *100)	960, 1920
Temperature annual range (max temp – min temp)	960, 1920
Annual precipitation (mm)	960, 1920
Precipitation seasonality (coefficient of variation)	960, 1920

* MODIS phenology metrics used in the models were reduced to five principal components using PCA.

Table 3.3. Comparisons of parsimony between models containing different variable sets built at increasing extent. Proportions of model replicates that featured increased AICc values of greater than 2 and thus reduced parsimony (-), proportions of replicates that featured decreased AICc values of greater than 2 and thus improved parsimony (+), and proportions of replicates that featured changes in AICc values less than 2 and thus no change in parsimony (0) are presented.

	Total extent 1			Total extent 2			Total extent 3			Total extent 4		
Comparison	-	+	No change	-	+	No change	-	+	No change	-	+	No change
Base: Base + Phenology	0.63	0.23	0.14	0.84	0.14	0.02	0.36	0.64	0	0	1	0
Base + Phenology: Base + Phenology + Climate	0.43	0.36	0.21	0.69	0.29	0.02	0.08	0.93	0	0	1	0
Base: Base + Phenology + Climate	0.53	0.29	0.18	0.84	0.13	0.02	0.05	0.95	0	0	1	0

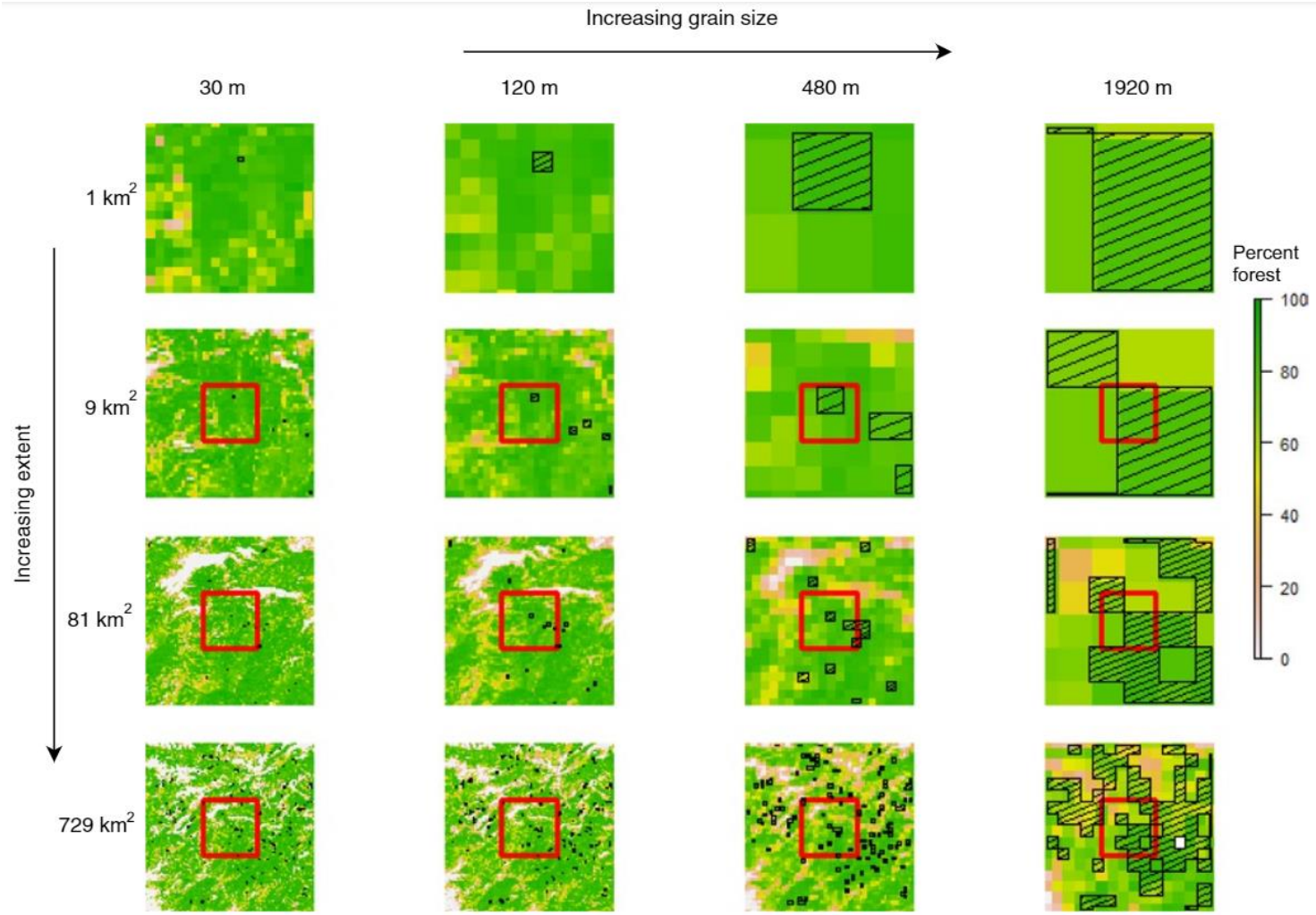

Figure 3.1. Visual representation of increasing spatial scale along two main axes – total extent and grain size. Percent forest is characterized at 4 extents and 4 grain sizes in an area around the middle of the giant panda range, with giant panda presence cells depicted with a hatch pattern. The red squares in rows 2 to 4 indicate the previous total extent shown in the previous row. Note: these are not the total extents examined in this study, they are purely a visual demonstration of the effects of changing spatial scale on environmental and presence data.

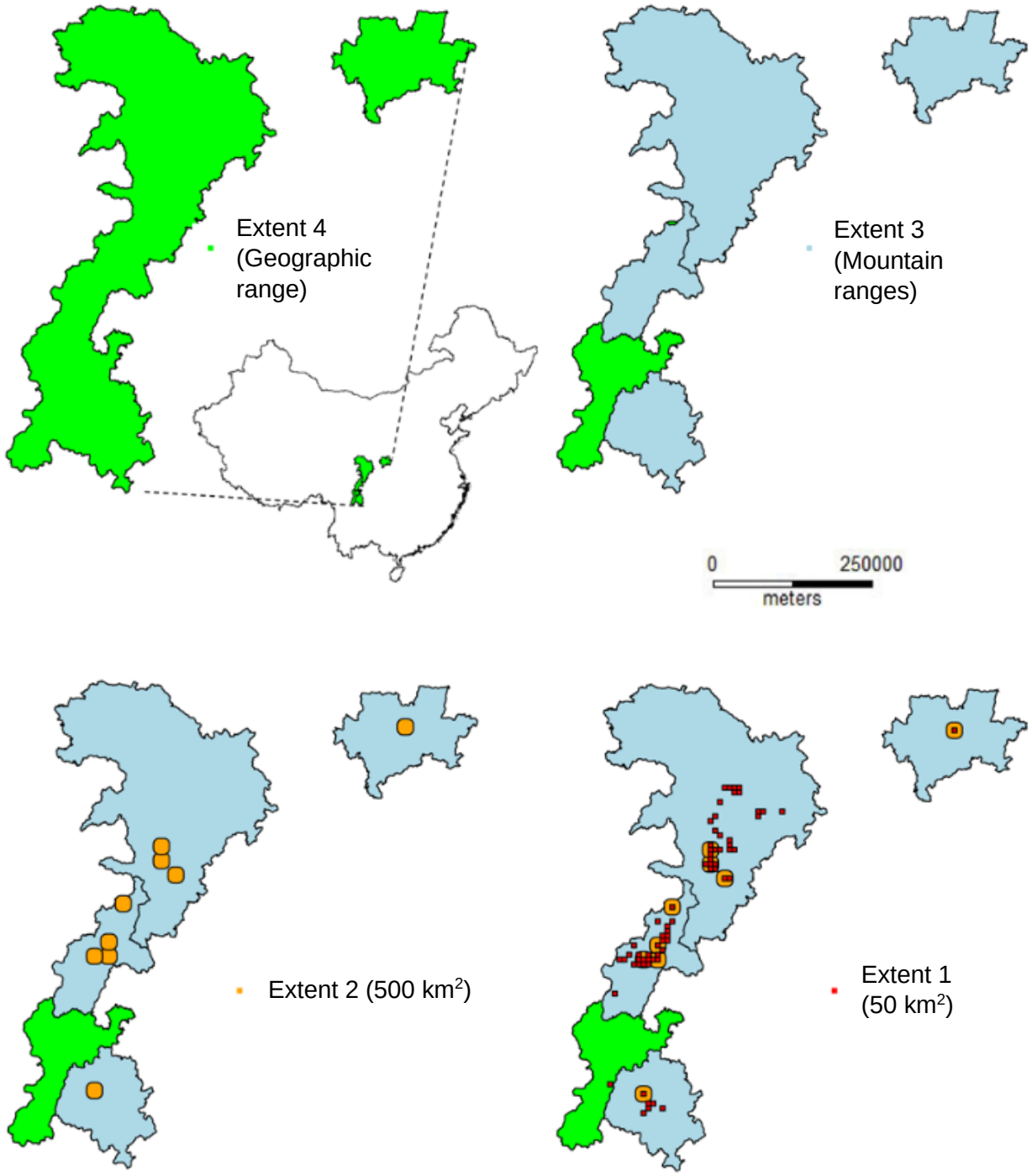

Figure 3.2. Total extents used in the analysis of scale effects on modeling giant panda habitat suitability and distribution. The smallest total extent was replicated 68 times, total extent 2 was replicated nine times, and total extent 3 was replicated four times across the current panda geographic range, which served as total extent 4.

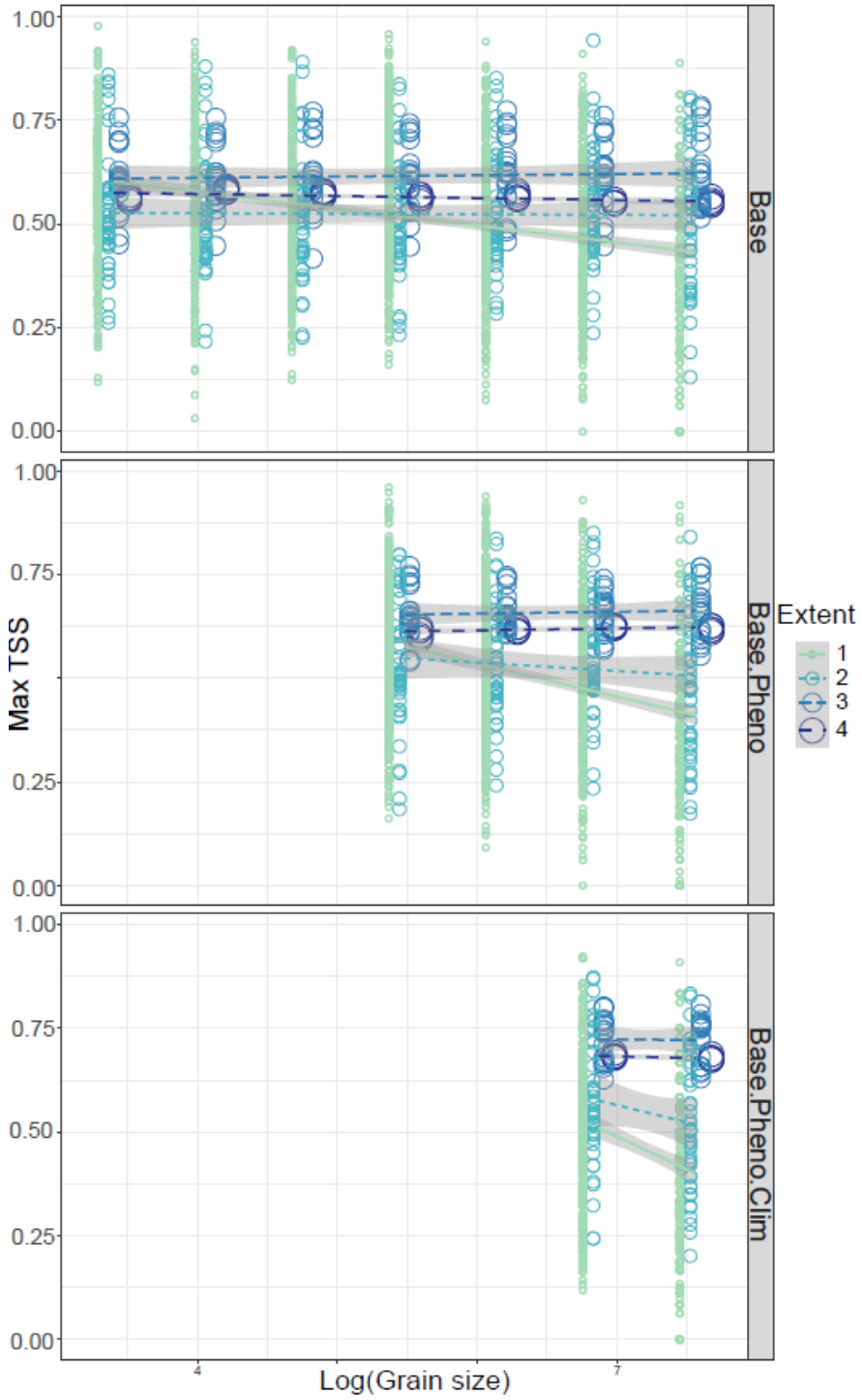

Figure 3.3. The effect of total extent, grain size, and variable combination on model accuracy (measured by the maximum True Skill Statistic (TSS)). Different variable combinations are presented in different panels - "Base" models include elevation, percent forest, slope, distance to main road, and distance to stream variables, "Base.Pheno" models include these five variables plus five phenology variables derived from MODIS data and PCA, and "Base.Pheno.Clim" models include all of these variables in addition to the five bioclimatic variables listed in Table 3.2.

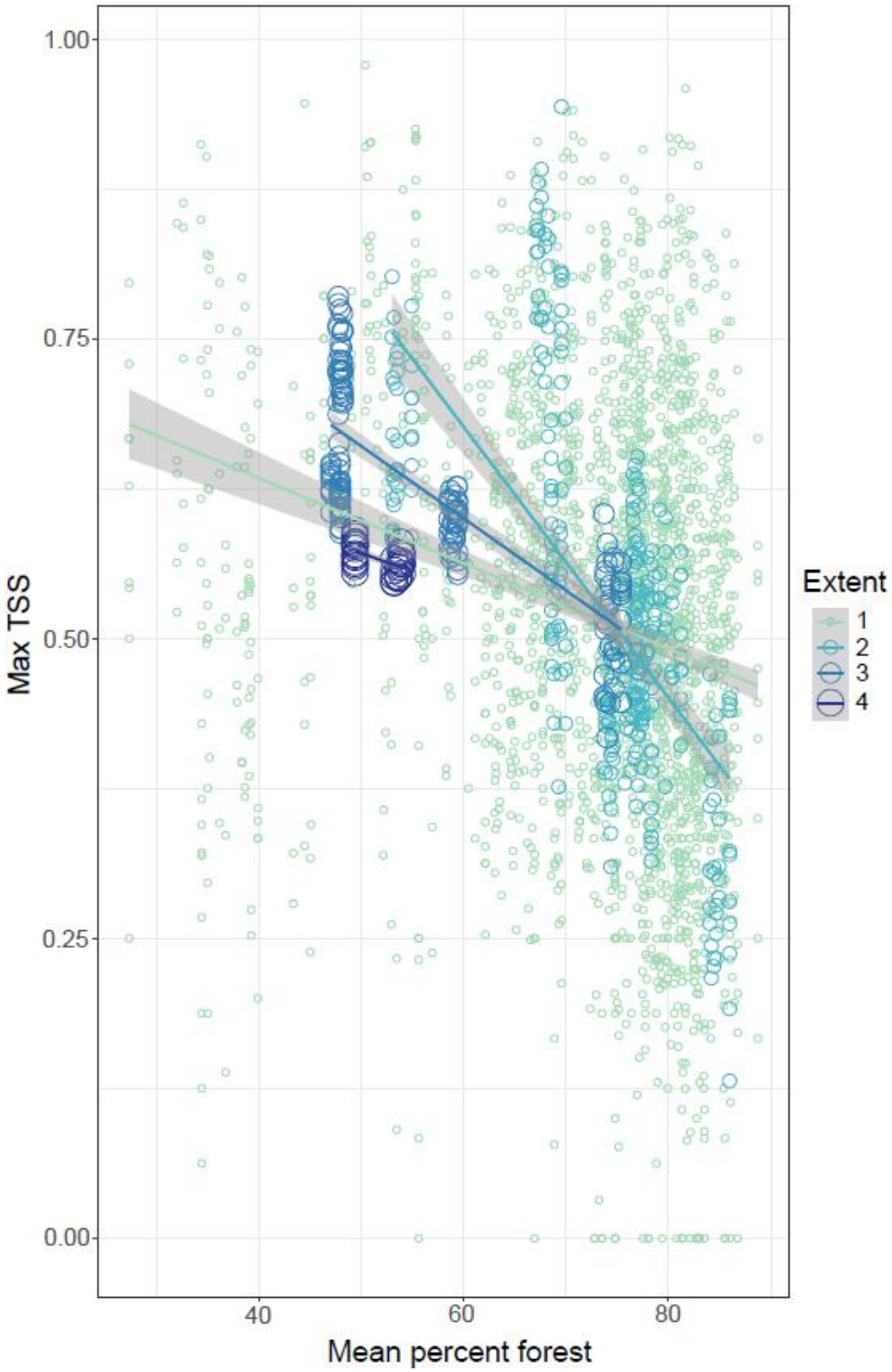

Figure 3.4. The effect of the position of a study area along an environmental gradient (percent forest cover) on model accuracy, as well as its impact on the effect of total extent on model accuracy. Evaluation scores of models built at different total extents are represented by different colors and differently sized circles.

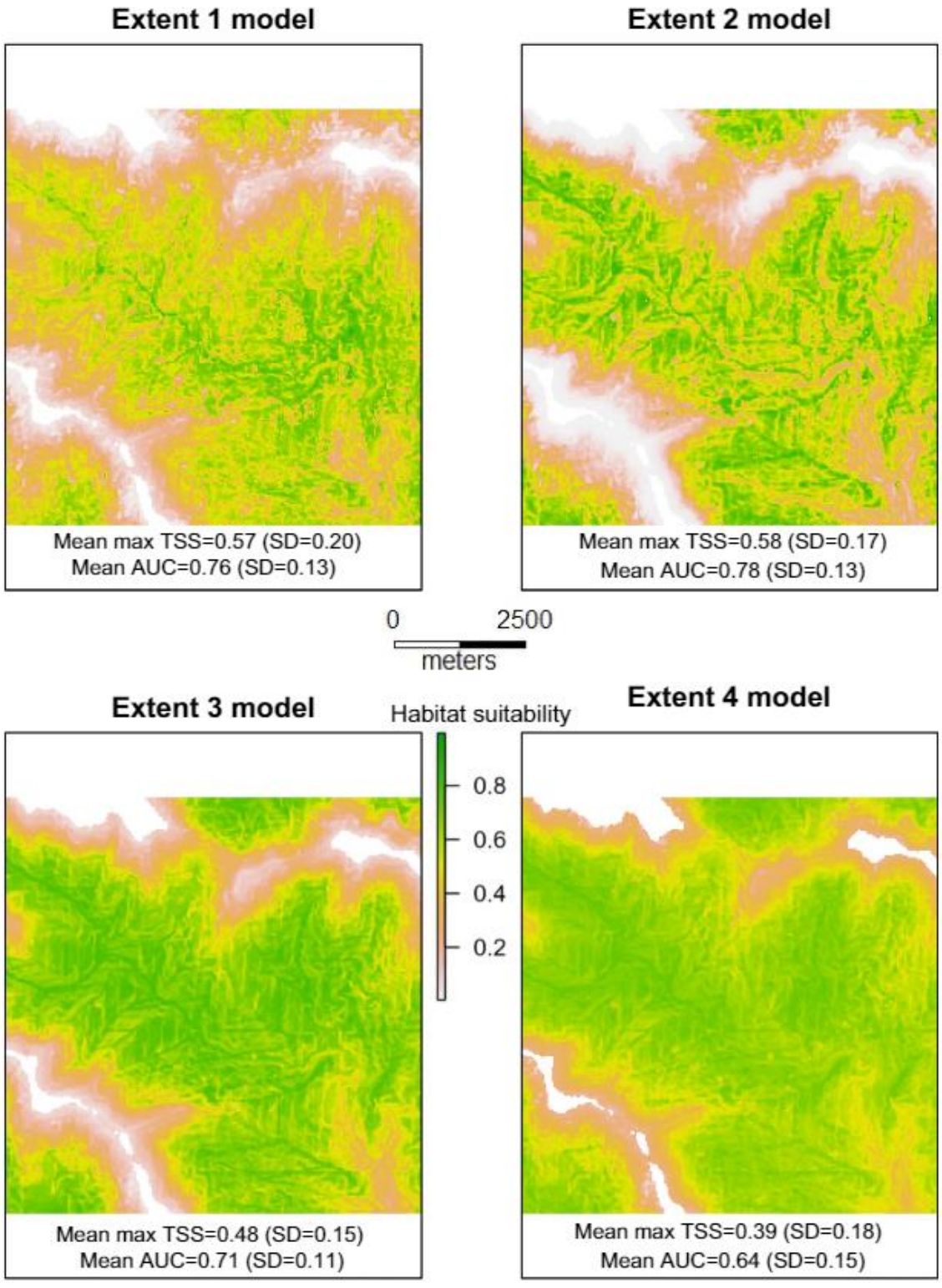

Figure 3.5. Giant panda habitat suitability predictions of one of the study area replicates of the smallest extent produced by models using the base variables at the 30-m grain size and trained at increasingly larger extents. Accuracy statistics for models trained at a given total extent are reported under the predicted suitability maps produced by that respective model.

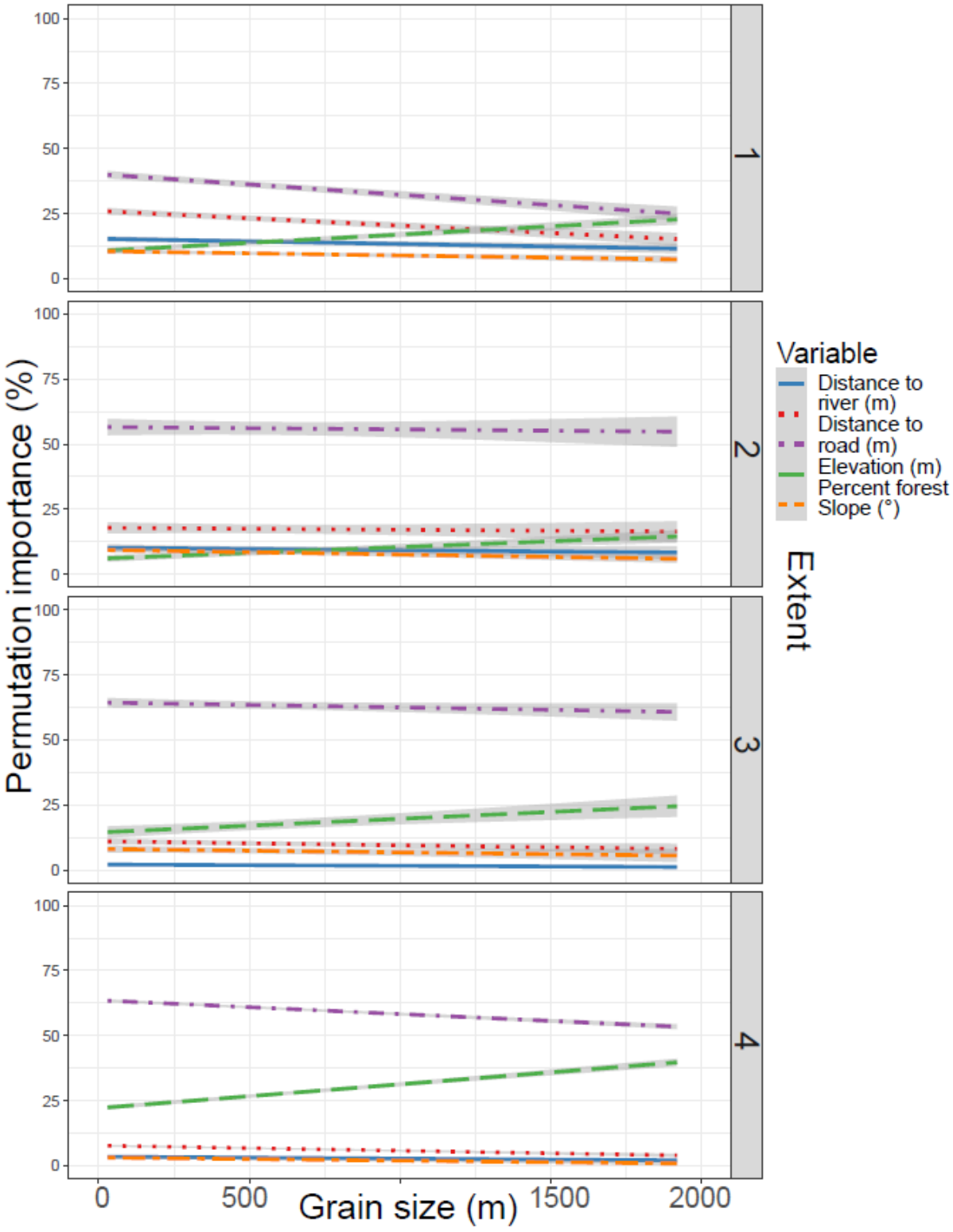

Figure 3.6. Variable permutation importance values (in percent) for the variables included in the MaxEnt habitat suitability models incorporating the base variable set. Grain size is plotted on the x-axis and models built at the different total extents are separated by panel.

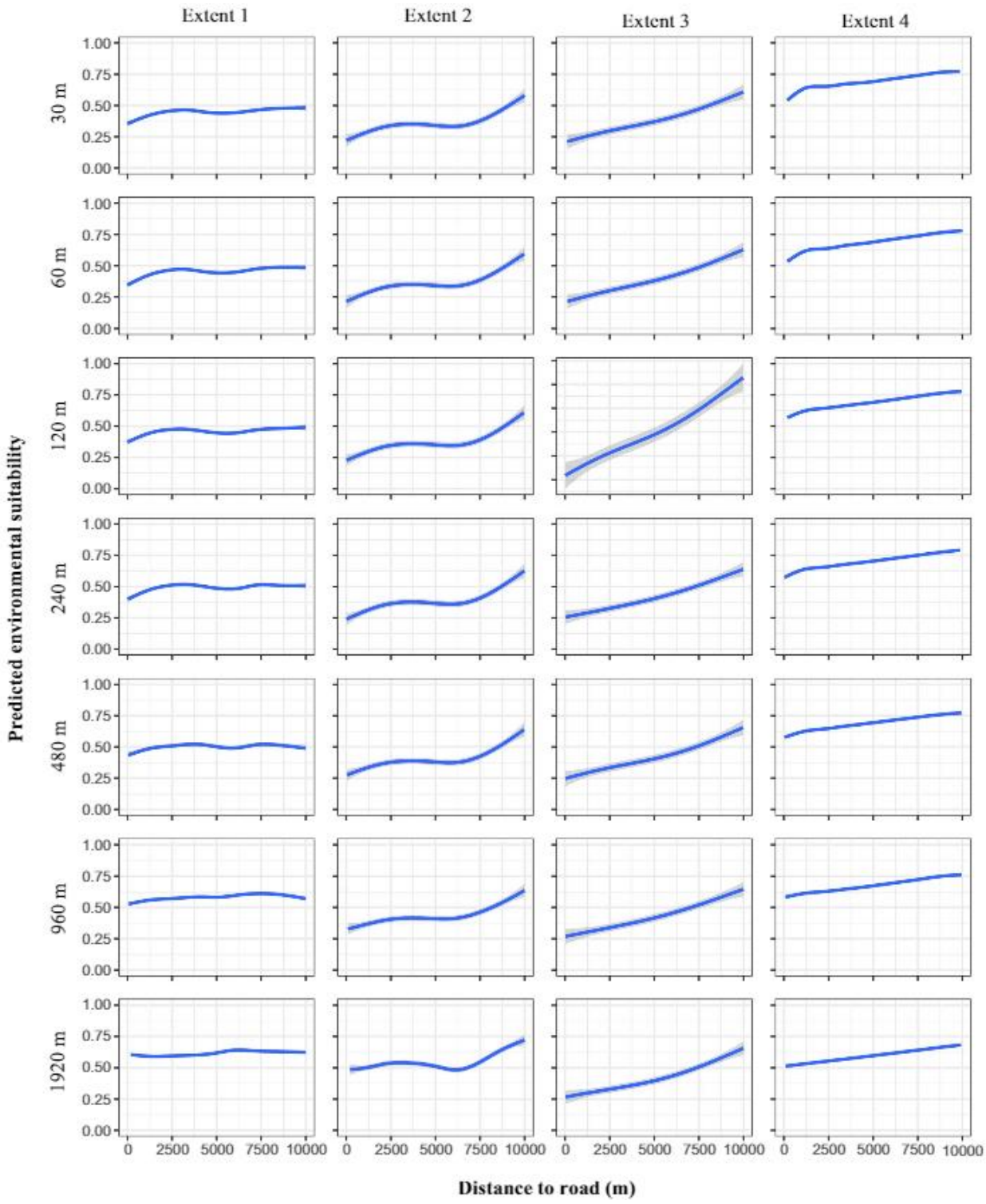

Figure 3.7. Responses of the base model predictions of giant panda habitat suitability in one of the sampled study areas to changes in the distance to road variable at increasing (left to right) total extent and increasing (top to bottom) grain size.

REFERENCES

REFERENCES

Akaike, H. 1974. STOCHASTIC THEORY OF MINIMAL REALIZATION. Ieee Transactions on Automatic Control **AC19**:667-674.

Allouche, O., A. Tsoar, and R. Kadmon. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology **43**:1223-1232.

Anderson, R. P., and A. Raza. 2010. The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus Nephelomys) in Venezuela. Journal of Biogeography **37**:1378-1393.

Austin, M. P., and K. P. Van Niel. 2011. Improving species distribution models for climate change studies: variable selection and scale. Journal of Biogeography **38**:1-8.

Bearer, S., M. Linderman, J. Y. Huang, L. An, G. M. He, and J. Q. Liu. 2008. Effects of fuelwood collection and timber harvesting on giant panda habitat use. Biological Conservation **141**:385-393.

Burnham, K. P., and D. R. Anderson. 2004. Multimodel inference - understanding AIC and BIC in model selection. Sociological Methods & Research **33**:261-304.

Connor, T., V. Hull, and J. G. Liu. 2016. Telemetry research on elusive wildlife: A synthesis of studies on giant pandas. Integrative Zoology **11**:295-307.

Connor, T., V. Hull, A. Vina, A. Shortridge, Y. Tang, J. D. Zhang, F. Wang, and J. G. Liu. 2018. Effects of grain size and niche breadth on species distribution modeling. Ecography **41**:1270-1282.

Cushman, S. A., and E. L. Landguth. 2010. Scale dependent inference in landscape genetics. Landscape Ecology **25**:967-979.

Daly, C. 2006. Guidelines for assessing the suitability of spatial climate data sets. International Journal of Climatology **26**:707-721.

Deblauwe, V., V. Droissart, R. Bose, B. Sonke, A. Blach-Overgaard, J. C. Svenning, J. J. Wieringa, B. R. Ramesh, T. Stevart, and T. L. P. Couvreur. 2016. Remotely sensed temperature and precipitation data improve species distribution modelling in the tropics. Global Ecology and Biogeography **25**:443-454.

DeCesare, N. J., M. Hebblewhite, F. Schmiegelow, D. Hervieux, G. J. McDermid, L. Neufeld, M. Bradley, J. Whittington, K. G. Smith, L. E. Morgantini, M. Wheatley, and M.

Musiani. 2012. Transcending scale dependence in identifying habitat with resource selection functions. Ecological Applications **22**:1068-1083.

Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carre, J. R. G. Marquez, B. Gruber, B. Lafourcade, P. J. Leitao, T. Munkemuller, C. McClean, P. E. Osborne, B. Reineking, B. Schroder, A. K. Skidmore, D. Zurell, and S. Lautenbach. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography **36**:27-46.

Elith, J., S. J. Phillips, T. Hastie, M. Dudik, Y. E. Chee, and C. J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions **17**:43-57.

Fourcade, Y., A. G. Besnard, and J. Secondi. 2018. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. Global Ecology and Biogeography **27**:245-256.

Guan, T. P., J. R. Owens, M. H. Gong, G. Liu, Z. Y. Ouyang, and Y. L. Song. 2016. Role of New Nature Reserve in Assisting Endangered Species Conservation - Case Study of Giant Pandas in the Northern Qionglai Mountains, China. Plos One **11**.

Guisan, A., C. H. Graham, J. Elith, F. Huettmann, and N. S. Distri. 2007a. Sensitivity of predictive species distribution models to change in grain size. Diversity and Distributions **13**:332-340.

Guisan, A., N. E. Zimmermann, J. Elith, C. H. Graham, S. Phillips, and A. T. Peterson. 2007b. What matters for predicting the occurrences of trees: Techniques, data, or species' characteristics? Ecological Monographs **77**:615-630.

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.

Hernandez, P. A., C. H. Graham, L. L. Master, and D. L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography **29**:773-785.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology **25**:1965-1978.

Hijmans, R. J., S. Phillips, J. Leathwick, and Ja Elith. 2017. dismo: Species Distribution Modeling. R package version 1.1-4. https://CRAN.R-project.org/package=dismo.

Hull, V., G. Roloff, J. D. Zhang, W. Liu, S. Q. Zhou, J. Y. Huang, W. H. Xu, Z. Y. Ouyang, H. M. Zhang, and J. G. Liu. 2014. A synthesis of giant panda habitat selection. Ursus **25**:148-162.

Hull, V., J. D. Zhang, J. Y. Huang, S. Q. Zhou, A. Vina, A. Shortridge, R. G. Li, D. A. Liu, W. H. Xu, Z. Y. Ouyang, H. M. Zhang, and J. G. Liu. 2016. Habitat Use and Selection by Giant Pandas. Plos One **11**:18.

Johnson, D. H. 1980. THE COMPARISON OF USAGE AND AVAILABILITY MEASUREMENTS FOR EVALUATING RESOURCE PREFERENCE. Ecology **61**:65-71.

Khosravi, R., M. R. Hemami, M. Malekian, A. L. Flint, and L. E. Flint. 2016. Maxent modeling for predicting potential distribution of goitered gazelle in central Iran: the effect of extent and grain size on performance of the model. Turkish Journal of Zoology **40**:574-585.

Levin, S. A. 1992. THE PROBLEM OF PATTERN AND SCALE IN ECOLOGY. Ecology **73**:1943-1967.

Liu, J., V. Hull, W. Yang, A. Viña, X. Chen, and H. Zhang. 2016. Pandas and People - Coupling Human and Natural Systems for Sustainability. Page 304. Oxford University Press, U.K.

Liu, J. G., M. Linderman, Z. Y. Ouyang, L. An, J. Yang, and H. M. Zhang. 2001. Ecological degradation in protected areas: The case of Wolong Nature Reserve for giant pandas. Science **292**:98-101.

Liu, J. G., Z. Ouyang, W. W. Taylor, R. Groop, K. C. Tan, and H. M. Zhang. 1999. A framework for evaluating the effects of human factors on wildlife habitat: the case of giant pandas. Conservation Biology **13**:1360-1370.

Lobo, J. M., A. Jimenez-Valverde, and R. Real. 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography **17**:145-151.

Manzoor, S. A., G. Griffiths, and M. Lukac. 2018. Species distribution model transferability and model grain size - finer may not always be better. Scientific Reports **8**:9.

Merow, C., M. J. Smith, and J. A. Silander. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography **36**:1058-1069.

National Geomatics Center of China. 2018. National Fundamental Geographic Information Database.

Osborne, P. E., G. M. Foody, and S. Suarez-Seoane. 2007. Non-stationarity and local approaches to modelling the distributions of wildlife. Diversity and Distributions **13**:313-323.

Pan, W. 2014. A chance for lasting survival: Ecology and behavior of wild giant pandas. Smithsonian Institution.

Phillips, S. J., R. P. Anderson, M. Dudik, R. E. Schapire, and M. E. Blair. 2017. Opening the black box: an open-source release of Maxent. Ecography **40**:887-893.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling **190**:231-259.

Phillips, S. J., M. Dudik, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications **19**:181-197.

Qi, D. W., S. N. Zhang, Z. J. Zhang, Y. B. Hu, X. Y. Yang, H. J. Wang, and F. W. Wei. 2012. Measures of Giant Panda Habitat Selection Across Multiple Spatial Scales for Species Conservation. Journal of Wildlife Management **76**:1092-1100.

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/.

Renner, I. W., J. Elith, A. Baddeley, W. Fithian, T. Hastie, S. J. Phillips, G. Popovic, and D. I. Warton. 2015. Point process models for presence-only analysis. Methods in Ecology and Evolution **6**:366-379.

Renner, I. W., and D. I. Warton. 2013. Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology. Biometrics **69**:274-281.

Saab, V. 1999. Importance of spatial scale to habitat use by breeding birds in riparian forests: A hierarchical analysis. Ecological Applications **9**:135-151.

Schaller, George B. 1985. Giant pandas of Wolong. University of Chicago press, Chicago.

Seo, C., J. H. Thorne, L. Hannah, and W. Thuiller. 2009. Scale effects in species distribution models: implications for conservation planning under climate change. Biology Letters **5**:39-43.

State Forestry Administration. 2006. The 3rd National Survey Report on Giant Panda in China. Science Publisher, Beijing, China (in Chinese).

Stockwell, D. R. B., and A. T. Peterson. 2002. Effects of sample size on accuracy of species distribution models. Ecological Modelling **148**:1-13.

Tang, Y., J. A. Winkler, A. Vina, J. G. Liu, Y. B. Zhang, X. F. Zhang, X. H. Li, F. Wang, J. D. Zhang, and Z. Q. Zhao. 2018. Uncertainty of future projections of species distributions in mountainous regions. Plos One **13**.

Thomas, K., T. Keeler-Wolf, and J. Franklin. 2002. A comparison of fine- and coarse-resolution environmental variables toward predicting vegetation distribution in the Mojave Desert. Predicting Species Occurrences: Issues of Accuracy and Scale:133-139.

Thompson, C. M., and K. McGarigal. 2002. The influence of research scale on bald eagle habitat selection along the lower Hudson River, New York (USA). Landscape Ecology **17**:569-586.

Trivedi, M. R., P. M. Berry, M. D. Morecroft, and T. P. Dawson. 2008. Spatial scale affects bioclimate model projections of climate change impacts on mountain plants. Global Change Biology **14**:1089-1103.

Tuanmu, M. N., A. Vina, S. Bearer, W. H. Xu, Z. Y. Ouyang, H. M. Zhang, and J. G. Liu. 2010. Mapping understory vegetation using phenological characteristics derived from remotely sensed data. Remote Sensing of Environment **114**:1833-1844.

Tuanmu, M. N., A. Vina, W. Yang, X. D. Chen, A. M. Shortridge, and J. G. Liu. 2016. Effects of payments for ecosystem services on wildlife habitat recovery. Conservation Biology **30**:827-835.

Turner, M. G., R. V. O'Neill, R. H. Gardner, and B. T. Milne. 1989. Effects of changing spatial scale on the analysis of landscape pattern. Landscape Ecology **3**:153-162.

USGS Shuttle Radar Topography Mission. 2000. 1 Arc Second scene. Global Land Cover Facility, University of Maryland. College Park, Maryland.

Vale, C. G., P. Tarroso, and J. C. Brito. 2014. Predicting species distribution at range margins: testing the effects of study area extent, resolution and threshold selection in the Sahara-Sahel transition zone. Diversity and Distributions **20**:20-33.

Vina, A., W. Liu, S. Q. Zhou, J. Y. Huang, and J. G. Liu. 2016. Land surface phenology as an indicator of biodiversity patterns. Ecological Indicators **64**:281-288.

Vina, A., M. N. Tuanmu, W. H. Xu, Y. Li, Z. Y. Ouyang, R. DeFries, and J. G. Liu. 2010. Range-wide analysis of wildlife habitat: Implications for conservation. Biological Conservation **143**:1960-1969.

Warren, D. L., and S. N. Seifert. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecological Applications **21**:335-342.

Wheatley, M. 2010. Domains of scale in forest-landscape metrics: Implications for species-habitat modeling. Acta Oecologica-International Journal of Ecology **36**:259-267.

Wiens, J. A. 1989. SPATIAL SCALING IN ECOLOGY. Functional Ecology **3**:385-397.

Yang, H. B., A. Vina, Y. Tang, J. D. Zhang, F. Wang, Z. Q. Zhao, and J. G. Liu. 2017. Range-wide evaluation of wildlife habitat change: A demonstration using Giant Pandas. Biological Conservation **213**:203-209.

Zeller, K. A., T. W. Vickers, H. B. Ernest, and W. M. Boyce. 2017. Multi-level, multi-scale resource selection functions and resistance surfaces for conservation planning: Pumas as a case study. Plos One **12**.

Zhang, Y. K., P. D. Mathewson, Q. Y. Zhang, W. P. Porter, and J. H. Ran. 2018. An ecophysiological perspective on likely giant panda habitat responses to climate change. Global Change Biology **24**:1804-1816.

Zhang, Z. J., R. R. Swaisgood, S. N. Zhang, L. A. Nordstrom, H. J. Wang, X. D. Gu, J. C. Hu, and F. W. Wei. 2011. Old-growth forest is what giant pandas really need. Biology Letters **7**:403-406.

Zhao, S. C., P. P. Zheng, S. S. Dong, X. J. Zhan, Q. Wu, X. S. Guo, Y. B. Hu, W. M. He, S. N. Zhang, W. Fan, L. F. Zhu, D. Li, X. M. Zhang, Q. Chen, H. M. Zhang, Z. H. Zhang, X. L. Jin, J. G. Zhang, H. M. Yang, J. Wang, and F. W. Wei. 2013. Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. Nature Genetics **45**:67-U99.

CHAPTER 4: THE EFFECTS OF HABITAT AMOUNT AND FRAGMENTATON ON FUNCTIONAL CONNECTIVITY IN A GIANT PANDA POPULATION

In collaboration with:

Maiju Qiao, Jindong Zhang, Kim Scribner, Wenke Bai, Jianguo Liu

Abstract

Habitat loss and fragmentation are widely recognized threats to wildlife populations, but the relative importance of habitat fragmentation and habitat amount is debated. Functional connectivity, defined here as the successful dispersal and reproduction of animals (gene flow) across landscapes, is an example of an important factor contributing to the persistence of species in the environment. The relative effects of habitat amount and habitat fragmentation on functional connectivity need further investigation. To accomplish this, we combined species distribution modeling and landscape genetics techniques to estimate the effects of habitat amount and habitat fragmentation on functional connectivity in a giant panda (*Ailuropoda melanoleuca*) population. We found that a metric of habitat fragmentation which measures the variation in the amount of core vs. edge habitat area across multiple patches (CAI_SD) alone was the best predictor of functional connectivity, while habitat amount alone resulted in the second-ranked model. Importantly, functional connectivity was maximized when about 80 percent of the local landscape was habitat, above which further increases in the amount of habitat rapidly reduced functional connectivity. There was a similar threshold in the top fragmentation variables – functional connectivity was maximized at an intermediate level of habitat fragmentation. This suggests that areas with large tracts of contiguous habitat act as a "trap" in which individuals choose not to disperse while areas of more habitat heterogeneity force individuals to disperse further on the landscape. Our results offer new insight for the prioritization of habitat for conservation and restoration across our study area, and our methods can be applied to any species and landsdcape in which habitat and gene flow can be estimated.

4.1 Introduction

Habitat loss and fragmentation are widely considered the greatest threats to wildlife populations in the modern era (Brooks et al. 2002, Krauss et al. 2010). The relative importance of habitat fragmentation per se, defined as the effects of changing the configuration of existing habitat without changing the amount of habitat in the system, has been the subject of increasing debate (Fletcher et al. 2018, Fahrig et al. 2019). Fahrig (2017) argues that the widespread view that habitat fragmentation is negative and should be minimized is an outdated idea that should be revised. In a review of 118 studies, she found that the majority of ecological responses to habitat fragmentation were not significant, and most of those that were significant were positive (Fahrig 2017). Other studies and reviews have argued that habitat fragmentation has clear negative effects on important ecological responses and functions (Haddad et al. 2015, Fletcher et al. 2018).

One important ecological response to habitat loss and fragmentation is functional connectivity – the successful dispersal and reproduction of individuals across landscapes. Functional connectivity between populations is important for the long-term health and survival of species because it helps maintain levels of genetic diversity necessary to adapt to changing environmental conditions, allows for the spread of adaptive genes when they become advantageous, and helps to suppress deleterious inbreeding effects that arise in small, isolated populations (Frankham 1995, (Eizaguirre and Baltazar-Soares 2014, Bay et al. 2017). It is also important to maintain functional connectivity in systems which feature discrete habitat patches across the landscape so that they can be recolonized following localized extinctions (Hanski 1998).

Like other ecological phenomena, the response of functional connectivity to habitat fragmentation is debated. Keyghobadi (2007) conducted a review of 32 empirical studies that compared population genetic relationships in both a control and "fragmented" landscape and found that 22 (69%) showed increased genetic differentiation (reduced functional connectivity) in the fragmented landscapes and 10 (31%) showed either no effect or greater differentiation in the control landscapes. She noted that most of these studies conflated habitat loss and fragmentation and did not disentangle their effects, however, a frequent problem in any ecological research on habitat fragmentation (Keyghobadi 2007, Smith et al. 2009). Opposite conclusions regarding the relative importance of habitat amount and fragmentation on functional connectivity have also been reported in simulation studies - Cushman et al. (2012) found that habitat fragmentation was the main predictor of genetic distance between individuals simulated on varied habitat/non-habitat landscapes while Jackson and Fahrig (2016) concluded that the amount of habitat in a landscape was a much stronger predictor of inter-individual genetic distance than its level of fragmentation.

Methodology from the field of landscape genetics, which combines techniques from landscape ecology and population genetics (Manel et al. 2003), is ideally suited to studying the effects of habitat variation on functional connectivity. A common method in landscape genetics studies is to transform environmental variables into surfaces based on hypothesized resistances to animal movements that result in an isolation by resistance (IBR) pattern (Manel and Holderegger 2013). Some of these studies have included measures of fragmentation of these environmental variables in their models. For example, Howell et al. (2015) found that the correlation length, a measure of contiguity, of coniferous forests positively predicted gene flow in reintroduced American marten populations. Draheim et al. (2018) developed estimates of change in the spatial

genetic structure in a black bear population over time and found that this change was best predicted by fragmentation metrics describing landscape change. The use of direct habitat estimates in landscape genetics has been less common, but several studies have used the inverse of habitat suitability predictions from species distribution models (SDMs) or resource selection functions (RSFs) to estimate resistances to gene flow (Weckworth 2013, Sacks et al. 2016).

The use of estimates of habitat amount and/or habitat fragmentation as spatially continuous environmental variables to investigate IBR patterns has been exceedingly rare. We posit that a novel combination SDM and landscape genetics methods in a novel workflow offers a powerful way to investigate the effects of habitat amount and fragmentation on functional connectivity. Here, we demonstrate our approach using a wild giant panda population in Southwest China. Giant pandas are habitat specialists that require understory bamboo forests and are sensitive to anthropogenic disturbance (Schaller et al. 1985). They also feature small remnant populations facing competing effects of anthropogenic habitat degradation and intensive conservation effort (Xu et al. 2017). We hypothesized that habitat fragmentation would have a greater effect on functional connectivity in giant pandas than habitat amount. We also hypothesized that habitat fragmentation would exhibit a parabolic relationship where increases in above zero would increase connectivity until a threshold at which further increases would reduce connectivity. Our reasoning for this is that individuals will be forced to disperse farther when habitat is broken apart, increasing gene flow on that landscape. We posit that this relationship is likely not linear, however, and that there exists a threshold at which further fragmentation reduces successful dispersal events between habitat patches and thus reduces gene flow. Our results should be directly applicable to conservation planning of giant panda protected areas in

order to maximize functional connectivity across the landscape. Our methods could also be applied to any other threatened wildlife species, with specific relevance to habitat specialists.

4.2 Methods

4.2.1 Overview

To investigate the effects of habitat amount and fragmentation on functional connectivity of a continuous giant panda population, we followed the general steps illustrated in Figure 4.1 and described here: 1) estimate interindividual genetic distances between pandas across our study area, 2) develop models of habitat suitability using relevant environmental variables and giant panda presence data, 3) predict habitat suitability across our study area, 4) convert continuous habitat suitability predictions to a binary habitat/nonhabitat map, 5) determine a relevant scale of effect and calculate the amount of habitat and habitat fragmentation metrics in the resulting local landscape in a moving window approach, 6) transform the resulting spatially continuous habitat amount/fragmentation variables into resistance surfaces, 7) utilize Circuitscape, a program that employs circuit theory to model probabilities of animal movement based on resistance surfaces, to estimate resistance distances between panda genetic sample locations, 8) build maximum likelihood population effects (MLPE) models predicting interindividual genetic distances with the resistance distances, 9) repeat steps 6-8 with different resistance transformations of the habitat amount and fragmentation variables in order to maximize the MLPE model likelihood using a genetic algorithm implemented through the ResistanceGA package (Peterman 2018).

4.2.2 Study area

We conducted noninvasive fecal genetics sampling across Wolong Nature Reserve, Sichuan, China. Wolong is an approximately 2000 km^2 national-level protected area centrally situated in the extant panda geographic range (Fig. 4.2). Recent population surveys suggest there

are approximately 150 giant pandas in the reserve (Qiao et al. 2019). Panda habitat in the reserve consists of understory bamboo forests occurring below 4,000 meters. In addition to forest and elevation requirements, important habitat variables include terrain ruggedness and anthropogenic disturbance levels (Hull et al. 2014). Rugged terrain induces additional energy expenditure that pandas preferentially avoid (Nie et al. 2015), and human activity has been shown to negatively affect panda occurrence (Zeng et al. 2019). Spatial variation in habitat factors has resulted in differences in the amount of habitat and level of habitat fragmentation in the landscape surrounding panda occurrence locations. Data available make Wolong an ideal study system to analyze the effects of habitat amount and fragmentation on functional connectivity.

4.2.3 Noninvasive fecal sampling

Fecal samples were collected throughout 2015 and 2016 in the reserve with a systematic sample design. Specifically, we subdivided known and potential habitat areas throughout the reserve, based on suitable elevation ranges (1100 – 4000 meters) into 2 km^2 survey cells that were searched in a zig-zag manner by experienced field workers and local guides to maximize the area searched. Fresh panda feces, judged by the status of the outer mucosal membrane, were collected whenever encountered, stored in sterile plastic bags, and frozen within eight hours of collection.

4.2.4 Genetic analyses

Seven microsatellite loci were selected for analysis from a large candidate set based on levels of polymorphism, a lack of genotyping error, and high amplification success rate even when feces were exposed to natural weather conditions for extended periods (Huang et al. 2015). To ensure locus utility for fecal samples, a pilot study on captive pandas was conducted to compare the results of DNA extracted from fecal samples with DNA extracted from blood

samples. All seven loci performed well – fecal and blood samples featured exact genotype matches in the n = 15 captive pandas (Huang et al. 2015). DNA was extracted using QIAamp stool mini kits according to the manufacturer's protocols (Qiagen, Germany). The seven loci were amplified separately by PCR using 25 μ reaction mixtures consisting of 50 ng of template DNA, 2 mm $MgCl_2$, 200 μmol of dNTP, 15 pmol of each primer, 1.0 μg of bovine serum albumin (BSA), and 0.3 units of Hotstart DNA polymerase. The following PCR protocol was used for amplifications: an initial denaturation step of at 95°C for 5 minutes, followed by 35 cycles of 95°C for 45 s, 30 s at a locus-specific annealing temperature and 50 s at 72°C, and a final elongation for 10 min at 72°C. PCR amplification products were separated by capillary electrophoresis using a denaturing acrylamide gel matrix on an ABI 3730xl Genetic Analyzer in order to genotype samples. Specific alleles were detected using Genemapper 3.2 software. Loci were independently scored by at least two trained individuals. A subset of the samples were analyzed for sex information, but because the full dataset lacked results we did not include sex in our analyses.

As a measure of quality control, we used a multiple-tubes approach in which we amplified every sample at least three times for each locus (Taberlet et al. 1996). To avoid the 'shadow effect', in which heterozygotes are falsely genotyped as homozygotes due to failed amplifications, we did not accept a single-locus genotype until we had at least three identical homozygote or two identical heterozygote profiles (Mills et al. 2000). If two individual genotypes differed by only one allele across the seven loci, we re-extracted DNA and repeated the PCR procedure three additional times to confirm results. As an additional method of quality control, we used MicroChecker to look for evidence of allelic dropout and FreeNA to estimate null allele frequency (Oosterhaut et al. 2004, Chapuis and Estoup 2007). Because two unique

individuals may share the same genotype purely by chance, we calculated this probability of identity (PID) and probability of identity of full siblings (PSIB) across our seven loci based on the method developed by Waits et al. (2001) and implemented in the GenAlEx program (Peakall and Smouse 2006, 2012) to ensure we used different individuals in our analyses. We calculated observed and expected heterozygosity for each locus in GenAlEx. We also tested each locus for significant divergence from Hardy-Weinberg Equilibrium (HWE) using a Bonferroni correction in GenAlEx. In order to calculate genetic distance between individuals we used Smouse and Peakall's (1999) definition of codominant genotypic distance implemented through GenAlEx, which has been shown to effectively measure genetic structure when the number of loci available is relatively small (Smouse and Peakall 2006, 2012, Draheim et al. 2015). Missing information at one locus was allowed to maximize our sample size. If an individual was captured multiple times across the study area, its location for the landscape genetic analyses was defined as the median X and Y coordinates from its capture locations. We used the median as opposed to the mean location to avoid excessive placement of individuals in unrealistic habitat locations on the landscape.

4.2.5 Modeling habitat

4.2.5.1 Giant panda presence data

We built habitat models across our study area with giant panda presence gathered from the 4th national giant panda survey, a consistently sampled geographic range-wide effort to determine panda distributions throughout Southwest China (State Council 2015). We selected only presence locations that fell within our noninvasive genetics survey extent in Wolong. This subset of 4^{th} national survey presence data was collected throughout Wolong in 2012. We then tested the ability of the models to predict giant panda habitat across our study area with the

locations of fecal samples collected for our genetic analyses in 2015-2016. Our models were thus trained with an independently collected dataset and evaluated with our genetic sample locations – the most relevant locations for our later landscape genetics analyses.

4.2.5.2 Environmental predictor data

To predict giant panda habitat suitability, we used a variable set consisting of elevation (m), aspect (degrees), terrain ruggedness, tree cover (%), the distance to a main road (m), and the distance to stream and river (m) (Table 4.1). We obtained elevation data at a grain size of 27.35 m from the National Space Agency's Shuttle Radar Topography (SRTM) mission that was void-filled by the USGS (USGS 2000). We calculated aspect and terrain ruggedness from this elevation data using the 'terrain' function in the 'raster' R package (Hijmans et al. 2019, R Core team 2019). Terrain ruggedness is defined in this function as the mean of the absolute differences between the elevation value of a cell and the elevation values of its surrounding eight cells. We obtained percent tree cover data at a grain size of 27 m from the global dataset developed by Hansen et al. (2013), which we resampled to 27.35 m to match the grain of elevation using bilinear interpolation. We obtained main road and stream maps at a scale of 1:250000 from the Chinese National Fundamental Geographic Information Database and created distance-to-feature rasters with a grain size of 27.35 m (National Geomatics Center of China 2018).

An animal's selection of a given location may depend largely on the surrounding environment (Zeller et al. 2014). To best describe the environmental conditions relevant to a given location, we used empirical movement distributions to weight and incorporate each environmental predictor's values in the surrounding area to the final value for that location. To derive these distributions, we used the 'moveHMM' package to build Hidden Markov Models (HMMs) of panda movement behavior using GPS-collar information (Michelot et al. 2016). Five

giant pandas were captured, fitted with GPS collars, and tracked between 2010 and 2011 to acquire the necessary movement data (Hull et al. 2015). Although this is a very small subset of Wolong's panda population, both sexes were represented and we determined that it may be useful to incorporate their movement behaviors into the environmental predictors. This research followed ASM guidelines and was approved by the MSU Institutional Animal Care and Use Committee. For further details on the GPS-collaring process and data collection see Hull et al. (2015). HMMs are based on the theory that animals have one or more unobserved behavioral modes that drive their movement patterns (Michelot et al. 2016). We used the distance traveled between consecutive location fixes (t = 3 hours) as step-length input into the HMMs and found that three movement "states" with their own step-length distributions best described panda movement behavior. We used these three step-length distributions to derive three different environmental predictor sets in which the variables were transformed with a moving window which weighted the surrounding are a according to each step-length distribution. We thus tested four different environmental variable sets when habitat modeling – the raw environmental variables and three sets of step-length-weighted environmental variables.

4.2.5.3 MaxEnt modeling

We used the MaxEnt modeling algorithm implemented through the 'dismo' R package to relate giant panda presence locations to environmental predictors (Phillips et al. 2006, Hijmans et al. 2017). It's one of the most popular methods of species distribution modeling and has performed well in comparisons between different techniques for predictive accuracy (Elith and Graham 2009). Another advantage of MaxEnt in our study is its formulation as a presence-only algorithm that does not require known absence locations but instead compares species presence locations to "background" environmental conditions described by random points across the study

area (Phillips et al. 2006). The algorithm makes use of machine learning techniques to minimize the relative entropy in the predicted suitability between the presence and background locations (Elith et al. 2011). We randomly selected 100,000 points across our study area to serve as background locations. This sample was ten times the default number generated by MaxEnt, which we chose so that we could better estimate background conditions in our study area (Renner and Warton 2013) due to the large number of raster cells within it (more than 10 million). We built four different MaxEnt models using these background locations, the presence locations from the 4th national survey, and the four different environmental variable sets described in the previous section. We then used the models and environmental variable sets to predict habitat suitability (0-1) across the study area and tested their accuracy using the presence locations from our fecal genetics survey and 68,270 random points to serve as background locations. This number of background locations was chosen so that our model training background locations vs. model testing background locations ratio matched the model training presence locations vs. model testing presence locations ratio used in our modeling procedure. The total model training and model testing extents remained the same – only the number of presence and background points differed. Several accuracy statistics including the area under the receiver operator curve (AUC), the True Skill Statistic (TSS), and the correlation between predicted suitability values between test presences and test background locations (Cor) were used to evaluate model performance in predicting habitat. Based on these results we then chose the model and associated habitat predictions which employed the set of environmental predictors that produced the highest accuracy. Finally, we converted the continuous habitat suitability surface into a binary habitat/nonhabitat map around the suitability threshold that maximized the value of the TSS. This threshold was thus that at which omission and commission errors in

predicting panda presences were minimized in order to create the most accurate depiction of habitat across the landscape (Allouche et al. 2006)

4.2.6 Habitat amount and fragmentation

In order to create variables measuring habitat amount and habitat fragmentation that could be used to predict gene flow, we needed spatially explicit estimates that would capture these values in the relevant surrounding landscape for each cell in our study area. To achieve this, we used a moving window approach on the binary habitat/nonhabitat map in which we calculated the amount of habitat and its level of fragmentation (as described by several metrics) in the area within a certain distance threshold surrounding the focal cell, which then took those values. We optimized this threshold at 42 cells (1149 m) away from the focal cell, resulting in an 85 x 85 cell (2325 m x 2325 m) matrix for the calculation of habitat amount and fragmentation. This threshold was optimized using the 'SS_optim.scale' function in the 'ResistanceGA' package, for details see section 4.2.7 (Peterman 2018). The resulting local landscape window of 5.40 km^2 also falls within the range of typical home range sizes of pandas in this region (Connor et al. 2016).

We chose to evaluate eight habitat fragmentation variables including the total edge contrast index (TECI), clumpiness index (CLUMPY), proximity index coefficient of variation (PROX_CV), connectance index (CONNECT), mean core area index (CAI_MN), core area index coefficient of variation (CAI_CV), core area index standard deviation (CAI_SD), and core area coefficient of variation (CORE_CV, McGarigal et al. 2012). We chose seven of these metrics based on analyses by Wang et al. (2014) that showed they have low correlation with the amount of habitat in a given landscape and the ability to differentiate between landscapes featuring a wide range of habitat fragmentation levels. We added the CONNECT metric because

it has relevance to habitat connectivity modeling (Wang et al. 2014). We tested the performance of two definitions of edge depth in calculating the core area metrics: a depth of one cell, corresponding to 27.35 meters, and a depth of seven cells, corresponding to 191.45 meters. These definitions assume that core habitat starts at either 27.35 meters away from an edge or 191.45 meters away from an edge, respectively. We have observed panda sign less than 50 meters from the edge of habitat patches in the study area (pers. obs.) which supports the former edge depth assumption, while the latter edge depth assumption incorporates the range of movement described by the second HMM step-length distribution (Fig. 4.3). We also used the 191.45-meter threshold to define separate habitat patches as joined or un-joined for calculation of the CONNECT metric, which sums the total joined patches and divides by the total possible number of joined patches to derive a proportion of connected patches. The CAI metrics measure the percent of habitat that is core habitat across habitat patches, while the CORE_CV metric is simply the coefficient of variation of the amount of core area in each patch. For the edge contrast metric (TECI), we defined the contrast between habitat and nonhabitat as the maximum possible (contrast = 1). Because we only had two classes in our landscape, TECI is a measure of the number of edges in each local window. Finally, the CLUMPY metric measures the number of like adjacencies observed between habitat cells in the landscape compared to the number that would be expected given a random distribution of the habitat cells. We used the moving window functionality in FRAGSTATS to calculate all metrics with the 85 x 85 cell window (McGarigal 2012). The amount of habitat in the window was calculated by summing the number of habitat cells in the same 85 x 85 cell window using the ‘focal’ function in the ‘raster’ R package (Hijmans 2019).

4.2.7 Landscape genetics analyses

In order to relate landscape variables to the observed pattern of genetic distances between individuals, they must first be transformed into a "resistance" surface (Spear et al. 2010). This surface represents the hypothesized effect of the landscape variable in question on gene flow. The relative support of a given variable and its transformation to a resistance surface must then be evaluated compared to other hypothesized variables and/or transformations using the observed genetic distances as the response variable to resistance distances between the genetic sample locations. We used the 'ResistanceGA' package to streamline this process and evaluate the effects of habitat amount and fragmentation on genetic distance (Peterman 2018). This package uses genetic algorithms and maximum likelihood population mixed effects (MLPE) models to test multiple transformations of input variables into resistance surfaces and evolve to better solutions based on how well the proposed resistance surface predicts genetic distance (McRae 2006, Peterman et al. 2014). In our case, we modeled movement using the program Circuitscape, which evaluates probabilities of movement at each cell between sample locations based on the resistance values of the surrounding cells (McRae 2006). By employing MLPE models, the effects of environmental resistance surfaces on movement can be separated from the random effects specified as the pairwise dependence of observations, an issue that has plagued correlative techniques like the Mantel test and derivations (Guillot and Rousset 2013, Peterman et al. 2014). Likely due to this flexibility, MLPE models were recently found to be the best method of model selection among those commonly used landscape genetics research (Shirk et al. 2018). ResistanceGA has also performed well in recovering the correct resistance surface compared to other landscape genetics methods in simulated environments (Peterman et al. 2019). We considered all eight possible resistance transformation equations available in the package,

which each take a ‘maximum’ and ‘shape’ parameter to define the transformation and can be varied to explore a wide variety of transformation curves (Peterman et al. 2018). The genetic algorithm selects randomly generates new transformations through the mutation and recombination of the ‘maximum’ and shape’ parameters in every generation, and we used the loglikelihood of the resulting MLPE model as the fitness function to optimize. In evaluating models, we defined the number of parameters included in each as three per surface – the surface itself, the maximum value of the transformation, and the shape of the transformation.

Specifically, we evaluated the individual habitat amount and eight habitat fragmentation metric surfaces individually using the ‘SS_optim’ function. We also evaluated both a model only including Euclidean distance between sample locations (an isolation-by-distance (IBD) model) and a completely Null model that assumes panmixia. Before optimization, we aggregated the cell size of these surfaces by a factor of eight to a size of 218.8 m using the mean value of the smaller cells due to the computational demand of the smaller 27.35 m cell size. This is unlikely to affect results due to previous research that has shown that Circuitscape and landscape genetics analyses are robust to increasing cell size (McRae 2006, Cushman and Landguth 2010). We ranked the performance of these single surface models according to their Akaike’s Information Criterion corrected for the number of sample individuals and parameters in each MLPE and considered any surface that had an AICc of less than four higher than the top surface as competitive (Akaike 1974, Burnham and Anderson 2004). We then built multi-surface resistance surfaces by transforming and adding together every possible combination of these competitive surfaces using the ‘MS_optim’ function in ResistanceGA. In this function, the relative weights of each surface’s contribution to the final resistance surface is also optimized. If any pair of the competitive surfaces had a Pearson’s correlation value higher than 0.5, we only considered the higher ranked

surface of the pair in the multi-surface model due to collinearity issues and the need to disentangle habitat amount vs. fragmentation effects (Dormann et al. 2013, Wang et al. 2014). For all resistance transformations we used the log-likelihood of the model's fit to the genetic distance data as the objective function to maximize using the genetic algorithm. We ranked all final single-surface and multi-surface variable combinations and transformations according to AICc. If there were more than one top-supported models with ΔAICc values of less than four, we calculated a weighted average of the resistance surfaces resulting from those models as a form of model-averaging. Finally, we used the Circuitscape algorithm to calculate the cumulative probabilities of movement across the landscape between all sample locations given the final resistance surface.

4.2.8 Optimizing the spatial scale of effect

ResistanceGA includes functionality for investigating the scale of effect of a given environmental variable on genetic distance through a Gaussian smoothing parameter. This parameter is applied to the surface before transformation and is also optimized through the genetic algorithm. In order to select the landscape window size for calculating the amount of habitat and our habitat fragmentation metrics, we first ran a single surface optimization on the aggregated (218.8 m grain) binary habitat map with the scaling parameter using the 'SS_optim.scale' function. The scaling parameter represents the standard deviation, in cells, surrounding where a focal cell is smoothed via a Gaussian function. This scaling parameter was optimized at 2.63 for our aggregated 218.8 m-grain habitat map. We multiplied this by two to incorporate 95% of the optimized scale of effect in one direction, which translated to forty-two 27.35 m-grain cells in the original habitat map. Considering the focal cell and 95% of the scale

of effect in each direction resulted in the final 85 x 85 matrix focal window used in the calculation of our habitat amount and fragmentation variables.

4.3 Results

4.3.1 Noninvasive genetics survey

142 unique pandas were genotyped across Wolong Nature Reserve. The probability of two different individuals sharing the same genotype (PID) across the seven loci was estimated at less than 3.9×10^{-6} and the same probability for identical siblings (PSIB) was estimated at 0.005. Microchecker found no evidence of allelic dropout at any locus, and FreeNA estimated an average of less than a 0.04 frequency of null alleles across the seven loci. Observed and expected heterozygosity ranged from 0.39-0.74 and 0.36-0.78 and averaged 0.60 and 0.63 across the seven loci, respectively. No loci were found to be significantly outside HWE at $p = 0.01$ using the Bonferroni correction. Full results for each locus are reported in Table 1 of Qiao et al. (2019).

4.3.2 Habitat models

The MaxEnt habitat models built with the unadjusted environmental variables and those which weighted the surrounding landscape based on each of the three HMM step-length distributions did not vary greatly in predictive accuracy (Table 4.2). Because the highest TSS (0.71) was achieved with the environmental variable set based on the second HMM step-length distribution, we selected that model and associated habitat predictions to use for the calculation of the habitat amount and fragmentation metrics (Fig. 4.3). This maximum TSS and an AUC of 0.90 when predicting to our fecal genetics sample locations indicated good accuracy in predicting the relevant panda distribution for our study. The binary habitat map and resulting habitat amount variable depict large swathes of habitat through the middle region of our study

area with breaks in habitat that follow the main road through Wolong and associated human settlement and alpine zones to the West (Fig. 4.1).

4.3.3 Landscape genetics models

Based on the ResistanceGA results, the CAI_SD, habitat amount, and CONNECT variables were the best predictors of inter-individual genetic distance (Table 4.3). Optimizing the transformation of multi-surface combinations of these top single surfaces into a resistance surface did not result in better-fitting MLPE models (Table 4.3). The optimized resistance surface based on a transformation of CAI_SD resulted in the top-ranked MLPE model with 0.57 of the AICc weight, while the next-most supported models based on the habitat amount and CONNECT surfaces had weights of 0.14 and 0.09, respectively (Table 4.3). Because these top models were within ΔAICc of four, we calculated a weighted-average resistance surface based on AICc weight. All models incorporating habitat amount and/or fragmentation were more highly ranked than the Null panmixia model, but several single surface and multi-surface models were outperformed by the distance model (Table 4.3). That said, the MLPE model of distance alone was ranked 10th and had an AICc weight of only 0.01. The optimal transformations of the top variables into resistance surfaces varied, with an Inverse Ricker equation applied to the CAI_SD and CONNECT variables and an Inverse-Reverse Ricker transformation applied to the habitat amount variable (Fig. 4.4). These optimized transformations followed our hypothesis in that increasing habitat amount and fragmentation at low values increased functional connectivity until a threshold, at which point further increases reduced functional connectivity. The Circuitscape conductance map representing cumulative probabilities of movement between all sample locations across the weighted average-resistance of the CAI_SD, habitat amount, and CONNECT surfaces is presented in Figure 4.5.

4.4 Discussion

Results suggest that levels of habitat fragmentation in the surrounding landscape is more important to functional connectivity in panda populations than the amount of habitat that is available. That said, the resistance surface based on habitat amount was the second-best predictor of genetic structure and thus should be considered in future analyses of giant panda population functional connectivity. The fact that a measure of habitat fragmentation with low correlation with habitat abundance outperformed predictive models of gene flow based on the amount of habitat in leads us to reject the "Habitat Amount" hypothesis as it relates to our study system and ecological response, however (Fahrig 2013). Briefly, Fahrig (2013) proposed that increasing habitat amount results in increasing species richness, and that this richness is not affected by patch size or isolation (fragmentation effects). Although we looked at a different ecological response (functional connectivity) in a single species, we found that habitat amount, though well-supported, was second to the standard deviation of the core area index across patches as a predictor of functional connectivity. It is important to consider the effects of habitat amount and fragmentation on functional connectivity for multiple species and how this in turn drives metacommunity structure and evolution (Leibold et al. 2004, Urban et al. 2011). Better understanding this relationship between habitat configuration and functional connectivity across multiple species could also be an important step for understanding the ecological process behind observed patterns in biodiversity and the role of habitat amount and fragmentation in driving them (Thompson et al. 2017).

Another important consideration is that we found the relationship between the amount of habitat and its level of fragmentation to not be linear. Specifically, while increasing habitat amount was initially associated with increased functional connectivity in our studied panda

population and area, this relationship reversed when there was a high proportion of habitat in the local landscape. This threshold that maximized functional connectivity occurred at about 5800 raster cells (4.34 km^2) in the surrounding 85 x 85-cell landscape (7225 cells, 5.40 km^2). This means that increasing the amount of habitat in the surrounding landscape up to 80% resulted in increased functional connectivity, but that further increases in the amount of habitat rapidly decreased connectivity (Fig. 4.4). Jackson and Fahrig (2015) also found nonlinear effects of habitat amount on genetic structure in simulated environments, but in their case differences in genetic responses to habitat amount were measured at the landscape level across different simulated landscape replicates. Ours is the first study, to our knowledge, to demonstrate nonlinear effects of habitat amount on functional connectivity in an empirical system and across a single landscape. We posit that high abundances of habitat in the surrounding local area result in more resource availability and in turn lower dispersal rates and smaller dispersal distances, driving the negative relationship between habitat amount and functional connectivity that we observed at habitat amounts of 80 % or more.

It is interesting to consider our findings in relation to animal dispersal theory and source-sink dynamics (Clobert et al. 2012). At high population densities and growth rates, areas of high resource availability may serve as source populations from which individuals disperse across the landscape (Draheim et al. 2016). Either panda densities in our landscape are not high enough to drive this behavior, or the number of non-dispersing individuals is high enough to mask the genetic signal of individuals dispersing from the high-resource source areas. In either case, it may make more conceptual sense to think of the local landscapes containing more than 80% habitat as exhibiting an “attractancy” to individuals compared to a “resistance” to their movement (Clobert et al. 2012). As opposed to total population density, effective population

density may be a more important predictor of genetic structure in certain circumstances, particularly for small subpopulations that are affected by lower dispersal rates at low densities and experience high rates of genetic drift (Weckworth et al. 2013). Effective population density can also be used, along with estimates of demographic rates, to calculate Wright's neighborhood size for continuously distributed populations (Waples et al. 2018). Spatially explicit estimates of both total and effective population densities in the relevant surrounding landscape would thus be a valuable measure to account for in analyses of functional connectivity (Poethke and Hovestadt 2002, Efford 2011, Pfluger and Balkenhol 2014). The interaction between population density and habitat amount/fragmentation is also an important consideration (Lutscher 2008, Cushman et al. 2010). For example, we posit that high local densities would weaken our finding that high amounts of habitat in the local landscape reduce gene flow. This is because individuals would be forced to disperse due to resource competition arising not from limited habitat but instead high densities (Draheim et al. 2016).

In addition to habitat amount, the transformation of the fragmentation metric that resulted in the top-performing MLPE model (CAI_SD), into a resistance surface was also nonlinear. Our hypothesized threshold did manifest, but it occurred at a very low level of habitat heterogeneity. Specifically, local landscape windows with no variation in the relative amount of core area between patches resulted in low connectivity, but this connectivity rapidly increased until it was maximized at a standard deviation in the core area index of about 2.5. As an example, if a local landscape contained two habitat patches containing 60 % and 55 % core area, this local landscape would have a standard deviation of 2.5 and the highest possible functional connectivity predicted by our model. Further increases in heterogeneity rapidly decreased functional connectivity until about CAI_SD = 15, which represents large variation in the

proportion of core area between habitat patches. The transformation of the other top surface, the connectance index (CONNECT) metric, had a similar optimized transformation in which maximum functional connectivity was achieved when only about 12 % of habitat patches in a local landscape were connected. The CAI_SD transformation indicates that functional connectivity in our studied giant panda population is maximized at relatively low levels of habitat patch heterogeneity across the landscape while the CONNECT transformation suggests that this functional connectivity is maximized at surprisingly low levels of habitat patch connectedness.

Regarding the relative influence of habitat amount and habitat fragmentation on functional connectivity, our results support the conclusions made by Cushman et al. (2012) that habitat configuration is more important to genetic structure than habitat amount through repeated simulations. We argue that the opposite conclusion reached by Jackson and Fahrig's (2015) simulation study is likely due to their analysis of only one configuration metric – CLUMPY, which we also found to be a poor predictor of genetic structure (Table 4.3). They chose this metric because it was designed to be independent of habitat amount, but other metrics of habitat configuration also feature low correlations with habitat amount (Wang et al. 2014). The selection and analysis of a range of these metrics should be done to explore a wide range of potential habitat configuration effects.

Other studies have used fragmentation metrics to investigate genetic structure. For example, Bull et al. (2011) used correlation length in forested land cover, an estimate of contiguity, to predict gene flow in black bear populations across several landscape replicates and found it to be a significant predictor in those landscapes which featured lower forest cover. Howell et al. (2016) also used correlation length of forests, which they found positively predicted

gene flow in reintroduced American marten populations. Draheim et al. (2018) used several fragmentation metrics to describe landscape change and predict changes to spatial genetic structure in black bear populations. This latter approach, if combined with explicit estimations of habitat like done here, may be especially important to determine the effects of ongoing habitat loss and fragmentation on populations over time.

4.4.1 Conservation implications

Maintaining functional connectivity within and among populations is a key component in the survival of many species through the preservation of genetic diversity, the spread of adaptive genes, and the recolonization of areas after localized extinctions (Hanski 1998, Bay et al. 2017). There is an ongoing debate in the literature about the importance or not of habitat fragmentation to biodiversity and other ecological responses relative to and after accounting for habitat amount (Fahrig 2017, Haddad et al. 2017, Fletcher et al. 2018, Fahrig et al. 2019). Others have argued that the debate is largely irrelevant to conservation – that habitat amount should always be maximized, and habitat fragmentation should be managed to indirectly conserve populations through responses like functional connectivity (Villard and Metzger 2014). Our results warrant further consideration of this assertion – in our studied landscape and species, both habitat amount and habitat fragmentation may be optimized at intermediate levels to maximize functional connectivity. Specifically, the threshold at 80% habitat in the surrounding landscape that we found suggests that there is an optimum to the amount of habitat that should be restored or preserved in a given area. The configuration of this habitat may be even more important, and in our study system, patches that have relatively similar core area ratios (low standard deviation between them) and are only moderately connected should be maintained to maximize the functional connectivity response.

With most conservation projects limited in funding, our findings that there are thresholds in the response of functional connectivity to both habitat amount and fragmentation have significant implications for the spatial planning of habitat protection and restoration efforts. This is especially the case for threatened species facing subpopulation fragmentation like the giant panda (Xu et al. 2017). Although our study has been purported to contain two discrete subpopulations, recent research has found relatively high rates of contemporary gene flow between them and the pandas in Wolong likely exist in a relatively continuous population (Forestry Department of Sichuan Province 2015, Qiao et al. 2019). Our results are thus more relevant to within-population functional connectivity but could be tentatively extrapolated to increase connectivity between isolated subpopulations. This may be especially important for pandas due to their relatively low fecundity rate, meaning that migration may play an important role in both demographic and genetic connectivity (Lowe and Allendorf 2010). Further research should be conducted in other areas of the panda range to confirm our observed relationships between habitat amount/fragmentation and functional connectivity in the species or describe how they vary based on different conditions. Finally, sex has been shown to be an important variable for gene flow in giant panda populations in that they feature female-biased natal dispersal (Hu et al. 2010). Thus, future work should include sex information to determine if there are important differences in environmental effects on gene flow between males and females.

The methods employed in our study could be applied to any species/population and landscape in which habitat and estimates of successful dispersal can be effectively measured. Our analyses allowed for a detailed investigation into the effects of habitat amount and habitat fragmentation on functional connectivity and are likely especially important to consider for habitat specialist species. We posit that these methods should be applied to species with a wide

range of traits, however, to better understand the relationships between habitat amount, habitat fragmentation and functional connectivity. In an era when wildlife populations face increasing anthropogenic pressure across much of the globe, detailed investigations into functional connectivity, such as detailed here, should be used to optimize the use of limited conservation resources for the persistence of species (Dirzo et al. 2014).

APPENDIX

Table 4.1. Environmental variables used in habitat modeling

Variable name	Source
Elevation (m)	NASA's SRTM mission (NASA 2013)
Terrain ruggedness	Derived from elevation (see text)
Aspect (degrees)	Derived from elevation (see text)
Tree cover (%)	Hansen et al. (2013)
Distance to river (m)	National Geomatics Center of China (2018)
Distance to road (m)	National Geomatics Center of China (2018)

Table 4.2. Predictive accuracy statistics of MaxEnt models trained with either the raw environmental variables or those that incorporated the surrounding local landscape based on weights derived from step-length distributions derived from Hidden Markov Model (HMM) movement states. TSS is the true skill statistic, AUC is the area under the receiver operating curve, and Cor is the correlation between predicted suitability values between test presences and test background locations. The step-length distribution used to weight the environmental variables in the top-performing model (HMM movement state 2) is shown in Figure 4.3.

Variable set	TSS	AUC	Cor
HMM movement state 2	0.71	0.90	0.15
HMM movement state 1	0.70	0.90	0.15
HMM movement state 3	0.70	0.91	0.15
Raw	0.70	0.91	0.15

Table 4.3. MLPE results from ResistanceGA, ranked by increasing AICc. CAI = Core Area Index, CORE = Core area, CONNECT = Connectance Index, CLUMPY = Clumpiness index, MN = Mean, CV = Coefficient of Variation, SD = Standard Deviation. * denotes metric calculated with edge depth = 191.45 m, while † denotes metric calculated with edge depth = 27.35 m.

Transformed surfaces in model	Parameters	AICc	Log-likelihood	ΔAICc	AICc Weight
CAI_SD*	4	44206.18	-22098.94	0.00	0.57
Habitat amount	4	44208.91	-22100.31	2.74	0.14
CONNECT*	4	44209.78	-22100.75	3.61	0.09
CAI_SD†	4	44211.83	-22101.77	5.66	0.03
CAI_CV†	4	44211.89	-22101.80	5.72	0.03
CAI_mn†	4	44212.65	-22102.18	6.48	0.02
TECI	4	44212.85	-22102.28	6.68	0.02
CAI_SD* CONNECT*	7	44212.95	-22099.06	6.77	0.02
CAI_SD* Habitat amount	7	44213.12	-22099.14	6.95	0.02
Distance	2	44213.96	-22104.94	7.79	0.01
CORE_CV†	4	44214.77	-22103.24	8.60	0.01
CAI _CV†	4	44214.90	-22103.30	8.72	0.01
CONNECT* Habitat amount	7	44215.03	-22100.10	8.85	0.01
CLUMPY	4	44215.06	-22103.39	8.89	0.01
CAI_mn*	4	44215.55	-22103.63	9.37	0.01
CAI_CV*	4	44216.08	-22103.89	9.90	0.00
CORE_CV*	4	44216.16	-22103.94	9.99	0.00
CAI_SD* CONNECT* Habitat amount	10	44221.87	-22100.10	15.70	0.00

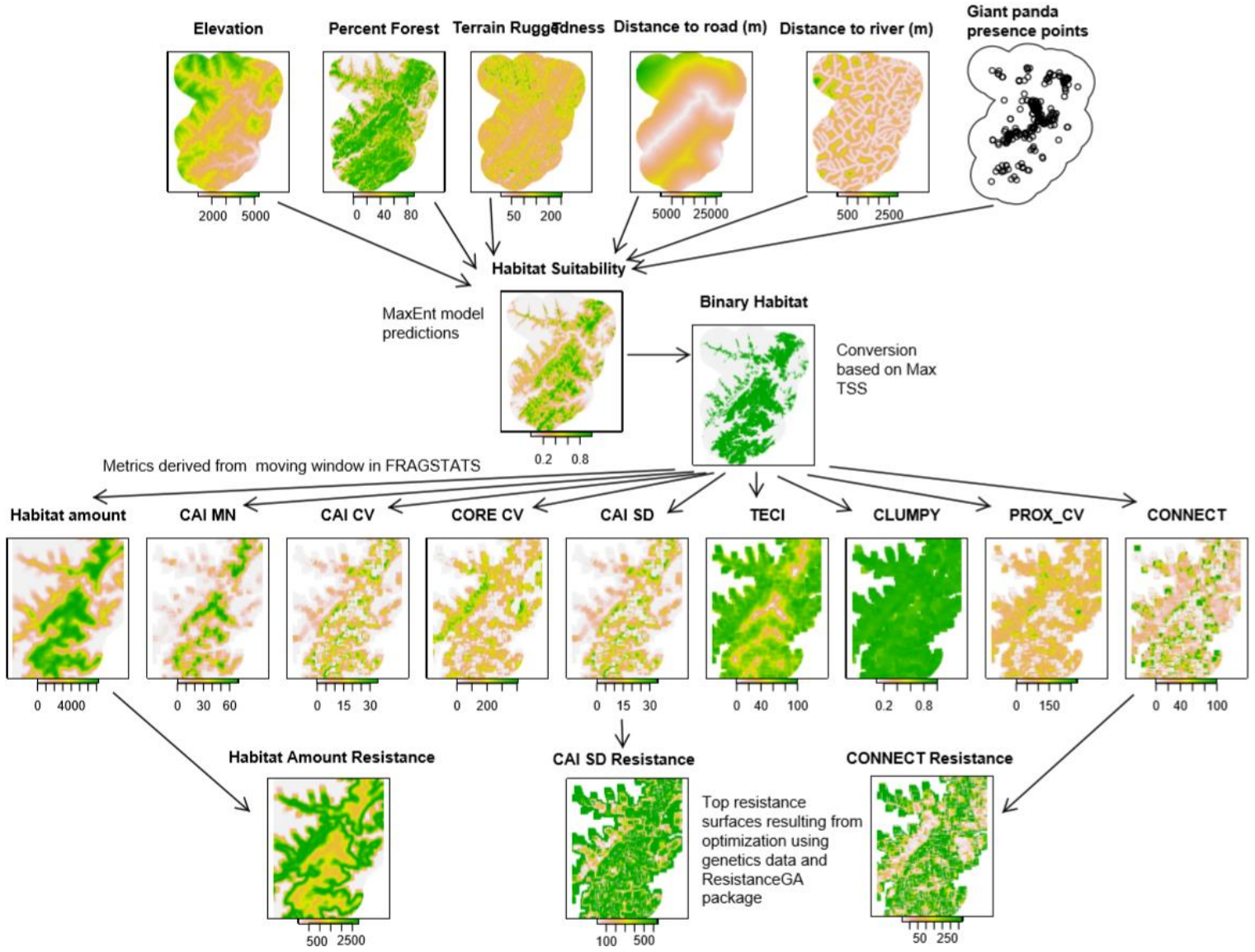

Figure 4.1. Flowchart of the process and intermediary outputs in the development of a map of landscape resistance across the study area, starting with raw environmental variables and giant panda presence dataset. For more details on each step, see the methods.

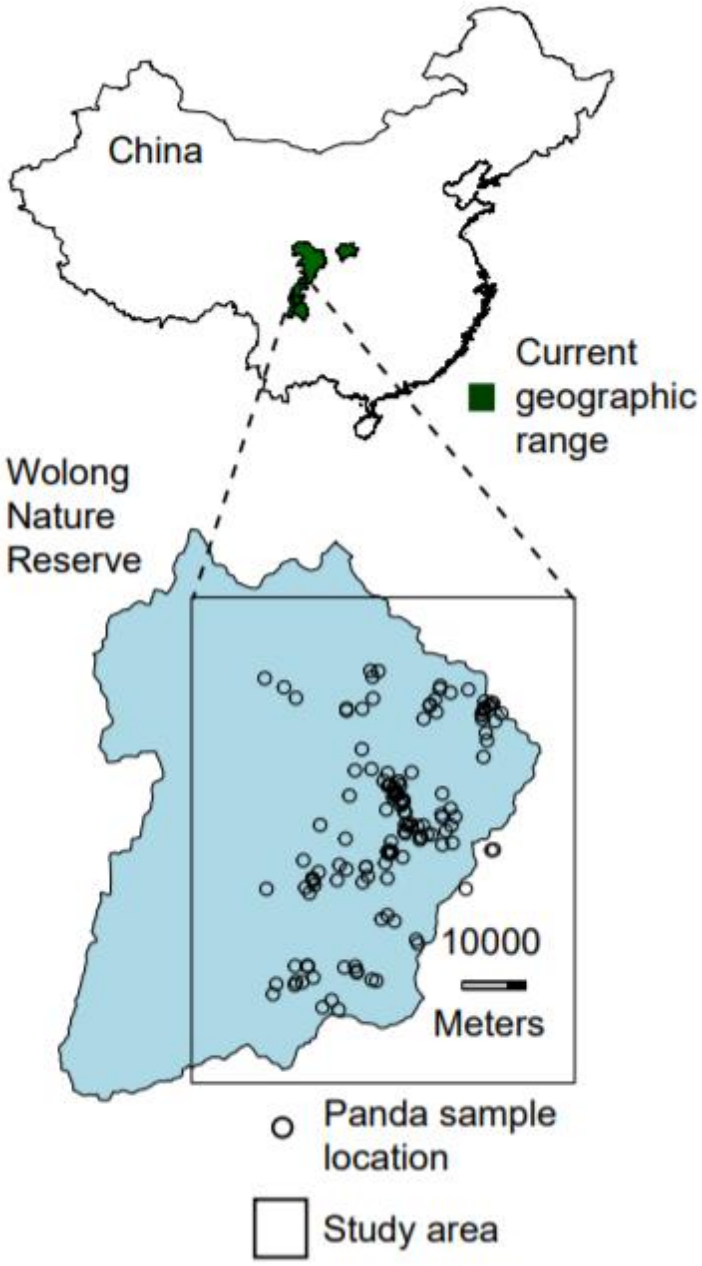

Figure 4.2. Study area and sample locations, inset on map of China and the current giant panda geographic range.

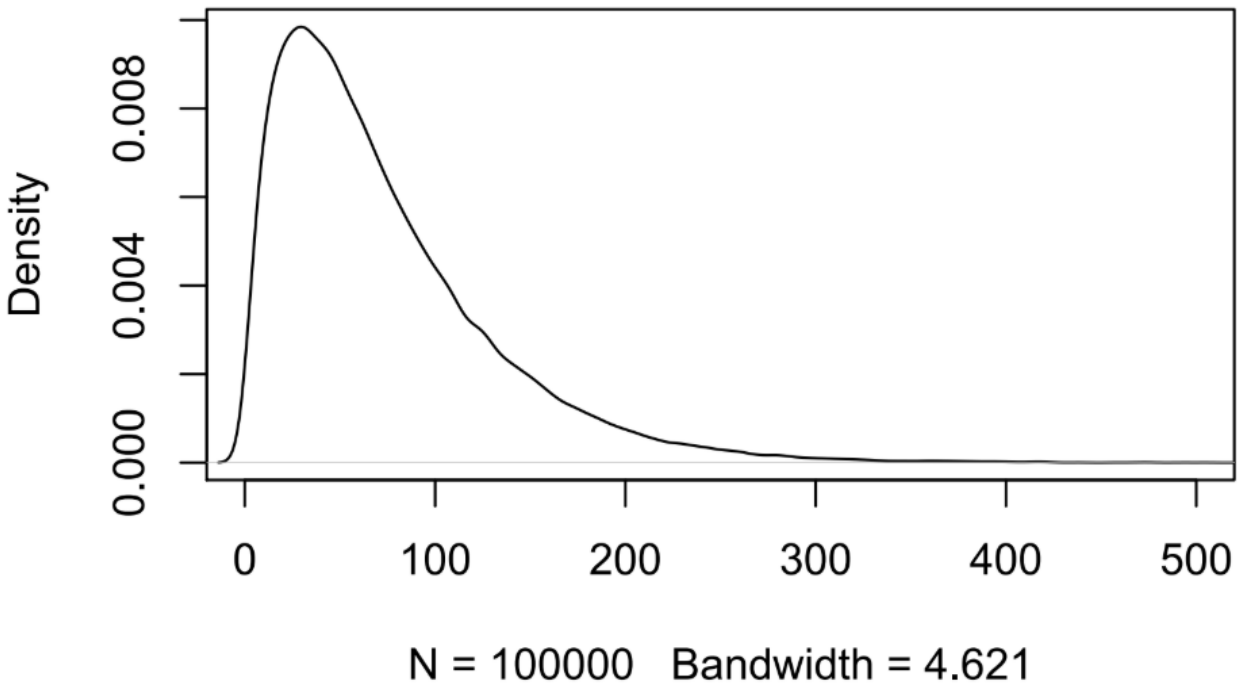

Figure 4.3. Kernel density of panda movement steps at t = 3-hour intervals. This plot represents the second movement "state" that resulted in transformations of the environmental predictor variables that best predicted habitat. This distribution was derived from GPS collar data from n=5 pandas and includes steps from all hours due to the fact that giant pandas have multiple activity peaks throughout of the day and night (Zhang et al. 2016).

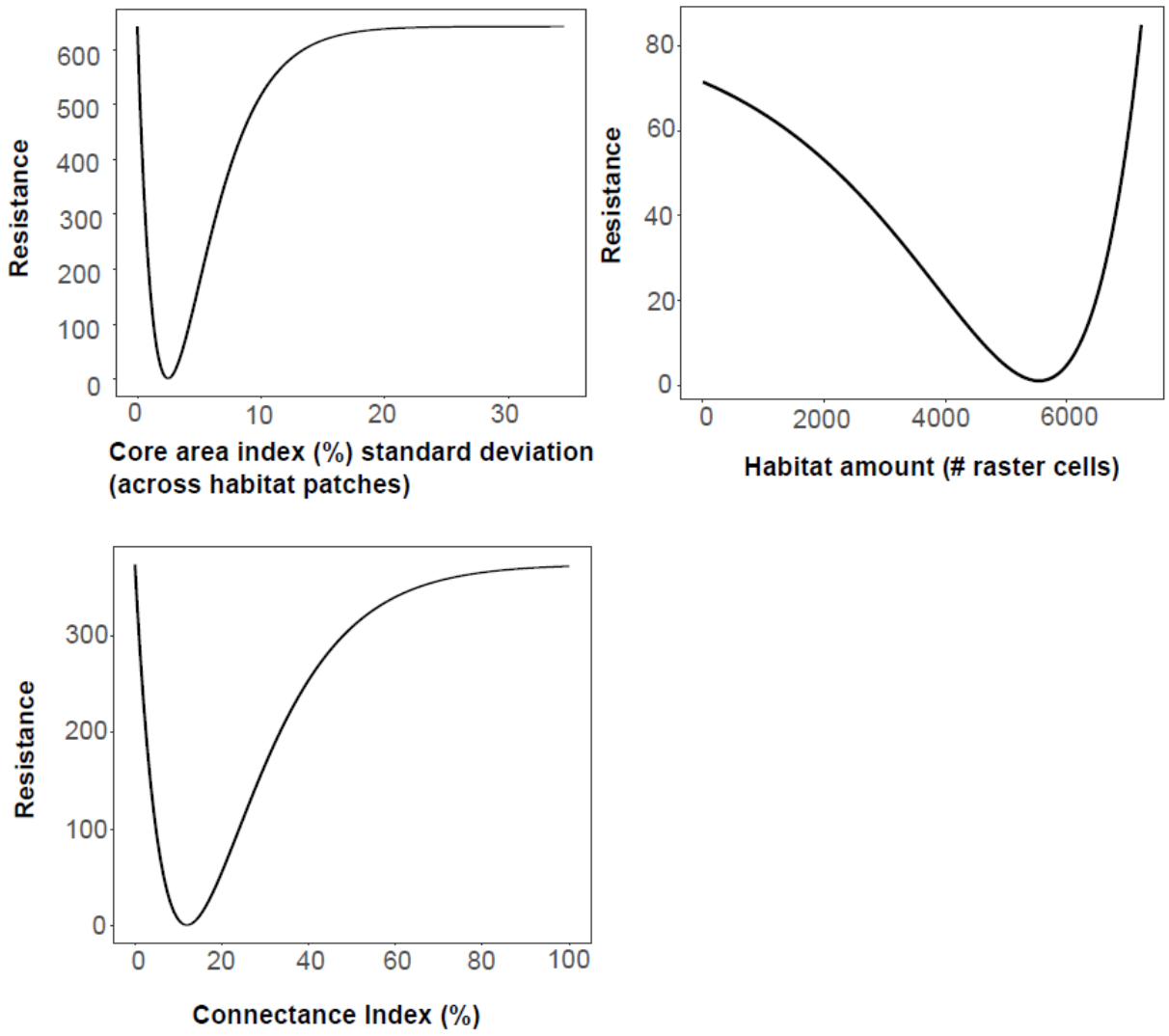

Figure 4.4. Transformations of the core area index standard deviation across habitat patches (CAI_SD), habitat amount and CONNECT metrics optimized through ResistanceGA. The resistance surface resulting from CAI_SD resulted in the model with the best fit to the genetic data and thus was the best predictor of functional connectivity, followed by habitat amount and CONNECT.

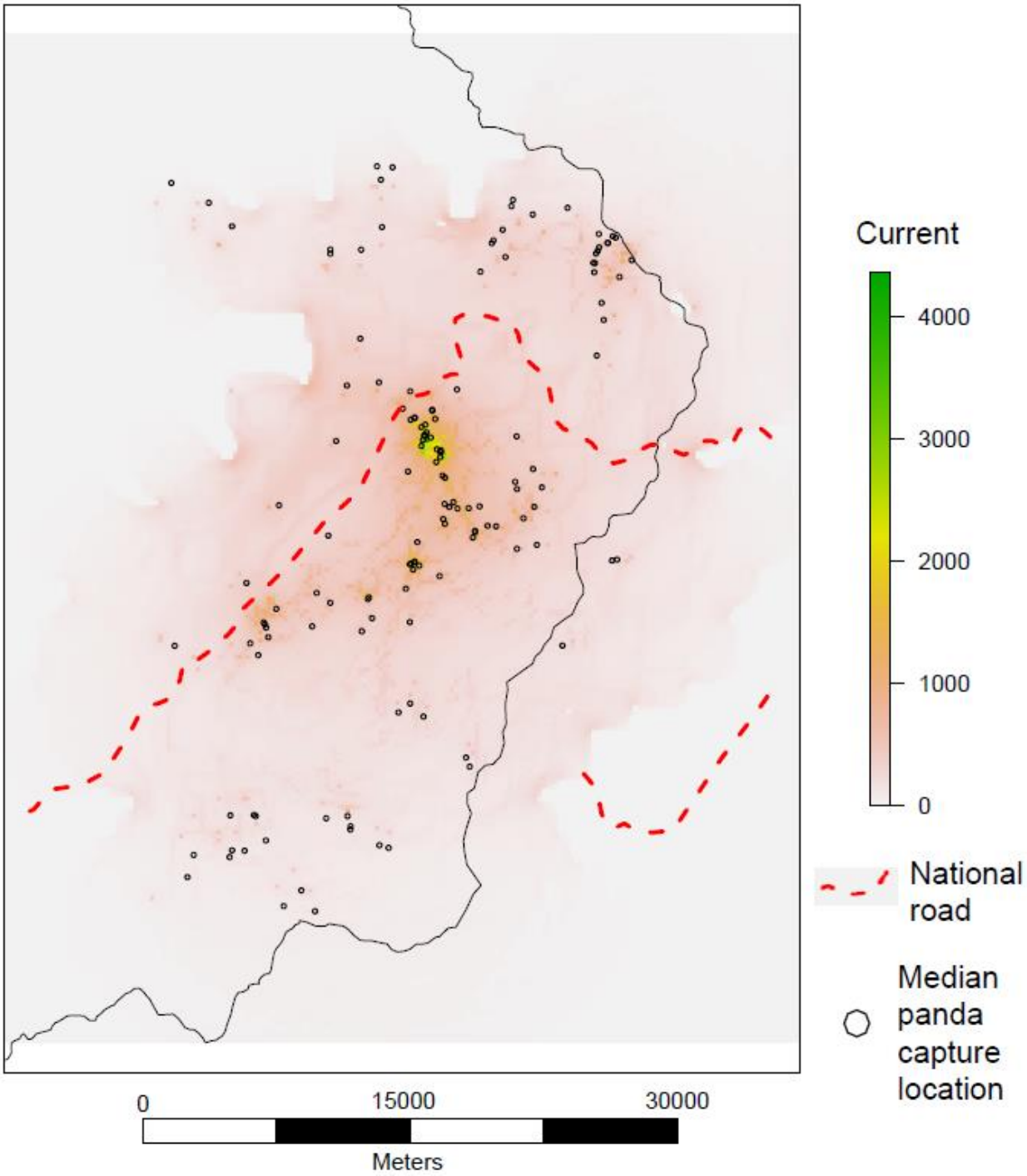

Figure 4.5. Conductance map derived from a weighted average (based on AICc weight) of the top resistance surfaces resulting from the transformation of the standard deviation in the core area index across patches (CAI_SD), habitat amount, and the connectance index (CONNECT). The Circuitscape algorithm was used to calculate the conductance map, which represents the cumulative probabilities of movements between all panda sample locations. Points represent the median locations of sampled individuals that were used in the landscape genetics analysis.

REFERENCES

REFERENCES

Akaike, H. 1974. STOCHASTIC THEORY OF MINIMAL REALIZATION. Ieee Transactions on Automatic Control **AC19**:667-674.

Allouche, O., A. Tsoar, and R. Kadmon. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology **43**:1223-1232.

Bay, R. A., N. Rose, R. Barrett, L. Bernatchez, C. K. Ghalambor, J. R. Lasky, R. B. Brem, S. R. Palumbi, and P. Ralph. 2017. Predicting Responses to Contemporary Environmental Change Using Evolutionary Response Architectures. American Naturalist **189**:463-473.

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.

Bull, R. A. S., S. A. Cushman, R. Mace, T. Chilton, K. C. Kendall, E. L. Landguth, M. K. Schwartz, K. McKelvey, F. W. Allendorf, and G. Luikart. 2011. Why replication is important in landscape genetics: American black bear in the Rocky Mountains. Molecular Ecology **20**:1092-1107.

Burnham, K. P., and D. R. Anderson. 2004. Multimodel inference - understanding AIC and BIC in model selection. Sociological Methods & Research **33**:261-304.

Chapuis, M. P., and A. Estoup. 2007. Microsatellite null alleles and estimation of population differentiation. Molecular Biology and Evolution **24**:621-631.

Clobert, J., Baguette, M., Benton, T.G. and Bullock, J.M. eds., 2012. Dispersal ecology and evolution. Oxford University Press.

Connor, T., V. Hull, and J. G. Liu. 2016. Telemetry research on elusive wildlife: A synthesis of studies on giant pandas. Integrative Zoology **11**:295-307.

Cushman, S. A., B. W. Compton, and K. McGarigal. 2010. Habitat Fragmentation Effects Depend on Complex Interactions Between Population Size and Dispersal Ability: Modeling Influences of Roads, Agriculture and Residential Development Across a Range of Life-History Characteristics. Spatial Complexity, Informatics, and Wildlife Conservation:369-385.

Cushman, S. A., and E. L. Landguth. 2010. Scale dependent inference in landscape genetics. Landscape Ecology **25**:967-979.

Cushman, S. A., A. Shirk, and E. L. Landguth. 2012. Separating the effects of habitat area, fragmentation and matrix resistance on genetic differentiation in complex landscapes. Landscape Ecology **27**:369-380.

Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carre, J. R. G. Marquez, B. Gruber, B. Lafourcade, P. J. Leitao, T. Munkemuller, C. McClean, P. E. Osborne, B. Reineking, B. Schroder, A. K. Skidmore, D. Zurell, and S. Lautenbach. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography **36**:27-46.

Draheim, H. M., J. A. Moore, D. Etter, S. R. Winterstein, and K. T. Scribner. 2016. Detecting black bear source–sink dynamics using individual-based genetic graphs. Proceedings of the Royal Society B: Biological Sciences **283**:1835.

Draheim, H. M., J. A. Moore, M. J. Fortin, and K. T. Scribner. 2018. Beyond the snapshot: Landscape genetic analysis of time series data reveal responses of American black bears to landscape change. Evolutionary Applications **11**:1219-1230.

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.

Efford, M. G. 2011. Estimation of population density by spatially explicit capture-recapture analysis of data from area searches. Ecology **92**:2202-2207.

Elith, J., and C. H. Graham. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography **32**:66-77.

Elith, J., S. J. Phillips, T. Hastie, M. Dudik, Y. E. Chee, and C. J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions **17**:43-57.

Fahrig, L. 2013. Rethinking patch size and isolation effects: the habitat amount hypothesis. Journal of Biogeography **40**:1649-1663.

Fahrig, L. 2017. Ecological Responses to Habitat Fragmentation Per Se. Pages 1-23 *in* D. J. Futuyma, editor. Annual Review of Ecology, Evolution, and Systematics, Vol 48. Annual Reviews, Palo Alto.

Fahrig, L., V. Arroyo-Rodriguez, J. R. Bennett, V. Boucher-Lalonde, E. Cazetta, D. J. Currie, F. Eigenbrod, A. T. Ford, S. P. Harrison, J. A. G. Jaeger, N. Koper, A. E. Martin, J. L. Martin, J. P. Metzger, P. Morrison, J. R. Rhodes, D. A. Saunders, D. Simberloff, A. C. Smith, L. Tischendorf, M. Vellend, and J. I. Watling. 2019. Is habitat fragmentation bad for biodiversity? Biological Conservation **230**:179-186.

Fletcher, R. J., R. K. Didham, C. Banks-Leite, J. Barlow, R. M. Ewers, J. Rosindell, R. D. Holt, A. Gonzalez, R. Pardini, E. I. Damschen, F. P. L. Melo, L. Ries, J. A. Prevedello, T.

Tscharntke, W. F. Laurance, T. Lovejoy, and N. M. Haddad. 2018. Is habitat fragmentation good for biodiversity? Biological Conservation **226**:9-15.

Forestry Department of Sichuan Province (2015). The pandas of Sichuan. Sichuan Science and Technology Press.

Frankham, R. 1995. INBREEDING AND EXTINCTION - A THRESHOLD EFFECT. Conservation Biology **9**:792-799.

Guillot, G., and F. Rousset. 2013. Dismantling the Mantel tests. Methods in Ecology and Evolution **4**:336-344.

Haddad, N. M., L. A. Brudvig, J. Clobert, K. F. Davies, A. Gonzalez, R. D. Holt, T. E. Lovejoy, J. O. Sexton, M. P. Austin, C. D. Collins, W. M. Cook, E. I. Damschen, R. M. Ewers, B. L. Foster, C. N. Jenkins, A. J. King, W. F. Laurance, D. J. Levey, C. R. Margules, B. A. Melbourne, A. O. Nicholls, J. L. Orrock, D. X. Song, and J. R. Townshend. 2015. Habitat fragmentation and its lasting impact on Earth's ecosystems. Science Advances **1**:9.

Haddad, N. M., A. Gonzalez, L. A. Brudvig, M. A. Burt, D. J. Levey, and E. I. Damschen. 2017. Experimental evidence does not support the Habitat Amount Hypothesis. Ecography **40**:48-55.

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.

Hanski, I. 1998. Metapopulation dynamics. Nature 396:41-49.

Hijmans, R. J., S. Phillips, J. Leathwick, and Ja Elith. 2017. dismo: Species Distribution Modeling. R package version 1.1-4. https://CRAN.R-project.org/package=dismo

Hijmans, R. J. 2019. raster: Geographic Data Analysis and Modeling. R package version 3.0-7. https://CRAN.R-project.org/package=raster

Hesselbarth, M. H. K., M. Sciaini, K. A. With, K. Wiegand, and J. Nowosad, 2019. landscapemetrics: an open-source R tool to calculate landscape metrics. Ecography **42**: 1648-1657.

Howell, P. E., E. L. Koen, B. W. Williams, G. J. Roloff, and K. T. Scribner. 2016. Contiguity of landscape features pose barriers to gene flow among American marten (Martes americana) genetic clusters in the Upper Peninsula of Michigan. Landscape Ecology **31**:1051-1062.

Hu, Y. B., X. J. Zhan, D. W. Qi, and F. W. Wei. 2010. Spatial genetic structure and dispersal of giant pandas on a mountain-range scale. Conservation Genetics **11**:2145-2155.

Huang, J., Y. Z. Li, L. M. Du, B. Yang, F. J. Shen, H. M. Zhang, Z. H. Zhang, X. Y. Zhang, and B. S. Yue. 2015. Genome-wide survey and analysis of microsatellites in giant panda (Ailuropoda melanoleuca), with a focus on the applications of a novel microsatellite marker system. Bmc Genomics **16**.

Hull, V., G. Roloff, J. D. Zhang, W. Liu, S. Q. Zhou, J. Y. Huang, W. H. Xu, Z. Y. Ouyang, H. M. Zhang, and J. G. Liu. 2014. A synthesis of giant panda habitat selection. Ursus **25**:148-162.

Hull, V., J. D. Zhang, S. Q. Zhou, J. Y. Huang, R. G. Li, D. Liu, W. H. Xu, Y. Huang, Z. Y. Ouyang, H. M. Zhang, and J. G. Liu. 2015. Space use by endangered giant pandas. Journal of Mammalogy **96**:230-236.

Jackson, N. D., and L. Fahrig. 2016. Habitat amount, not habitat configuration, best predicts population genetic structure in fragmented landscapes. Landscape Ecology **31**:951-968.

Keyghobadi, N. 2007. The genetic implications of habitat fragmentation for animals. Canadian Journal of Zoology **85**:1049-1064.

Krauss, J., R. Bommarco, M. Guardiola, R. K. Heikkinen, A. Helm, M. Kuussaari, R. Lindborg, E. Ockinger, M. Partel, J. Pino, J. Poyry, K. M. Raatikainen, A. Sang, C. Stefanescu, T. Teder, M. Zobel, and I. Steffan-Dewenter. 2010. Habitat fragmentation causes immediate and time-delayed biodiversity loss at different trophic levels. Ecology Letters **13**:597-605.

Leibold, M. A., M. Holyoak, N. Mouquet, P. Amarasekare, J. M. Chase, M. F. Hoopes, R. D. Holt, J. B. Shurin, R. Law, D. Tilman, M. Loreau, and A. Gonzalez. 2004. The metacommunity concept: a framework for multi-scale community ecology. Ecology Letters 7:601-613.

Lowe, W.H. and F. W. Allendorf, 2010. What can genetics tell us about population connectivity? Molecular ecology. **19**:3038-3051.

Lutscher, F. 2008. Density-dependent dispersal in integrodifference equations. Journal of Mathematical Biology **56**:499-524.

Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends in Ecology & Evolution **18**:189-197.

Manel, S., and R. Holderegger. 2013. Ten years of landscape genetics. Trends in Ecology & Evolution **28**:614-621.

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. Available at the following web site: http://www.umass.edu/landeco/research/fragstats/fragstats.html

McRae, B. H. 2006. Isolation by resistance. Evolution **60**:1551-1561.

Michelot, T., R. Langrock, and T. A. Patterson. 2016. moveHMM: an R package for the statistical modelling of animal movement data using hidden Markov models. Methods in Ecology and Evolution **7**:1308-1315.

Mills, L. S., J. J. Citta, K. P. Lair, M. K. Schwartz, and D. A. Tallmon. 2000. Estimating animal abundance using noninvasive DNA sampling: Promise and pitfalls. Ecological Applications **10**:283-294.

National Geomatics Center of China. 2018. National Fundamental Geographic Information Database.

Nie, Y. G., J. R. Speakman, Q. Wu, C. L. Zhang, Y. B. Hu, M. H. Xia, L. Yan, C. Hambly, L. Wang, W. Wei, J. G. Zhang, and F. W. Wei. 2015. Exceptionally low daily energy expenditure in the bamboo-eating giant panda. Science **349**:171-174.

Peakall, R., and P. E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes **6**:288-295.

Peakall, R., and P. E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics **28**:2537-2539.

Peterman, W. E. 2018. ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. Methods in Ecology and Evolution **9**:1638-1647.

Peterman, W. E., G. M. Connette, R. D. Semlitsch, and L. S. Eggert. 2014. Ecological resistance surfaces predict fine-scale genetic differentiation in a terrestrial woodland salamander. Molecular Ecology **23**:2402-2413.

Pfluger, F. J., and N. Balkenhol. 2014. A plea for simultaneously considering matrix quality and local environmental conditions when analysing landscape impacts on effective dispersal. Molecular Ecology **23**:2146-2156.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling **190**:231-259.

Poethke, H. J., and T. Hovestadt. 2002. Evolution of density-and patch-size-dependent dispersal rates. Proceedings of the Royal Society B-Biological Sciences **269**:637-645.

Qiao, M. J., T. Connor, X. G. Shi, J. Huang, Y. Huang, H. M. Zhang, and J. H. Ran. 2019. Population genetics reveals high connectivity of giant panda populations across human disturbance features in key nature reserve. Ecology and Evolution **9**:1809-1819.

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/.

Sacks, B. N., J. L. Brazeal, and J. C. Lewis. 2016. Landscape genetics of the nonnative red fox of California. Ecology and Evolution **6**:4775-4791.

Schaller, G. B. 1985. Giant pandas of Wolong. University of Chicago press.

Shirk, A. J., E. L. Landguth, and S. A. Cushman. 2018. A comparison of regression methods for model selection in individual-based landscape genetic analysis. Molecular Ecology Resources **18**:55-67.

Smith, A. C., N. Koper, C. M. Francis, and L. Fahrig. 2009. Confronting collinearity: comparing methods for disentangling the effects of habitat loss and fragmentation. Landscape Ecology **24**:1271-1285.

Smouse, P. E., and R. Peakall. 1999. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. Heredity **82**:561-573.

Spear, S. F., N. Balkenhol, M. J. Fortin, B. H. McRae, and K. Scribner. 2010. Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. Molecular Ecology **19**:3576-3591.

State Council Information Office. 2015. The State Forestry Administration held the press conference for the fourth national giant panda survey results (In Chinese).

Taberlet, P., S. Griffin, B. Goossens, S. Questiau, V. Manceau, N. Escaravage, L. P. Waits, and J. Bouvet. 1996. Reliable genotyping of samples with very low DNA quantities using PCR. Nucleic Acids Research **24**:3189-3194.

Thompson, P. L., B. Rayfield, and A. Gonzalez. 2017. Loss of habitat and connectivity erodes species diversity, ecosystem functioning, and stability in metacommunity networks. Ecography **40**:98-108.

USGS Shuttle Radar Topography Mission. 2000. 1 Arc Second scene. Global Land Cover Facility, University of Maryland. College Park, Maryland.

Urban, M. C., L. De Meester, M. Vellend, R. Stoks, and J. Vanoverbeke. 2012. A crucial step toward realism: responses to climate change from an evolving metacommunity perspective. Evolutionary Applications **5**:154-167.

Van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes **4**:535-538.

Villard, M. A., and J. P. Metzger. 2014. Beyond the fragmentation debate: a conceptual model to predict when habitat configuration really matters. Journal of Applied Ecology **51**:309-318.

Waits, L. P., G. Luikart, and P. Taberlet. 2001. Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Molecular Ecology **10**:249-256.

Wang, X. L., F. G. Blanchet, and N. Koper. 2014. Measuring habitat fragmentation: An evaluation of landscape pattern metrics. Methods in Ecology and Evolution **5**:634-646.

Waples, R. S., K. T. Scribner, J. A. Moore, H. M. Draheim, D. Etter, and M. Boersen. 2018. Accounting for Age Structure and Spatial Structure in Eco-Evolutionary Analyses of a Large, Mobile Vertebrate. Journal of Heredity 109:709-723.

Weckworth, B. V., M. Musiani, N. J. DeCesare, A. D. McDevitt, M. Hebblewhite, and S. Mariani. 2013. Preferred habitat and effective population size drive landscape genetic patterns in an endangered species. Proceedings of the Royal Society B-Biological Sciences **280**.

Xu, W. H., A. Vina, L. Q. Kong, S. L. Pimm, J. J. Zhang, W. Yang, Y. Xiao, L. Zhang, X. D. Chen, J. G. Liu, and Z. Y. Ouyang. 2017. Reassessing the conservation status of the giant panda using remote sensing. Nature Ecology & Evolution **1**:1635-1638.

Zeller, K. A., K. McGarigal, P. Beier, S. A. Cushman, T. W. Vickers, and W. M. Boyce. 2014. Sensitivity of landscape resistance estimates based on point selection functions to scale and behavioral state: pumas as a case study. Landscape Ecology **29**:541-557

Zeng, Y., J. Zhang, V. and Hull. 2019. A mixed-method study on medicinal herb collection in relation to wildlife conservation: the case of giant pandas in China. Integrative zoology **00**: x–x

Zhang, J. D., V. Hull, J. Y. Huang, S. Q. Zhou, W. H. Xu, H. B. Yang, W. J. McConnell, R. G. Li, D. A. Liu, Y. Huang, Z. Y. Ouyang, H. M. Zhang, and J. G. Liu. 2015. Activity patterns of the giant panda (Ailuropoda melanoleuca). Journal of Mammalogy **96**:1116-1127.

CHAPTER 5: SOCIAL NETWORK ANALYSIS UNCOVERS HIDDEN SOCIAL COMPLEXITY IN 'SOLITARY' SPECIES

In collaboration with

Ken Frank, Jindong Zhang, Kim Scribner, Houjin Wan, Abbey Wilson, Jianguo Liu

Abstract

Sociality is an important part of many species' ecology. Analyses of animal social networks has traditionally been conducted on species that exhibit clear social behaviors such as group living. Species that are thought of as solitary and that communicate through delayed scent marking cues have been largely overlooked. Here, we employ noninvasive genetics sampling to identify wild giant panda individuals across a study population in Southwest China and use spatiotemporal proximity thresholds to infer yearly, mating season, and non-mating season association networks and conduct social network analyses. We found clear evidence of social clustering in which cluster members were associated more frequently with each other than expected by chance, as well as a small core group of two to seven individuals to which associations were concentrated. Members of different social clusters were found in proximity in the Northwest portion of our study area, meaning this area may facilitate interactions across clusters and be important to the overall network structure. We recommend this region be targeted for prioritized conservation. Distance between individual activity centers was a significant negative predictor while genetic relatedness and being male were significant positive predictors of associations between individuals throughout the year. We found that genetic relatedness was a negative predictor of associations in the mating season, however, and a positive predictor during the rest of the year. Our results provide evidence for social structure in a species conventionally viewed as solitary, and our methods can be applied on any rare or elusive species for which direct observation is a challenge.

5.1 Introduction

Measuring the degree to which species exhibit social behavior has been a major endeavor in animal ecology (Stephens and Sutherland 1999, Wilson 2000). Some level of social behavior is inherent to all sexually reproducing species, and differences in interaction rates have the potential to heavily influence individual fitness through changes in reproductive success and survival rates (West-Eberhard 1979, Farine and Whitehead 2015). Intraspecific sociality also affects population-wide processes such as disease transmission, the transfer of information, and overall vital rates (Nunn et al. 2015). An understanding of social behavior in a species or population is thus an important piece to understanding their ecology as a whole and informing effective conservation action.

Social network analysis (SNA) has emerged as a valuable tool to study animal social behavior and has been applied to a variety of species (Lusseau 2003, Kurvers et al. 2014, Spiegel et al. 2018). Social networks are defined by a series of "nodes" (e.g., individuals) that are connected by "edges" (e.g., observed social interactions). SNA methods allow for the examination of individual and population-level influences on animal sociality (Lusseau and Newman 2004), as well as the effects of sociality on individual and population-level processes such as the spread of disease (Sah et al. 2018). In order to employ SNA methods, some measure of social interaction between individuals must be quantified (Farine and Whitehead 2015). The potential ways that animal social networks have been constructed can broadly be defined as "interaction" networks and "proximity/association" networks (Castles et al. 2014). Interaction networks use direct observations of social interactions between individuals to define edges in the network, while association networks rely on assumptions of social interactions between individuals based on their proximity in space and time (Castles et al. 2014).

The use of association networks allows for the application of SNA methods to species and populations that are difficult to directly observe and/or only individual co-occurrence data are available (Farine 2014). Association as a proxy for interaction, if defined well, has been shown to reasonably recover the structure of interaction networks in some systems when both data types are available (Farine 2014). Despite this potential for animal social networks to be defined via association, the vast majority of SNA applications to animal social networks have occurred in species seen as inherently social and featuring behavior such as group-living (Wey et al. 2008). However, there are many species across diverse taxa that engage in intraspecific associations in non-group settings. For example, scent-communication through marking surfaces is a common behavior in mammals (Johnson 1973), amphibians (Burn and Keogh 2007), and insects (Morgan 2009). In many cases scents can persist in the environment for extended periods of time (Soso et al. 2014).

Considering the potential for populations of many species to form social networks over time-lags, Formica et al. (2010) developed a method in which the weight of a given social interaction between individuals is measured as the number of times that each individual is observed within the other's home range. While this method can be applied to any study system in which multiple individual locations are determined, it is imprecise in its estimate of association. This is particularly the case for situations in which an individual's home range is substantially larger than the spatial range at which a delayed scent communication might take place. Many mammals, for example, maintain large home ranges and only deposit scent marks in locations that cover a small subset of that area (Vogt et al. 2014). In these cases, a stricter measure of the spatial proximity of two individuals is required in order to define a connection between them.

Giant pandas are an example of a species that are generally considered solitary but exhibit extensive scent marking behavior (Schaller et al. 1985, Pan et al. 2014). Pandas live in small (5-13 km^2) overlapping home ranges but are believed to live alone and communicate with one another largely from a distance by leaving marks created by rubbing their anogenital scent glands on the bark of 'scent post' trees located along well-traveled animal pathways throughout the habitat. In addition to scent marking, there is some observational and GPS-telemetry evidence for the potential formation of social groups in studied populations (Pan et al. 2014, Hull et al. 2015). But such evidence is anecdotal, leaving many unanswered questions about the degree of sociality in this threatened species. Due to the thick bamboo understory that makes up their habitat and elusive nature, direct observational studies of wild giant pandas are exceedingly difficult (Schaller et al. 1985). It is also challenging and cost-prohibitive to capture and fit global positioning system (GPS) collars on wild individuals, particularly for a large enough sample size for SNA. Due to these challenges and the perception of pandas as a solitary species, SNA of pandas has never been completed. Here, we address these challenges by using noninvasive genetic sampling to repeatedly observe panda individuals across a study population and use temporal and spatial thresholds to build association networks for the first time. We then use a social clustering algorithm to detect the presence of panda clusters within the population, a core-periphery algorithm to detect evidence of a core structure in the population, and additive and multiplicative effects (AME) social network models to analyze factors that may explain associations between individuals. Our results advance knowledge of giant panda ecology through the exploration of whether social ties exist and plausible mechanisms driving them. These methods can be used on any species that is difficult to observe and/or communicates with scent marks.

5.2 Methods

5.2.1 Study area

We studied the giant panda population in an area within Wolong Nature Reserve, Sichuan Province, China. Wolong is situated in a global biodiversity hotspot in the Qionglai Mountains and features a wide range of habitats and plant and animal species due to large elevation gradients spanning from 1,300 to over 6,000 meters (Myers 2000). It is also China's flagship national-level giant panda nature reserve and contains a wild panda population of around 150 individuals (Qiao et al. 2019). In addition to abundant wildlife and plant populations, Wolong is home to about 5,000 local residents who mainly reside in two towns situated along a national level road that runs through a valley that bisects the reserve (Fig. 5.1, Liu et al. 2016). We chose a region called Hetaoping (55 km^2) to focus our study of giant panda social networks, as this area contains extensive giant panda habitat featuring mixed deciduous/coniferous forests with vast swathes of understory F. *robusta* and B. *fabri* bamboo, resulting in high panda densities (Hull et al. 2015). Hetaoping represents a relatively discrete region of habitat to the North, West, and East due to the road, but is connected to relatively continuous habitat to the South (Fig. 5.1).

5.2.2 Sampling and genetic analyses

Noninvasive sampling of panda feces occurred from March 2015 to February 2016. The Hetaoping area was divided into contiguous 24 2 km^2 sample polygons, and teams of experienced field technicians and local guides conducted area searches of the polygons at multiple time points throughout the sampling period so that each polygon was sampled evenly. Any fresh feces detected during sampling was collected for analysis. Fecal samples were flagged as potential mother-cub pairs if they were found in close proximity and were of clearly different sizes. Samples were collected with care to avoid contamination, stored in sterile plastic bags, and

frozen within 6 hours of collection. Frozen samples were then taken to the genetics laboratory of the China Center for Research and Conservation of the Giant Panda (CCRCGP) in Daguan, Sichuan, China for storage and analysis.

DNA was extracted from the layer of epithelial tissue that coats the outside of fecal pellets using QIAamp DNA stool mini kits according to manufacturer protocols (Qiagen, Germany). The following seven microsatellite loci were then targeted for PCR: GPL-60, gpz-20, GPL-29, gpz-6, GPL-53, GPL-44, and gpz-47. These loci were selected from a large candidate set that showed levels of high polymorphism, lack of genotyping errors, and high amplifications success rate even after long exposure to weather elements based on a pilot study that compared captive panda feces and blood samples (see Huang et al. (2015). See chapter four for details and PCR methods. To ensure accurate genotypes, each locus was amplified using PCR and sequenced a minimum of three times for each sample in a multi-tubes approach (Taberlet et al. 1996). If two samples resulted in genotypes that differed by only one allele, DNA was re-extracted and amplified three additional times to confirm the samples as separate genotypes.

The MStools plugin for Microsoft Excel was used to identify unique individuals and repeat-samples from the same individual from the full dataset of microsatellite loci (Park 2001). The program GenAlex (Peakall and Smouse 2006) was used to estimate the probability of identity (PID) and the probability of identity of two full siblings (PSIB) based on matching loci using the methods of Waits et al. (2001). These metrics measure the probability that two matching genotypes may be two related individuals or two full siblings, respectively. Finally, we used the programs MicroChecker (Oosterhout et al. 2004) and FreeNA (Chapuis & Estoup, 2006) to identify allelic dropout, scoring errors, and null alleles at each locus that might cause errors in individual identification. For more details on quality control, see chapter four.

We tested the performance of six different genetic relatedness estimators using the 'related' package by first simulating 306 parent-offspring, full sibling, half sibling, and unrelated individual pairs (nearly equating the 1225 unique pairs contained in our final dataset of 50 individuals). We then calculated the relatedness estimates between the simulated pairs using each estimator and calculated the correlation between those estimates and the expected relatedness values of 0.5 for parent-offspring and full sibling pairs, 0.25 for half-sibling pairs, and 0 for unrelated individual pairs. We selected Wang's (2002) estimator for use in our analyses because it maximized this correlation at 0.68 (Table 5.1).

5.2.3 Constructing association networks

In order to analyze the social structure of the panda population in Hetaoping, we first needed to set criteria for building our association network. A key assumption in the analyses was that giant pandas can detect other specific individuals from their scent marks. This assumption was supported both by behavioral evidence from captive pandas due to their increased interest in scents of new individuals over habituated scents (Swaisgood et al. 1999) and the finding that different panda individuals' scent secretions have distinct chemical signatures (Lee and Hagey 2003, Yuan et al. 2004). In order to have engaged in scent communication, at least one panda per pair must have left a scent mark and the other must have investigated it. It is impossible to confirm with certainty whether this occurred for every pair of individuals whose feces we found within close proximity to each other, but due to the frequency of scent marking behavior we have observed at scent trees (nearly 50 % of site photographic evidence involving scent communication behavior, unpublished data) and the high density of scent marking trees we have found throughout the area (Fig. S8), we can conclude that these communications occur on a frequent basis. Scent trees are identifiable because they are stained a dark green from the

anogenital gland secretions and urine used in panda scent marks (Fig. S9, Nie et al. 2012). Both males and females are known to scent mark, but females may do so at a lower frequency than males outside the mating season (Nie et al. 2012).

We constructed our association network based on spatial and temporal proximity between individuals. We weighted connections between individuals found within 10 meters, 50 meters, 100 meters, and 200 meters as decreasing in strength from 4 to 3, 2, and 1, respectively. The justification for the 200-m spatial threshold was twofold: it has precedence in the literature in defining shared space-use (Hull et al. 2015) and captures the majority of panda movements recorded by GPS collars between 3 hour fixes (unpublished data). By weighting connections higher between individuals found within closer distances, we consider the increased likelihood of an association and strength of association between those individuals. We then used three different temporal thresholds of 1 month, 3 months, and 6 months in between sample collections of two different individuals falling within the above spatial proximity zones to define the presence or not of a connection. The justification for these temporal thresholds came from behavioral evidence that pandas can detect conspecific scent marks at least four months after deposition in indoor environments (Swaisgood et al. 2004) and at least three months after deposition in outdoor environments based on observations of pandas investigating the marks (Swaisgood, pers. comm.). The persistence of scent marks in the environment is heavily dependent on the relatively large size of the molecules and their lipid component (Soso et al. 2014), and panda anogenital gland secretions have both properties (Lee and Hagey 2003, Zhang et al. 2008). Although panda scent marks thus likely last a substantial amount of time, there is no certainty regarding how long unique individuals are distinguishable from their marks. We therefore created the three different networks based on decreasing strictness in the temporal

threshold (1 month, 3 months, and 6 months) used to define associations. Though the inferred associations come with some uncertainty, this spatiotemporal framework for defining networks is more precise than assuming that individuals with sample locations falling within each other's home ranges are associated, as has been done before (Formica et al. 2010). Finally, we adjusted the final weighted association matrices by counting the number of times that two individuals had samples fall within the spatial and temporal thresholds (Farine et al. 2015).

In order to investigate differences between panda social networks in and around the mating season compared to the rest of the year, we also built season-specific association networks. Specifically, we loosely defined associations occurring in and around the mating season as those in which the later fecal sample was collected between March 1st and June 15th. This resulted in the earliest and latest mating season associations being defined as March 23 and May 16, respectively. This represents a wider time window than the typical April-May mating season but should capture associations occurring adjacent to and perhaps influenced by mating activity. This definition resulted in n= 39 and n = 111 captures for the mating and non-mating season networks, respectively. Because we cannot be sure if all the same pandas were present for both seasons and the sampling window for the mating season is small, we built the season-specific networks only with the pandas that were captured within the same season. Finally, because the mating season is only about two months, we built 1 and 2-month seasonal association networks.

5.2.4 Social clustering analyses

We defined social clusters as groups of individuals with associations with other individuals within the group at a higher frequency, and associations with individuals outside the group at a lower frequency, than is expected by chance. We used the KliqueFinder algorithm

(Frank 1995) to examine the evidence for social clustering by maximizing the odds ratio presented in equation 1:

$$Odds\ ratio = \frac{(absence\ of\ ties\ outside\ of\ cluster) * (presence\ of\ ties\ within\ cluster)}{(presence\ of\ ties\ outside\ of\ cluster) * (absence\ of\ ties\ within\ cluster)} \quad (1)$$

To maximize this objective function, the Kliquefinder algorithm iteratively creates clusters, dissolves clusters, and reassigns actors - in this case panda individuals - to different clusters as needed. After final cluster assignments, we conducted n=1000 simulations in which we randomly reassigned ties between individuals and then applied the algorithm and measured the odds ratio in each. This allowed us to statistically test the significance of the clustering results by comparing our observed odds ratio to the distribution of those that arose from the simulations.

5.2.5 Core-periphery structure analyses

In addition to looking for evidence of cohesive clusters within our study population, we also conducted an analysis of core-periphery structure using UCINET (Borgatti et al. 2002). This method identifies nodes that make up the core of a network through a hill-climbing algorithm that redefines nodes to core or periphery values and maximizes the correlation between the permuted data matrix and an 'ideal' core-periphery structure matrix (Borgatti and Everett 2000).

5.2.6 Additive and multiplicative effects models

We built additive and multiplicative effects (AME) models to estimate the effects of several individual and pairwise variables on the probability of association weights between pandas using the 'ame' function in the 'amen' package in R (Hoff 2018). AME models are an extension of social relations regression models, in which a social relations covariance model is fit to account for random effects within individuals (e.g. age, size) and across pairs of individuals in addition to any fixed individual and pairwise effects. The AME extension can include additive

and multiplicative effects specified as latent factors estimated for each individual to account for dependencies inhering in the underlying latent social space (Hoff 2009). We modeled the number of associations observed between panda individuals as a ranked ordinal response with a probit link function. We used a Bayesian approach and employed Monte Carlo Markov Chain (MCMC) simulations to derive a posterior distribution of parameter estimates. We used a burn-in of 10,000 iterations and sampled 100,000 MCMC iterations to derive posterior estimates. The model was hierarchical in that both pairwise and individual-level covariates were considered and took the form:

$$Log\left(\frac{p(Y_{i,j})}{1-p(Y_{i,j})}\right) = \beta_i + \beta_j + \gamma_2 \,\mathrm{sqrt}\left(Distance_{ij}\right) + \gamma_3 Genetic\ Relatedness_{ij}$$

$$\beta_i = x_i sex + v_i \qquad (2)$$

$$\beta_j = x_j sex + v_j$$

Where $Y_{i,j}$ is the weighted association between individuals i and j, β_i and β_j are individual-level covariates that indicate sex plus random effects v associated with the individual, and the numbered γ terms are coefficients on the pairwise variables detailed in equation (2). An intercept was not included as it cannot be estimated using the ordinal form of the AME due to its treatment as a hazard function (Hoff 2018). We tested two different additions to the above model in order to account for social dependencies – one was to include 2-dimensional 'latent factors' introduced above (Hoff 2009). We also build models without these latent factors but including a pairwise variable denoting membership or not in the same KliqueFinder-derived social cluster. An individual's sex was identified using the DNA extracted from fecal samples and a 210-bp sex-specific SRY gene targeted with primers designed by Zhan et al. (2006). DNA from one male

and one female captive panda were used as positive and negative controls, respectively. The pairwise distance variable was calculated on the mean x and y coordinates of all sample locations for each individual, and the square root was taken for better model convergence. We used Wang's (2002) estimator of genetic relatedness to calculate the pairwise genetic relatedness values, as described above. To analyze potential spatial genetic structure in our samples, we plotted these genetic relatedness values against pairwise distances and looked for a trend. In order to examine potential differences in social behavior in the mating season vs. the non-mating season, we built separate AME models for the season-specific networks. These seasonal AME models took the same form as the model built using the entire dataset and the same Bayesian parameter estimation methods were employed. We evaluated model performance with goodness-of-fit (GOF) plots comparing histograms of network statistics simulated from the posterior distribution to the observed network statistics.

5.2.7 Sensitivity analyses

In addition to the standard p-value estimate of significance for parameters that is calculated from the posterior mean, standard deviation, and number of observations derived from the AME models, we calculated a percent bias that would be needed to invalidate the inference (Frank et al. 2013). Specifically, we employed the methods of Frank et al. (2013) to estimate this percent bias given the parameter posterior mean, standard deviation, the number of observations (possible panda pairs in network), and the number of parameters in the model. The result can be interpreted as the number of observations that would have to be replaced by null observations (parameter effect = 0) for the inference to be invalidated. We also compared the parameter estimates from the AME models of the separate seasonal networks to examine potential behavioral differences. Specifically, we tested for significant differences in the seasonal

parameter estimates using Cohen and Cohen's (1983) test implemented through the KonFound-it application (Frank 2014).

5.3 Results

5.3.1 Noninvasive genetics survey

The feces of 50 pandas were genotyped a total of 150 times. The number of recaptures per panda ranged from 0 to 17. Twenty-four pandas were only captured once. There was an approximately even sex ratio of 26 males and 24 females, but males were captured a total of 99 times while females were captured just 51 times. No mother-cub pairs were recovered, likely because there is a small window that cubs both consume bamboo (starting at around 1.5 years old) and remain with their mother (1.5 – 2 years old, Pan et al. 2014). The probability of two full siblings sharing identical genotypes across six of the seven loci analyzed was 0.01, so we included several individuals which had missing data at one locus. Microchecker results showed no evidence for large allele dropout or scoring issues at any locus, while FreeNA estimated an average null allele rate of less than 0.04 across the seven loci. Altogether, this suggests that the identification of individuals through our noninvasive genetics sampling was very accurate. Observed heterozygosity values for individual loci ranged between 0.40 and 0.80, expected heterozygosity values ranged from 0.34 to 0.79, and fixation indices ranged from -0.17 to 0.15. Genetic relatedness estimates among the 50 unique pandas followed a normal distribution around 0 (Fig. 5.2).

5.3.2 Networks, clustering, and core-periphery analysis

The association networks varied substantially based on the temporal proximity thresholds used to define ties between individuals, ranging from 32 ties in the association network with the strictest threshold assumptions of 1 month to 72 ties in the network with the least strict threshold

assumption of 6 months (Fig. 5.3). The KliqueFinder algorithm found strong evidence of social clustering in all association networks, with two distinct groups in the networks built under the 1 and 6-month thresholds and three distinct groups in the 3-month threshold network. The KliqueFinder algorithm maximized odds ratios of 1.62, 1.02, and 1.06 for clusters identified in the 1, 3, and 6-month threshold association networks, respectively. Significance testing through randomly distributing connections among individuals and reapplying the algorithm resulted in p-values of less than 0.01 for all three networks (Table 5.2). Core-periphery structure analyses of the 1, 3 and 6-month threshold networks resulted in three, two, and five pandas making up the core and correlation coefficients of 0.45, 0.60, and 0.67, respectively. In each case there was one female panda in the core and the rest were males.

5.3.3 Additive and multiplicative effects models

The AME models for the full association networks had fairly similar results. The model built using the 1-month threshold network had the most significant parameter estimates for nodal sex effects and genetic relatedness (Table 5.3). The 3 and 6-month threshold networks had only mildly significant effects of nodal sex, and the 6-month network did not have a significant effect of genetic relatedness. The square root of pairwise distance between activity centers was a highly significant variable in all networks (Table 5.3). Genetic relatedness remained a significant positive predictor of associations outside the mating season but had a negative, insignificant coefficient estimate in the mating season (Table 5.4). Goodness-of-fit plots indicated decent model fit in all models, but the models including 2-dimensional latent factors had slightly better fits than those with the KliqueFinder clustering parameter (Fig. 5.4). There was little evidence for spatial genetic structure in our study area (Fig. S10).

The percent bias sensitivity analyses suggested that between 77 and 90% of the observations would haver to be due to bias to invalidate the inference that the effect of distance has a negative effect on panda associations, depending on the network definition (Table 5.3, 5.4). Between 4 and 26% of the observations would need to be due to bias to invalidate the reference that males form more associations than females in the full networks (Table 5.3). In the full networks, between 25 and 38% of the observations would have to be due to bias to invalidate the inference that genetic relatedness has a positive effect on associations forming between pandas (Table 5.3). In the non-mating season 77.37% of the observations would need to be due to bias to invalidate the inference that genetic relatedness had a positive effect on associations forming between pandas. The Cohen and Cohen test for significant differences between the mating and non-mating season parameter estimates for genetic relatedness effects was highly significant ($p = 0.009$).

5.4 Discussion

Our results suggest that there is a higher level of social structuring present in wild giant panda populations than previously thought (Schaller et al. 1985, Gilad et al. 2016). Our findings of significant non-random associations throughout the year contradict the traditional view that pandas associate with conspecifics only during the mating season in April-May (Schaller et al. 1985, Pan et al. 2014). The joint-attraction to the same areas has been observed between panda individuals before with GPS telemetry data (Hull et al. 2015), and our clustering analysis suggests that this social behavior extend beyond individual pairs and that there is strong evidence for larger groups of pandas that make up social clusters. Although we found variation in the extent of this social clustering (i.e., number of clusters, the number of individuals making up each cluster, and the level of cohesion of these clusters) based on the temporal thresholds used to

define the association networks, all three networks featured at least two clusters with strong evidence for cohesion. This cohesion within clusters means that the 5-12 individuals making up each cluster were significantly more associated with each other than they were with any of the other 50 individuals outside their cluster.

The seasonal and full association networks deserve further consideration. Because the mating season is generally the only time of the year that giant pandas, as well as many other species viewed as solitary, are considered to engage in significant social interactions (Sandell 1989, Odden and Wegge 2004, Nie et al. 2012), other times of the year are often not considered. We found that most associations in our dataset occurred outside the mating season, however, and were spread across the year. Our seasonal AME model results also suggest important differences in behavior in the mating vs. non-mating season. Specifically, genetic relatedness was a significant positive predictor of associations outside of the mating season but was a significant negative predictor of associations in the mating season. This finding suggests some cryptic level of family-structure outside the mating season, potentially driven by multigenerational female-offspring associations that have been anecdotally observed long after offspring reach adulthood (Pan et al. 2014). The fact that this relationship reversed in the mating season makes sense from an evolutionary perspective – mating associations between individuals of lower relatedness will reduce the potential for negative inbreeding effects (Keller and Waller 2002). Though our seasonal AME models featured lower sample sizes and less of our observations would need to be replaced by null observations to invalidate our inferences compared to the full networks, our findings lead us to question the recent suggestion that pandas largely lack kin-recognition abilities outside mother-male offspring pairs and rely on female-biased natal dispersal to avoid

inbreeding events (Hu et al. 2017). We suggest that future work focus specifically on mating season association networks to better understand the behavioral differences we found.

Our AME model results on the full network bolster and provide new context for several previous observations in panda ecology. For example, the sex-covariate on panda individuals indicated that males interact with other pandas at a higher rate than females. Previous studies have shown that female home ranges feature minimal overlap with other females, while male home ranges overlap with both female and other male home ranges (Pan et al. 2014, Connor et al. 2016). It has also been found that adult males tend to make longer-distance movements and maintain larger home ranges than females, perhaps in order to check the status of other males and females in the area (Schaller et al. 1985, Hull et al. 2015). Research on captive pandas suggests that male pandas have better spatial memory than females (Perdue et al. 2011), which would better allow them to revisit scent-marking trees and specific areas perhaps frequented by other individuals. Our results support this general conclusion that males are instigating and maintaining more connections between themselves and other pandas than females. That said, we did infer a number of female-female associations. This could reflect a tendency for both males and females to make long-distance movements around the mating season (Zhang et al. 2014). Although the population of individuals that we detected featured an exactly even sex ratio, males were recaptured at a much higher rate than females. This finding may be related to fact that past research has shown that sub-adult females disperse from and immigrate to populations at a higher rate than males (Zhan et al. 2007, Connor et al. 2016).

Our results also provide new insights for understanding social dominance in pandas. The fact that the core of the networks only included one female and one to six males was interesting and likely reflect a strong male-male competition in this species. Most males were on the

periphery of the network structure and many (likely subordinate) were were never found in association with other individuals. Even though the correlations resulting from the core-periphery structure analysis suggest that this structure was generally weak and the cluster structure better describes associations in the network, it would be useful to analyze traits that may influence the core-periphery pattern such as age, size, and dominance-status (Nie et al. 2012, Moore et al. 2015). These traits could be experimentally teased apart in future noninvasive studies (see future directions).

Regarding other model results, the fact that distance been panda activity centers was a significant negative predictor of associations was expected. Given the fact that our association networks were built in part based on spatial proximity, it follows that distance would be a large factor in predicting those associations. There were examples of associations between individuals whose activity centers were distant, however, and there were some individuals holding membership in the same social cluster that were also far apart (around 7 km, Fig. 5.5). This suggests that associations may in some cases be maintained over large distances. The fact that genetic relatedness was a significant positive predictor in the absence of mother-cub pairs in our dataset suggests that family structure is likely an important determinant of associations in panda populations. Anecdotal observational evidence that male offspring may maintain relationships with their mother after sub-adulthood, and behavioral study on captive pandas has shown that females discriminate between close-kin and unrelated individuals after long periods of separation (Pan et al. 2014, Gilad et al. 2016). Additionally, we accounted for the distance between individual activity centers in our models, meaning that a genetic isolation by distance pattern is not a confounding factor (Wright 1943). This is also evidenced by the lack of spatial genetic structure in our relatively small study area (Fig. S10).

The spatial distribution of individuals belonging to the observed clusters in our panda networks has important conservation implications. The Northwest portion of our study area held activity centers of pandas belonging to up to three different clusters, suggesting that this may be a particularly important habitat area for conservation (Fig. 5.5). Possible reasons for its high value include that this area includes a wide elevational range (multiple bamboo species). Because this area borders the main road and associated pastures and human communities, its functionality as a habitat block that serves to connect separate social clusters may be threatened (Fig. 5.1). This area also contained the activity centers of individuals that were found to be members of the core strucutre, indicating their importance in the network structure. Wolong Nature Reserve currently designates areas as one of three levels of decreasing conservation strictness – "core," "buffer," and "experimental" zones. We recommend that, despite its closeness to the main road, the area depicted in Fig. 5.5 be transitioned to and maintained as a "core" conservation zone.

5.4.1 Limitations/Future Directions

It is important to note that our panda association networks and models are depictions and estimates of only that – associations. Because we used inferred scent-mark communications through spatiotemporal proximity thresholds as our measure for connections between individuals, we cannot conclude whether or not the individuals had direct physical and/or vocal interactions. There is evidence that these direct interactions do occur both inside and outside the mating season at some rate that is likely driven by panda density and resource availability (Reid and Hu 1991). There is also likely feedbacks between different modes of communication, as vocalizations have been shown to affect scent marking behavior in captive pandas (Xu et al. 2012). There is anecdotal evidence of direct interactions between pandas in the wild being at times peaceful and at others aggressive, with females potentially exhibiting more aggressive

territoriality with other females and males fighting for mating opportunities in the spring (Schaller et al. 1985, Pan et al. 2014). Social networks of panda populations that capture these distinctions would require extensive observational data that is unrealistic to collect at a large scale but has been done at a small scale (Nie et al. 2012).

We also could not reliably account for directionality. A scent mark communication network, as built here, is by nature directed in that the individual that deposits the scent mark advertises its presence and status to all other individuals that investigate the mark before it fades. The uncertainty of the relative timing of fecal deposition for individuals captured within days of each other, and which one or both detected the other's presence, prevented us from building a directed network, however. Camera traps offer a promising opportunity to collect finer temporal-resolution data indicating which individuals scent mark a given location and which detect that scent mark, as long as individuals can be reliably identified from photographs. Although mark-recapture studies employing camera traps to identify individuals has mainly been done in patterned big cat species (Carbone et al. 2001), there is also the potential to identify giant panda individuals from photographs (Zheng et al. 2016). Camera traps could also document direct physical and potentially vocal interactions, but their field of view is a limiting factor.

Fecal genetics surveys also offer additional future directions for social network analysis in pandas and other species. Examining more microsatellite loci or SNPs with age-class information would allow for a parentage analysis and the construction of detailed pedigrees (Jones and Arden 2003). This provides a powerful tool to examine the factors influencing mate choice in wild populations (Moore et al. 2015). Although it is impossible to determine the exact age of panda individuals through their feces, a reliable estimate of age class up to adulthood is possible through measuring the size of bamboo fragments found in the feces (Pan et al. 2014). In

addition to helping to determine parentage, this estimate of size may be a useful node-level predictor of associations in the network. It is also possible to extract hormone information from fresh panda feces (Nie et al. 2012), which is likely an important indicator of dominance and may also be a useful node-level predictor. These methods could be used to build association networks and gain a better understanding of social behavior in a wide variety of species throughout the year which are traditionally viewed as solitary outside the mating season and/or are difficult to observe. Carnivores are an example of a taxa in which many species are considered "solitary" and social behavior research has previously focused on the mating season (Sandell 1989, Hawkins and Racey 2009).

5.5 Conclusions

In this chapter, we presented the first social network analysis of a wild giant panda population. Using noninvasive fecal DNA sampling, we were able to identify individuals over a 55 km^2 region and build an association matrix based on spatiotemporal proximity. We found strong evidence for two to three social clusters whose members associated with others in their cluster more often than expected by chance. There was a subset of our study area that held the activity centers of individuals from multiple clusters, which we recommend be established as an important conservation area. We found that there was only a small core group of one female and two to six male pandas that featured concentrated associations and drove network structure. Finally, we showed that genetic relatedness, distance between activity centers, and co-membership in the same subgroup are significant positive predictors of associations between pandas. Being male had a positive effect, likely leading to the core structure of the network being made up of largely males. We present paths forward for future study, and argue that more rare,

elusive, and/or traditionally considered “solitary” species should be investigated with social network analyses for a better understanding of their ecology and more effective conservation.

APPENDIX

Table 5.1. Performance of genetic relatedness estimators, as measured by the correlation between simulated pairwise relatedness values based on observed allele frequencies and expected pairwise values given the relationship (parent-offspring, full sibling, half sibling, and unrelated).

Relatedness estimator	Correlation
Wang (2002)	0.68
Queller and Goodnight (1989)	0.67
Lynch and Ritland (1999)	0.67
Li et al. (1993)	0.67
Milligan (2003)	0.67
Wang (2007)	0.65

Table 5.2. Evidence of social clusters based on the KliqueFinder algorithm. P-values were derived from random permutations of ties between actors and recalculating the odds ratio to estimate a null distribution.

Network	Clusters identified	Odds ratio	P-value
1-month threshold	3	1.62	<0.01
3-month threshold	2	1.02	<0.01
6-month threshold	3	1.06	<0.01

Table 5.3. Parameter estimates from the Additive and Multiplicative Effects (AME) models built using the full association networks and predicting the number of associations between pairs of pandas. Percent bias to invalidate inference is only reported for significant parameter coefficients.

Parameter	1-month threshold estimate (SD)	P-val	% bias needed to invalidate inference (n. obs.)	3-month threshold estimate (SD)	P-val	% bias needed to invalidate inference (n. obs.)	6-month threshold estimate (SD)	P-val	% bias needed to invalidate inference (n. obs.)
Sqrt(dist) (meters)	-0.06 (0.007)	0.00	77.12% (1889)	-0.06 (0.005)	0.00	83.66% (2050)	-0.06 (0.007)	0.00	77.12% (1889)
Genetic relatedness	1.35 (0.43)	0.00	37.53% (920)	0.89 (0.34)	0.01	25.09% (615)	0.42 (0.33)	0.20	-
Individual sex (male)	0.45 (0.22)	0.04	4.13% (101)	0.45 (0.26)	0.08	10.56% (259)	0.85 (0.32)	0.01	26.18% (641)

a) Sample size of n = 50 individual pandas and n = 2450 possible panda pairs

Table 5.4. Parameter estimates from the Additive and Multiplicative Effects (AME) models built using the seasonal association networks and predicting the number of associations between pairs of panda individuals. Percent bias to invalidate inference is only reported for significant parameter coefficients.

Parameter	Mating season network			Non-mating season network		
	1-month threshold estimate (SD)	P-val	% bias needed to invalidate inference (n. obs.)	1-month threshold estimate (SD)	P-val	% bias needed to invalidate inference (n. obs.)
Sqrt(dist) (meters)	-0.04 (0.02)	0.08	-	-0.05 (0.01)	<0.00	77.37% (1332)
Genetic relatedness	-1.265 (1.35)	0.35	-	1.10 (0.48)	<0.00	54.60% (940)
Individual sex (male)	0.88 (0.89)	0.32	-	-0.06 (0.27)	0.83	-

a) Sample sizes of n = 19 and n = 42 panda individuals and n = 342 and n = 1722 possible pairs of pandas in the mating season and non-mating season networks, respectively.

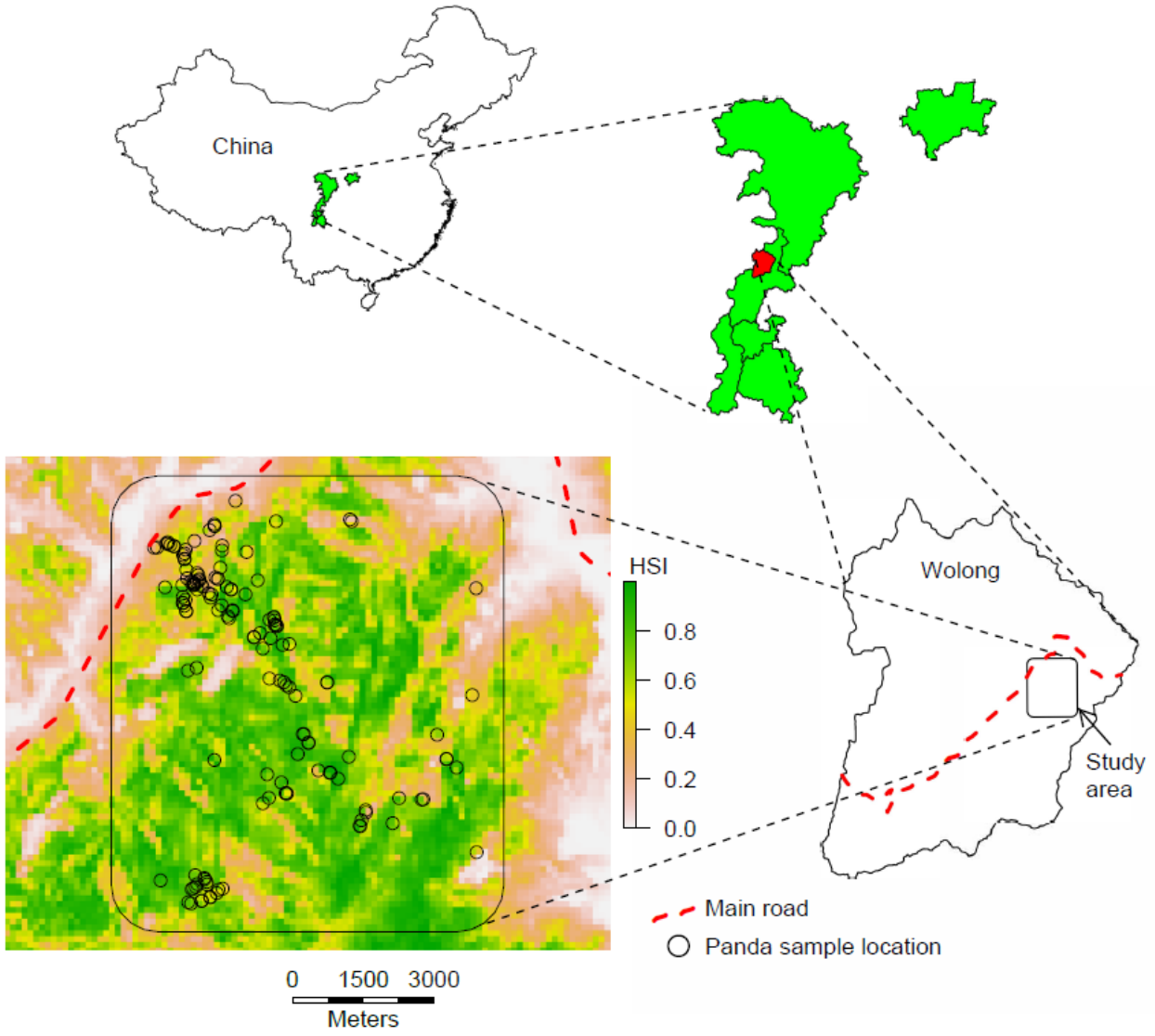

Figure 5.1. Study area inset on a map of China and the current giant panda geographic range. The finest scale inset depicts panda capture locations overlaid on a map of habitat suitability and the main road that bisects Wolong Nature Reserve. The habitat map was taken from Connor et al. (2019).

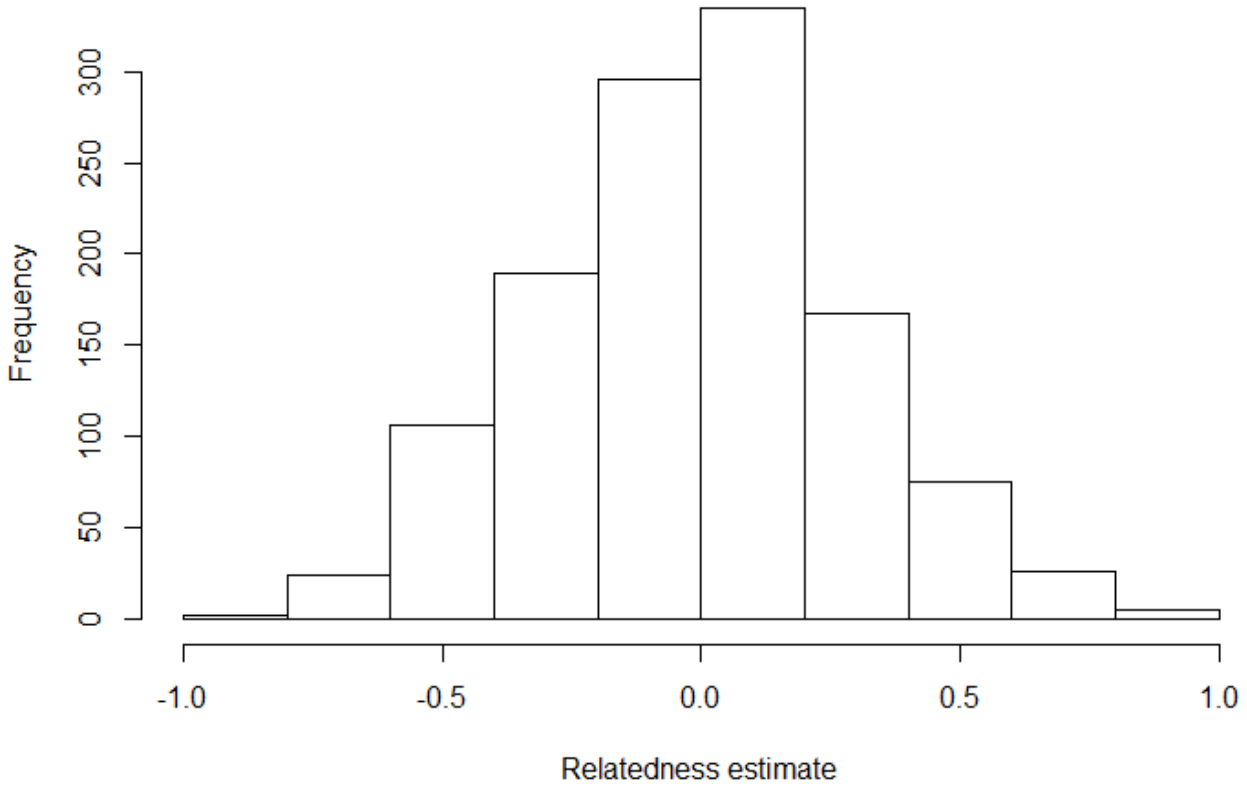

Figure 5.2. Histogram of pairwise genetic relatedness estimates between all unique pairs of the 50 panda individuals detected in the study. Genetic relatedness was estimated using Wang's (2002) estimator.

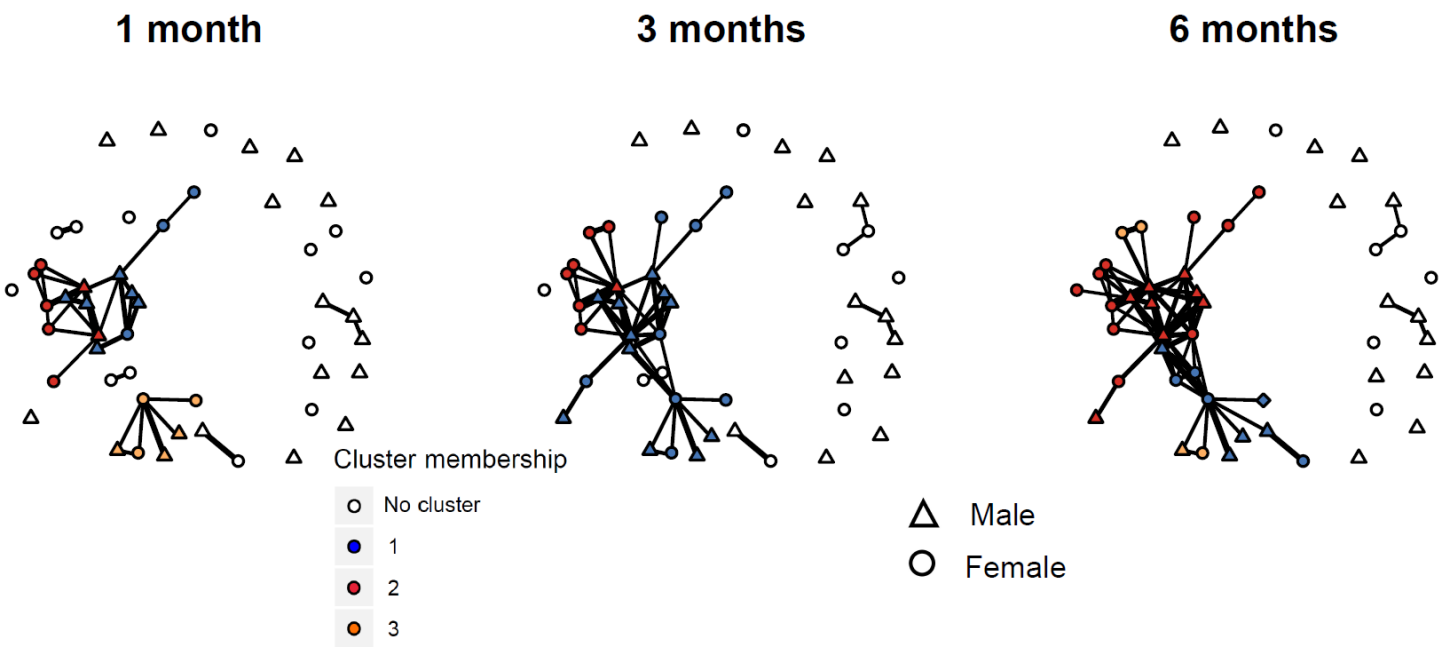

Figure 5.3. Panda association networks based on different temporal proximity thresholds for defining associations.

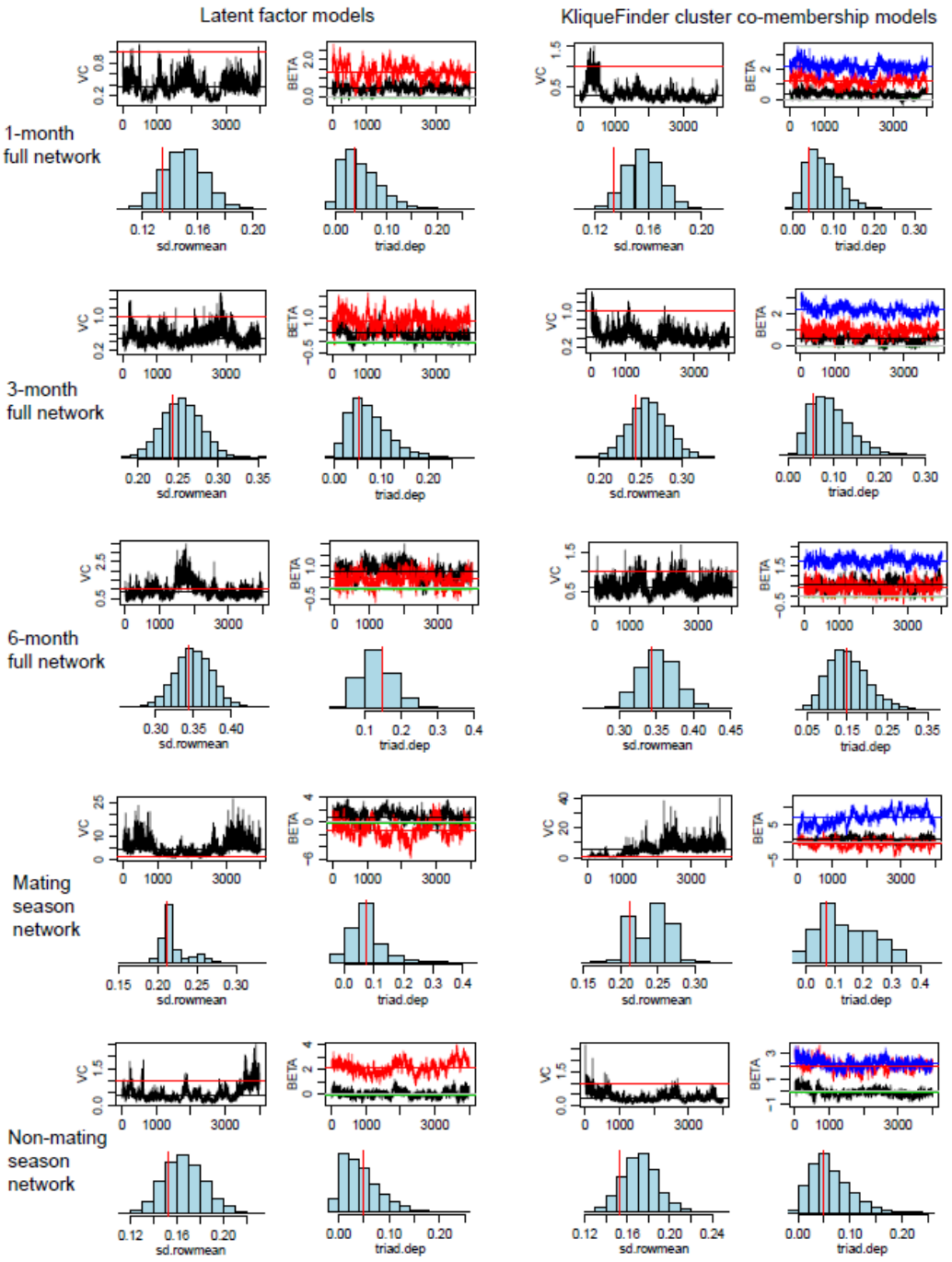

Figure 5.4. Goodness-of-fit plots of the AME models of the 1, 3, and 6-month full association networks and 1-month seasonal networks. Models fit with 2-dimensional latent factors are plotted in the left row and models fit with a co-membership in the same KliqueFinder cluster variable are plotted in the right row. Trace plots track the covariance parameter and specified fixed effects, while the histograms depict estimates of node random effects and triad dependence simulated from the posterior predictive distribution compared to the observed (vertical red line) statistics.

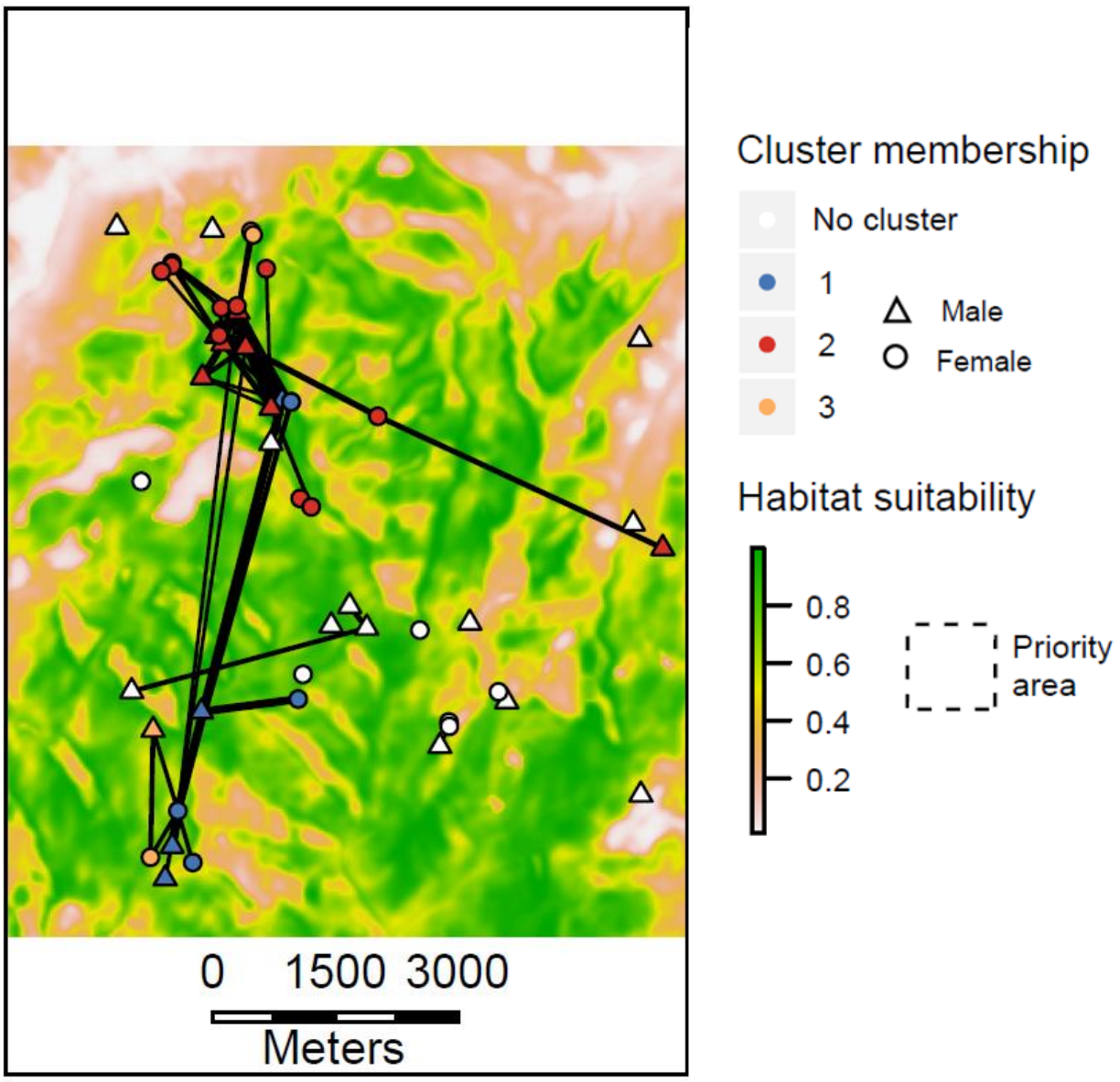

Figure 5.5. Spatial map of panda individual activity centers (mean of all X and Y coordinates of captures of a given individual) their social cluster membership, and sex. The social clusters depicted here are taken from the 3-month threshold association network results. The frequency of associations between individuals is represented by the thickness of the line connecting them. The network is overlaid on a habitat suitability map derived from Connor et al. (2019).

REFERENCES

REFERENCES

Borgatti, S. P., and Everett, M. G. 2000. Models of core/periphery structures. Social networks **21**:375-395.

Borgatti, S.P., Everett, M.G. and Freeman, L.C. 2002. Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.

Byrne, P. G., and J. S. Keogh. 2007. Terrestrial toadlets use chemosignals to recognize conspecifics, locate mates and strategically adjust calling behaviour. Animal Behaviour **74**:1155-1162.

Carbone, C., S. Christie, K. Conforti, T. Coulson, N. Franklin, J. Ginsberg, M. Griffiths, J. Holden, K. Kawanishi, and M. Kinnaird. 2001. The use of photographic rates to estimate densities of tigers and other cryptic mammals. Pages 75-79 in Animal Conservation forum. Cambridge University Press.

Chapuis, M., and A. Estoup, A. 2007. Microsatellite null alleles and estimation of population differentiation. Molecular Biology & Evolution **24**:621– 631.

Connor, T., V. Hull, and J. Liu. 2016. Telemetry research on elusive wildlife: a synthesis of studies on giant pandas. Integrative zoology **11**:295-307.

David Morgan, E. 2009. Trail pheromones of ants. Physiological entomology **34**:1-17.

Farine, D. R. 2015. Proximity as a proxy for interactions: issues of scale in social network analysis. Animal Behaviour:e1-e5.

Farine, D. R., and H. Whitehead. 2015. Constructing, conducting and interpreting animal social network analysis. journal of animal ecology **84**:1144-1163.

Formica, V. A., M. E. Augat, M. E. Barnard, R. E. Butterfield, C. W. Wood, and E. D. Brodie. 2010. Using home range estimates to construct social networks for species with indirect behavioral interactions. Behavioral Ecology and Sociobiology **64**:1199-1208.

Frank, K. A. 1995. Identifying cohesive subgroups. Social networks **17**:27-56.

Gilad, O., Swaisgood, R. R., Owen, M. A., and X. Zhou. 2016. Giant pandas use odor cues to discriminate kin from nonkin. Current zoology **62**:333-336.

Hagey, L., and E. MacDonald. 2003. Chemical cues identify gender and individuality in giant pandas (Ailuropoda melanoleuca). Journal of chemical ecology **29**:1479-1488.

Hawkins, C. E., and P. A. Racey. 2009. A novel mating system in a solitary carnivore: the fossa. Journal of Zoology **277**:196-204.

Hoff, P. D. 2009. Multiplicative latent factor models for description and prediction of social networks. Computational and Mathematical Organization Theory **15**:261-272.

Hoff, P.D. 2018. Additive and multiplicative effects network models. arXiv:1807.08038.

Hu, Y. B., Y. G. Nie, W. Wei, T. X. Ma, R. Van Horn, X. G. Zheng, R. R. Swaisgood, Z. X. Zhou, W. L. Zhou, L. Yan, Z. J. Zhang, and F. W. Wei. 2017. Inbreeding and inbreeding avoidance in wild giant pandas. Molecular Ecology **26**:5793-5806.

Huang, J., Y.-Z. Li, L.-M. Du, B. Yang, F.-J. Shen, H.-M. Zhang, Z.-H. Zhang, X.-Y. Zhang, and B.-S. Yue. 2015. Genome-wide survey and analysis of microsatellites in giant panda (Ailuropoda melanoleuca), with a focus on the applications of a novel microsatellite marker system. BMC genomics **16**:61.

Hull, V., J. Zhang, S. Zhou, J. Huang, R. Li, D. Liu, W. Xu, Y. Huang, Z. Ouyang, and H. Zhang. 2015. Space use by endangered giant pandas. Journal of Mammalogy **96**:230-236.

Johnson, R. P. 1973. Scent marking in mammals. Animal Behaviour **21**:521-535.

Jones, A. G., and W. R. Ardren. 2003. Methods of parentage analysis in natural populations. Molecular ecology **12**:2511-2523.

Keller, L. F., and D. M. Waller. 2002. Inbreeding effects in wild populations. Trends in Ecology & Evolution **17**:230-241.

Kurvers, R. H., J. Krause, D. P. Croft, A. D. Wilson, and M. Wolf. 2014. The evolutionary and ecological consequences of animal social networks: emerging issues. Trends in ecology & evolution **29**:326-335.

Li, C. C., D. E. Weeks, and A. Chakravarti. 1993. SIMILARITY OF DNA FINGERPRINTS DUE TO CHANCE AND RELATEDNESS. Human Heredity 43:45-52.

Liu, J., V. Hull, W. Yang, A. Viña, X. Chen, Z. Ouyang, and H. Zhang. 2016. Pandas and people: coupling human and natural systems for sustainability. Oxford University Press.

Lusseau, D. 2003. The emergent properties of a dolphin social network. Proceedings of the Royal Society of London. Series B: Biological Sciences **270**:S186-S188.

Lynch, M., and K. Ritland. 1999. Estimation of pairwise relatedness with molecular markers. Genetics 152:1753-1766.

Milligan, B. G. 2003. Maximum-likelihood estimation of relatedness. Genetics 163:1153-1167.

Moore, J. A., R. Xu, K. Frank, H. Draheim, and K. T. Scribner. 2015. Social network analysis of mating patterns in American black bears (Ursus americanus). Molecular ecology **24**:4010-4022.

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

Nie, Y. G., R. R. Swaisgood, Z. J. Zhang, Y. B. Hu, Y. S. Ma, and F. W. Wei. 2012. Giant panda scent-marking strategies in the wild: role of season, sex and marking surface. Animal Behaviour **84**:39-44.

Nie, Y., R. R. Swaisgood, Z. Zhang, X. Liu, and F. Wei. 2012. Reproductive competition and fecal testosterone in wild male giant pandas (Ailuropoda melanoleuca). Behavioral Ecology and Sociobiology **66**:721-730.

Nunn, C. L., M. E. Craft, T. R. Gillespie, M. Schaller, and P. M. Kappeler. 2015. The sociality–health–fitness nexus: synthesis, conclusions and future directions. Philosophical Transactions of the Royal Society B: Biological Sciences **370**:20140115.

Odden, M., and P. Wegge. 2007. Predicting spacing behavior and mating systems of solitary cervids: A study of hog deer and Indian muntjac. Zoology **110**:261-270.

Park, S.D.E. 2001. Microsatellite toolkit for MS Excel. Trinity College, Dublin.

Peakall, R., and P. E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes **6**:288-295.

Perdue, B. M., R. J. Snyder, Z. Zhihe, M. J. Marr, and T. L. Maple. 2011. Sex differences in spatial ability: a test of the range size hypothesis in the order Carnivora. Biology Letters **7**:380-383.

Pew, J., J. Wang, P. Muir, and T. Frasier. 2015. related: an R package for analyzing pairwise relatedness data based on codominant molecular markers. R package version 1.0.

Qiao, M., T. Connor, X. Shi, J. Huang, Y. Huang, H. Zhang, and J. Ran. 2019. Population genetics reveals high connectivity of giant panda populations across human disturbance features in key nature reserve. Ecology and Evolution **9**:1809-1819.

Queller, D. C., and K. F. Goodnight. 1989. ESTIMATING RELATEDNESS USING GENETIC-MARKERS. Evolution 43:258-275.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reid, D. G., and H. Jinchu. 1991. GIANT PANDA SELECTION BETWEEN BASHANIA-FANGIANA BAMBOO HABITATS IN WOLONG-RESERVE, SICHUAN, CHINA. Journal of Applied Ecology **28**:228-243.

Sah, P., J. Mann, and S. Bansal. 2018. Disease implications of animal social network structure: a synthesis across social systems. journal of animal ecology **87**:546-558.

Sandell, M. (1989). The mating tactics and spacing patterns of solitary carnivores. In Carnivore behavior, ecology, and evolution: 164–182. Gittleman, JL (Ed.). London: Chapman & Hall.

Schaller, G. B. 1985. Giant pandas of Wolong. University of Chicago press.

Schwarz, G. 1978. Estimating the dimension of a model. The annals of statistics **6**:461-464.

Soso, S. B., J. A. Koziel, A. Johnson, Y. J. Lee, and W. S. Fairbanks. 2014. Analytical methods for chemical and sensory characterization of scent-markings in large wild mammals: a review. Sensors **14**:4428-4465.

Stephens, P. A., and W. J. Sutherland. 1999. Consequences of the Allee effect for behaviour, ecology and conservation. Trends in ecology & evolution **14**:401-405.

Swaisgood, R. R., D. G. Lindburg, A. M. White, Z. Hemin, and Z. Xiaoping. 2004. Chemical communication in giant pandas. Giant pandas: biology and conservation 106.

Swaisgood, R. R., D. G. Lindburg, and X. Zhou. 1999. Giant pandas discriminate individual differences in conspecific scent. Animal Behaviour **57**:1045-1053.

Swaisgood, R. R. 2019, 9/20. Personal communication.

Van Oosterhout, C., W. F. Hutchinson, D. P. Wills, and P. Shipley. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes **4**:535-538.

Vogt, K., F. Zimmermann, M. Kölliker, and U. Breitenmoser. 2014. Scent-marking behaviour and social dynamics in a wild population of Eurasian lynx Lynx lynx. Behavioural processes **106**:98-106.

Waits, L. P., G. Luikart, and P. Taberlet. 2001. Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Molecular ecology **10**:249-256.

Wang, J. 2002. An estimator for pairwise relatedness using molecular markers. Genetics **160**:1203-1215.

Wang, J. 2007. Triadic IBD coefficients and applications to estimating pairwise relatedness. Genetics Research 89:135-153.

Wenshi, P. 2014. A chance for lasting survival: Ecology and behavior of wild giant pandas. Smithsonian Institution.

West-Eberhard, M. J. 1979. Sexual selection, social competition, and evolution. Proceedings of the American Philosophical Society **123**:222-234.

Wey, T., D. T. Blumstein, W. Shen, and F. Jordan. 2008. Social network analysis of animal behaviour: a promising tool for the study of sociality. Animal Behaviour **75**:333-344.

Wilson, E. O. 2000. Sociobiology. Harvard University Press.

Wright, S. 1943. Isolation by distance. Genetics **28**:114.

Xu, M., Z. Wang, D. Liu, R. Wei, G. Zhang, H. Zhang, X. Zhou, and D. Li. 2012. Cross-modal signaling in giant pandas. Chinese science bulletin **57**:344-348.

Yuan, H., D. Liu, L. Sun, R. Wei, G. Zhang, and R. Sun. 2004. Anogenital gland secretions code for sex and age in the giant panda, Ailuropoda melanoleuca. Canadian journal of zoology **82**:1596-1604.

Zhan, X., Z. Zhang, H. Wu, B. Goossens, M. Li, S. Jiang, M. W. Bruford, and F. Wei. 2007. Molecular analysis of dispersal in giant pandas. Molecular ecology **16**:3792-3800.

Zhang, J. X., Liu, D., Sun, L., Wei, R., Zhang, G., Wu, H., ... & Zhao, C. 2008. Potential chemosignals in the anogenital gland secretion of giant pandas, Ailuropoda melanoleuca, associated with sex and individual identity. Journal of chemical ecology, **34**: 398-407.

Zheng, X., M. Owen, Y. Nie, Y. Hu, R. Swaisgood, L. Yan, and F. Wei. 2016. Individual identification of wild giant pandas from camera trap photos–a systematic and hierarchical approach. Journal of Zoology **300**:247-256.

CHAPTER 6: CONCLUSIONS

Anthropogenic environmental degradation has become increasingly ubiquitous in the modern era, with wide reaching consequences for biodiversity and ecosystem function (Dirzo et al. 2014). Extinction rates have increased substantially in the Anthropocene, and the widespread decrease in wildlife abundance portends further increases to these rates (Ceballos et al. 2015, Rosenburg et al. 2019). In the face of these losses, it is important to optimize conservation strategies in order to maximize the chance for the survival of wildlife populations and species globally (Campagnaro et al. 2019). This requires accurate knowledge of species' distributions, habitats, and ecology in order to prioritize areas for protection and restoration. This knowledge is especially important for habitat specialist species that are sensitive to anthropogenic disturbance (Devictor et al. 2008).

In addition to the need to accurately define habitat according to a species' requirements, it may be important to consider the spatial configuration of said habitat when conserving landscapes. There is currently a widespread debate about the importance of considering habitat fragmentation, traditionally viewed as having negative effects on ecological responses like biodiversity, after the total amount of habitat is accounted for (Fahrig 2017, Fletcher et al. 2018). If the effects of habitat fragmentation are mostly negligible or even positive in more cases than negative (Fahig 2017), the spatial configuration of habitat is not a necessary consideration in conservation planning. Others argue that maximizing the amount of habitat should be a main goal of conservation but that its spatial configuration can be optimized to promote certain ecological responses like functional connectivity (Villard and Metzger 2014). Different relative effects of habitat amount and fragmentation on functional connectivity have been found in even simulation studies, however, resulting in uncertainty on how to optimize conservation efforts. Maintaining adequate functional connectivity in populations could be an effective way to

mitigate extinction risks in the face of further environmental degradations through the maintenance of genetic diversity, potential for the spread of adaptive genes, and recolonization of areas after localized extinctions (Allendorf et al. 2013, Hanski 1998).

In this dissertation, I and my colleagues endeavored to help solve the above conservation challenges by improving the estimation of species habitat and distribution, investigating the effects of habitat amount and fragmentation on functional connectivity, and employing novel analyses to better understand the ecology of elusive species. For this work, I used wild giant panda populations in Southwest China as a case study system. Pandas make an excellent study species for this work due to their habitat specialization and sensitivity to human disturbance, and through intensive fieldwork and collaborations I was able to put together an excellent dataset of panda presence locations, genetics data, and environmental variables. I also employed simulations of virtual species distributions on a real landscape in Scandinavia to investigate the effects of grain size and niche breadth on species distribution modeling. A common theme throughout the dissertation is that the spatial scale and environmental variables selected to measure ecological phenomena have substantial impacts on model accuracy and the interpretations of results. I argue that scale and variable selection should thus be optimized whenever possible using multiple iterations where these are varied. This will better inform our knowledge of species' ecologies and their responses to environmental variation and degradation in order to implement the most effective conservation action.

In Chapter 2, I investigated the effects of grain size on species distribution modeling by simulating four different virtual species that varied in niche breadth and the spatial grain size at which they were simulated across two landscapes of varying heterogeneity I then fit models with a wide range of grain sizes in order to measure grain size's effect on model predictive accuracy

and parameters. I found that there were significant decreases in accuracy as species distributions were modeled at grain sizes that grew further from that at which they were simulated. Importantly, the predicted area of a species' presence across the landscape became severely overestimated as grain size increased to a maximum of predicted distribution 14 times larger than the real distribution in the most extreme case. This implies that modeling at too large a grain size results in imprecise estimates of species presence and prevents the effective prioritization of areas for conservation.

In Chapter 3, I expanded on this investigation of spatial scale by looking at the interactive effects of grain size and the total extent of study areas on species distribution model accuracy and results in an empirical setting. I modeled giant panda distributions across four different total extents and seven different grain sizes ranging from 50 km^2 to the entire geographic range and 30 m to 1920 m, respectively. I found that increasing grain size reduced model accuracy at the smallest total extent, but that increases to extent negated the effect of increasing grain size. Increasing total extent generally increased model accuracy, but the most accurate distribution models were built at the second-largest extents consisting of separate mountain ranges. I also found that when predicting distributions in the smallest 50 km^2 extents, the models built at the next-larger (500 km^2) extent were the most accurate. I was thus able to show that the total extent of the area in which a species distribution model is trained may be best done at a different extent than the exact study area of interest. This implies that spatial scale should be more explicitly considered and optimized in future studies of species' habitat and distributions.

In Chapter 4, I further investigated ways to improve species distribution modeling accuracy and combined these habitat predictions with landscape genetic techniques to investigate the effects of habitat amount and fragmentation on functional connectivity in giant panda

populations. I found that a model incorporating a landscape resistance surface derived from the standard deviation of the core area index (CAI_SD) was the best predictor of functional connectivity, but that models based on the transformation of habitat amount and the connectance index (CONNECT) were also competitive. I also found that the relationship between habitat amount/fragmentation and functional connectivity was nonlinear – increases in the amount of habitat increased functional connectivity up to about 80% of the landscape consisting of habitat, at which further increases rapidly reduced connectivity. CAI_SD and CONNECT also featured thresholds that maximized functional connectivity which occurred at about SD = 2.5 and 12 % connected patches, respectively. These novel methods and results have important conservation implications in that we demonstrated that the restoration and preservation of habitat across a landscape can be optimized in such a way to target 80% habitat amount and reduced fragmentation in order to maximize functional connectivity in giant panda populations.

Finally, I leveraged the noninvasively collected genetics data used for the analyses in Chapter 4 to identify unique individuals across a core habitat area in my study area and conduct the first social network analysis of a giant panda population. Generally understood as a solitary species, I used spatiotemporal proximity thresholds to build association networks based on inferred scent communications. I found strong evidence for the presence of two to three social clusters consisting of individuals that associated significantly more with each other than individuals outside the cluster. There were some clusters featuring members that maintained associations across large distances between their activity centers, and one zone in the Northwest of our study area contained the activity centers of individuals from multiple clusters. I suggested that this zone be prioritized for more intensive conservation efforts as this would preserve important components of the network. I also built additive and multiplicative effects social

network models to determine what factors drive associations between pandas and found that genetic relatedness was a significant positive predictor of associations, indicating that some level of family structure to the associations is likely occurring. This relationship was reversed in the mating season, however, which is likely a product of inbreeding avoidance. My results showed that a predominantly solitary species like the giant panda forms social networks through delayed communications via scent marking, with implications for other species traditionally viewed and managed as solitary.

Collectively, this dissertation represents a thorough investigation into the factors influencing ecological inference from species distribution modeling and a step forward in understanding the effects of habitat amount and habitat fragmentation on functional connectivity. In this process, I advanced knowledge of giant panda ecology, which can be employed for more effective conservation of their remaining populations and habitat. I was also able to make generalizations about spatial scale effects on species distribution modeling that are important to consider in any system, and developed methodologies that can be widely applied in ecological research and conservation. I assert that the widespread practice of selecting spatial scale a priori based on expert opinion and/or native data formats should be replaced with rigorous optimization procedures like those we employed here. There is also more work needed to develop methods to better optimize spatial scale in ecological research – particularly in species distribution modeling. For example, finding the "scale of effect" for a given environmental variable on species presence has not been done in SDM studies. The need for this was recognized earlier and has been more explored in habitat selection research through the development of resource selection functions, methodology that should be integrated into SDMs (Zeller et al. 2017). Another avenue needing exploration is the inclusion and optimization of environmental variables

at different grain sizes in a single SDM, as opposed to resampling all variables to the same (and coarsest) grain size (Manzoor et al. 2018).

My research on the effects of habitat amount and fragmentation on functional connectivity, as well as the presence of social networks in giant pandas, needs to be reproduced in other species and systems. I posit that I have developed sound methodologies to investigate these topics in a wide variety of conditions, and their further use will help optimize conservation planning and further ecological knowledge in other systems. A better understanding of the effects of habitat amount and fragmentation on the functional connectivity in other species and landscapes has the potential to result in enormous conservation benefits. With increasing numbers of species and populations facing extinction risks and declines, respectively, pursuing conservation action to target habitat in a systematic manner to maintain functional connectivity is important to mitigate these threats. It is to this aim that we conducted the research detailed in this dissertation, and we encourage its widespread replication and application to conservation as the Anthropocene progresses.

REFERENCES

REFERENCES

Allendorf, F. W., Luikhart, G. H., and S. Aitken. 2013. Conservation and the Genetics of Populations. Wiley-Blackwell, West Sussex, UK.

Campagnaro, T., T. Sitzia, P. Bridgewater, D. Evans, and E. C. Ellis. 2019. Half Earth or Whole Earth: What Can Natura 2000 Teach Us? Bioscience **69**:117-124.

Ceballos, G., P. R. Ehrlich, A. D. Barnosky, A. Garcia, R. M. Pringle, and T. M. Palmer. 2015. Accelerated modern human-induced species losses: Entering the sixth mass extinction. Science Advances **1**.

Devictor, V., R. Julliard, and F. Jiguet. 2008. Distribution of specialist and generalist species along spatial gradients of habitat disturbance and fragmentation. Oikos **117**:507-514.

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.

Fahrig, L. 2017. Ecological Responses to Habitat Fragmentation Per Se. Pages 1-23 *in* D. J. Futuyma, editor. Annual Review of Ecology, Evolution, and Systematics, Vol 48. Annual Reviews, Palo Alto.

Fletcher, R. J., R. K. Didham, C. Banks-Leite, J. Barlow, R. M. Ewers, J. Rosindell, R. D. Holt, A. Gonzalez, R. Pardini, E. I. Damschen, F. P. L. Melo, L. Ries, J. A. Prevedello, T. Tscharntke, W. F. Laurance, T. Lovejoy, and N. M. Haddad. 2018. Is habitat fragmentation good for biodiversity? Biological Conservation **226**:9-15.

Hanski, I. 1998. Metapopulation dynamics. Nature **396**:41-49.

Manzoor, S. A., G. Griffiths, and M. Lukac. 2018. Species distribution model transferability and model grain size - finer may not always be better. Scientific Reports **8**:9.

Rosenberg, K. V., A. M. Dokter, P. J. Blancher, J. R. Sauer, A. C. Smith, P. A. Smith, J. C. Stanton, A. Panjabi, L. Helft, M. Parr, and P. P. Marra. 2019. Decline of the North American avifauna. Science **366**:120-+.

Villard, M. A., and J. P. Metzger. 2014. Beyond the fragmentation debate: a conceptual model to predict when habitat configuration really matters. Journal of Applied Ecology **51**:309-318.

Zeller, K. A., T. W. Vickers, H. B. Ernest, and W. M. Boyce. 2017. Multi-level, multi-scale resource selection functions and resistance surfaces for conservation planning: Pumas as a case study. Plos One **12**.