

Copyright
by
Cody James Schank
2018

The Dissertation Committee for Cody James Schank Certifies that this is the approved version of the following dissertation:

Investigation of novel methods to predict the distribution, abundance, and connectivity of rare species: a case study for the conservation of Baird's tapir

Committee:

Jennifer A. Miller, Supervisor

Eugenio Arima

Michael Daniels

Timothy Keitt

Weston W. Sechrest

Kenneth R. Young

**Investigation of novel methods to predict the distribution, abundance,
and connectivity of rare species: a case study for the conservation of
Baird's tapir**

by

Cody James Schank

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2018

Dedication

To Katherine, Callie, and Baxter.

Acknowledgements

As with any Dissertation, or really any project that lasts 5 years, there are so many people to thank and acknowledge for their support. First, I want to thank the Graduate School at UT Austin (especially Dr. Marvin Hackert, Elizabeth Korves, and Del Watson) for the Harrington Fellowship. I'd also like to thank my Supervisor, Dr. Jennifer Miller, and Committee (Drs. Arima, Daniels, Keitt, Sechrest, and Young) for their time and guidance. I also could not have made it through the last 5 years without the help of the dedicated and incredible staff of the Department of Geography and Environment: James Gunter, Teal Reid, and Tiffany Swonke. My fellow graduate students in the Department were also an endless source of inspiration, friendship, and a shoulder to commiserate on.

Outside of UT, I must first thank the Tapir Specialist Group (TSG), and all of those who specifically worked on this project or helped connect me to the right people, including Patricia Medici, Christopher Jordan, Armando Dans, Manolo Garcia, Diego Lizcano, Eduardo Mendoza, Ninon Meyer, Esteban Brenes-Mora, and Georgina O'Farrill. As well as many scientists not in the TSG, but who contributed to this research in important ways, including Mike Cove, Bob Dorazio, Marcella Kelly, and Clayton Nielsen (to name a few). Finally, I must thank Wes Sechrest and Global Wildlife Conservation for their long-standing support of my career as a conservationist and a scientist. I literally would not be here without them.

Last, I want to thank my family and friends for their love and support. This includes my oldest brother, John, who donated a computer to my research. Without those 12 cores

of high-speed processing power, the models presented in this Dissertation would still be running. My parents, for continuing to ask me when the dissertation would be finished. And my wife and daughter, for giving me the love to keep going through the hard days, and keep it all in perspective.

At risk of appearing incredibly corny and nerdy (though there is more of this in the Final Chapter), I must end with the wise words of Sam Hinkie, *trust the process*.

**Investigation of novel methods to predict the distribution, abundance,
and connectivity of rare species: a case study for the conservation of
Baird's Tapir**

Cody James Schank, PhD

The University of Texas at Austin, 2018

Supervisor: Jennifer A. Miller

Species Distribution Models (SDMs) are statistical tools used to develop continuous predictions of species occurrence. 'Integrated SDMs' (ISDMs) are an elaboration of this approach with potential as useful tools that allow for the dual use of opportunistically collected presence-only (PO) data and site-occupancy (SO) data from planned surveys. These models also account for survey bias and imperfect detection through the use of a hierarchical modeling framework that separately estimates the species-environment response and the detection process. This is particularly helpful for conservation applications and predictions for rare species, where data are often limited and prediction errors may have significant management consequences. Despite this potential importance, ISDMs remain largely untested under a variety of scenarios.

I applied the ISDM framework to Baird's tapir (*Tapirus bairdii*), an endangered large mammal that is found in Mesoamerica. The discrepancy between expert opinion and population estimates from the ISDM suggested further investigation was needed. Using a sensitivity analysis that tested modelling decisions, I found that site area had the strongest effect on the magnitude of population estimates and underlying intensity surface, which was driven by estimates of model intercepts. Selecting a site area that accounted for the

individual movements of the species within an average home range led to population estimates that coincided with expert estimates.

I used the lessons learned from the sensitivity analysis to employ an ISDM in a Range-wide Conservation Plan (RWCP) for Baird's tapir. The outputs from this model were combined with a hotspot analysis to delineate habitat patches for the species, and a graph network used to investigate the connectivity between these patches. In light of the results from this exercise, the following actions are recommended to facilitate the conservation of Baird's tapir: 1) protect existing habitat by strengthening enforcement in areas already under protection, and working with indigenous territories to preserve and enforce their land rights; 2) re-establish connections between habitat patches; 3) conduct additional non-invasive surveys in patches with little or no data on species presence; 4) collect more telemetry data on the species to better estimate home range size and dispersal capabilities; and 5) further develop robust models to more accurately predict true probabilities of presence, connectivity, and viability.

Table of Contents

| | |
|--|-----|
| List of Tables | xii |
| List of Figures | xiv |
| CHAPTER 1: Current practices and future directions in modeling the distribution of species to inform conservation..... | 1 |
| 1.1 Background | 1 |
| 1.2 Methodological Considerations | 5 |
| 1.2.1 Presence-absence Data..... | 6 |
| 1.2.2 Presence-only Data | 8 |
| 1.2.3 Spatially Biased Sampling of Presence Data | 10 |
| 1.2.4 Model Thresholding | 11 |
| 1.2.5 Accuracy Assessment | 12 |
| 1.2.6 Virtual Species Approach | 13 |
| 1.3 Theoretical Considerations | 13 |
| 1.3.1 Distributions and Niches..... | 14 |
| 1.3.2 Potential vs Realized Niche | 15 |
| 1.4 Point Process Models..... | 19 |
| 1.5 Frontiers in SDMs..... | 20 |
| 1.6 Research Objectives..... | 23 |
| 1.6.1 Baird’s tapir as a case study species | 24 |
| 1.6.2 Developing an international collaboration..... | 25 |
| CHAPTER 2: Using a novel model approach to assess the distribution and conservation status of the endangered Baird’s tapir | 27 |
| 2.1 Introduction..... | 27 |
| 2.2 Methods..... | 30 |
| 2.2.1 Study Area | 30 |
| 2.2.2 Model Formulation | 30 |
| 2.2.3 Description of Modeling Frameworks | 31 |
| 2.2.4 Presence Data..... | 34 |

| | |
|---|----|
| 2.2.5 Spatial Subsampling..... | 35 |
| 2.2.6 Model averaging | 35 |
| 2.2.7 Identification of habitat cores | 36 |
| 2.3 Results..... | 37 |
| 2.3.1 Coefficient Estimates | 37 |
| 2.3.2 Spatial Predictions | 40 |
| 2.3.3 Assessment of habitat cores and population estimates | 42 |
| 2.3.4 Comparison with other studies..... | 45 |
| 2.3.5 Habitat connectivity | 46 |
| 2.4 Conclusion | 47 |
| CHAPTER 3: A sensitivity analysis of Integrated Species Distribution Models (ISDMs): Investigating the effect of multiple spatio-temporal scales | 49 |
| 3.1 Introduction..... | 49 |
| 3.2 Methods..... | 51 |
| 3.2.1 Model Descriptions | 51 |
| 3.2.2 Model Convergence and Parameter Identifiability | 52 |
| 3.2.3 Presence Data..... | 53 |
| 3.2.4 Spatial Subsampling..... | 53 |
| 3.2.5 Predictor Variables..... | 54 |
| 3.2.6 Sampling Period Length | 56 |
| 3.2.7 Accuracy Assessment | 57 |
| 3.3 Results..... | 58 |
| 3.4 Discussion | 63 |
| 3.5 Conclusion | 66 |
| CHAPTER 4: Population status, connectivity, and conservation action for Baird's Tapir..... | 69 |
| 4.1 Introduction..... | 69 |
| 4.2 Data and methods..... | 71 |
| 4.2.1 Presence Data and Environmental Predictors | 71 |
| 4.2.2 Integrated SDM..... | 73 |

| | |
|--|-----|
| 4.2.3 Quantifying Habitat Connectivity | 75 |
| 4.2.4 Prioritization Criteria | 78 |
| 4.3 Results | 79 |
| 4.4 Discussion | 85 |
| 4.5 Conclusion | 92 |
| CHAPTER 5: Next steps - Linking SDMs, Connectivity analyses, and PVAs..... | 94 |
| 5.1 Introduction..... | 94 |
| 5.2 Linking SDMs, connectivity analyses, and PVAs | 95 |
| 5.2.1 Improving ISDM estimates..... | 96 |
| 5.2.2 Improving connectivity analyses | 99 |
| 5.2.3 Converting ISDM outputs to habitat patches..... | 101 |
| 5.3 Conclusion | 102 |
| 5.3.1 Including Stakeholders..... | 102 |
| 5.3.2 Public outreach, engagements, and sci-comm | 103 |
| 5.3.3 Expanding this research – applying these methods to other species | 103 |
| Appendix..... | 105 |
| Bibliography | 107 |

List of Tables

| | | |
|------------|--|----|
| Table 1.1: | Advantages and disadvantages of four SDM frameworks. The most common PO framework (i.e. Maxent), and the models which combine to create the Integrated SDM. | 10 |
| Table 2.1: | Categories of variables included in the candidate models. | 31 |
| Table 2.2: | Predictor variables in each category (* indicates quadratic term used in the model). | 32 |
| Table 2.3: | Mean of model-averaged coefficient estimates for each model framework, with mean standard errors in parentheses. Symbols indicate estimates greater than 1.96 standard errors from zero (i.e. 95% confidence intervals). Standard error estimates for PO model are not provided. | 39 |
| Table 2.4: | Mean of Akaike weights for each candidate model and model framework. Standard deviation of estimates in parentheses. Calculation of AIC is not straightforward for Maxent, thus we focus here on the other three frameworks. Cells left blank have no support (i.e. Akaike weight = 0). | 40 |
| Table 2.5: | Mean contribution, importance, and lambda values for each environmental variable. used in Maxent Model 9. | 41 |
| Table 3.1: | Model Convergence and Standard Estimation Rates for each model framework and combination of site area and number of samples. ... | 61 |
| Table 3.2: | Coefficient estimates and standard errors for the Integrated SDM (samples = 6) averaged across 100 model iterations fit on randomly subsampled presence data. | 62 |

| | | |
|------------|---|----|
| Table 4.1: | Coefficient estimates, standard errors, z-scores, and p-values. | 80 |
| Table 4.2: | Vulnerability, importance, and combined indices for conservation prioritization in the maximum dispersal = 160 km scenario..... | 84 |
| Table 4.3: | Forest loss rates (2000-2014), protected status, area, and location of protected areas that intersect tapir habitat patches..... | 88 |

List of Figures

- Figure 1.1: (A) Spatial distribution of presence points, (B) extent of occurrence and (C) area of occupancy delineated from the points (IUCN, 2001).....3
- Figure 1.2: BAM Diagram (Soberón & Nakamura, 2009). *A* represents the regions in geographic space that have abiotically suitable conditions for the species, corresponding to the potential, or Grinnellian, niche. *B* represents the regions that have biotically suitable conditions (e.g. competitors, predators, diseases, etc.), corresponding to the Eltonian niche. *M* represents the regions that the species can access. *G0* is the actual distribution of the species, while *G1* is the potential distribution which could be occupied if the structure of *M* changes.....17
- Figure 2.1: Predicted probability of presence for each model framework, (a) Maxent, (b) PO, (c) PA, and (d) Integrated.42
- Figure 2.2: Map of habitat cores for the Integrated model. Purple lines represent the proposed Nicaragua canal path and the Panama Canal, black points are presence-only sample locations, red points are detection history sample locations, and the area shaded light brown is the study area. Yellow points are study areas for independent abundance estimates (1. Corcovado, 2. Triunfo, 3. Los Katíos, 4. Oaxaca, and 5. Talamanca). Black circles are cores overestimated by the model (a. Encrucijada, b. Laguna de Términos, c. N and W of Selva Lacandona and d. Pixvae).43

Figure 3.1: Model accuracy using Boyce Index (BI) and Minimum Predicted Area (MPA). The difference calculated from the maximum (i.e. most accurate) Boyce Index, and from the minimum (i.e. most accurate) Minimum Predicted Area, within each combination of model settings. The greater the difference (from zero), the less accurate the result. MPA differences have been rescaled to positive numbers and are represented in units of 100,000 square kilometers.59

Figure 3.2: Population estimates from the Integrated SDM framework. Horizontal lines are placed at expert population estimates for the species, 3000 (current Red List assessment: Garcia et al 2017) and 16500 (Population Viability Assessment report: Medici et al 2005). Black triangles are the estimates for models run on the complete (not sub-sampled) set of presence data.60

Figure 3.3: Effective Sampling Area. A simplified diagram of two individuals with overlapping home ranges, and a camera trap in the area of their intersection. The effective sampling area is equal to the area that incorporates the activity centers of all individuals detected at the camera.65

Figure 4.1: Example of flow calculation for one source pixel going from patch 23 to patch 21. The source pixel (blue dot) contains 0.758 individuals. Black dots are additional pixels that fall within range of the destination patch, and will subsequently have LCPs and flow calculated. Blue lines are LCPs. Purple line is the Panama Canal.77

Figure 4.2: 3D response plot of intensity to EVI and forest cover.81

Figure 4.3: Map of habitat patches. Green areas are habitat patches. Brown areas are protected areas. Purple lines represent the Panama Canal and the planned path of the Nicaragua Canal. (a) - Santa Rosa, (b) - La Sepultura, (c) - El Triunfo, (d) - La Muralla.82

Figure 4.4: Maps of network connectivity. Thickness of edges indicates strength of flow between nodes. Size of nodes indicates estimated population size.83

CHAPTER 1: Current practices and future directions in modeling the distribution of species to inform conservation

1.1 BACKGROUND

Research and public attention on environmental conservation are growing rapidly as increasing amounts of evidence demonstrate a precipitous decline in biodiversity and ecosystem functioning, often due to land use change and other human-caused disturbances (Dirzo et al., 2014; Foley et al., 2005; Potts et al., 2010). Human influence on this dynamic requires the development of strategies to better manage natural resources, in order to preserve the immense benefit they provide to humanity (Millennium Ecosystem Assessment, 2005). The focus on species biodiversity is one of the central components of environmental conservation in general.

Species conservation revolves around two basic pieces of information: *what* are they? and *where* are they? Despite the simplicity of this information, the vast multitude of species and the difficult environments in which they occur have led to huge gaps in knowledge about biodiversity. These gaps are widely recognized as the Linnaean and the Wallacean shortfalls. Named after Carl Linnaeus (1707 - 1778), the father of modern taxonomy, the Linnean shortfall refers to the large number of species which remain unknown to science (Raven & Wilson, 1992). Some estimates have shown that approximately 90% of species are still awaiting discovery (Mora, Tittensor, Adl, Simpson, & Worm, 2011). While the Wallacean shortfall, named after Alfred Russel Wallace (1823 - 1913), refers to the limited information that exists about the distribution of species (Lomolino, 2004). This review will focus on issues surrounding the second problem - improving knowledge of species distributions.

Before diving into the specific concerns related to the study of species distributions for conservation purposes, it is important to note that there are other reasons to be interested in their study, including their use in examining biogeographic patterns and the macroecological theories that explain them. However, this review will seek to understand the science behind mapping the distribution of species as it relates to conservation. For this reason, the focus will often return to the challenges encountered when mapping rare species, or what to do in the case of limited information, as these scenarios are common when mapping the distribution of species for conservation. Species rarity can take many forms, with differing effects for how to understand and map species distributions (Gaston, 1994; D. Rabinowitz, 1981).

Several different strategies are used to map the spatial distribution of individual species. Range maps are one of the most basic of these methods. They are created by drawing a smooth outline around records of species occurrence, usually following expert opinion or boundaries of ecoregions (Lomolino, Riddle, Brown, & Whittaker, 2010). These range maps often grossly overestimate actual occupancy of the species (Hurlbert & Jetz, 2007), while the raw point locations on which they are based underestimate this property (Hurlbert & White, 2005). This issue (range maps vs. point locations) is also described by the differences between extent of occurrence (EOO) and area of occupancy (AOO), which are two ways to conceptualize a species' distribution (Gaston, 1991). EOO, the larger of the two, is generally delineated by drawing a minimum convex polygon (or some similar geometry) around the known populations of the species. This concept generally aligns with the process of creating range maps. On the other hand, AOO is defined as the area within the EOO that is occupied by the species (excluding cases of vagrancy) and is dependent on the resolution of the analysis (Figure 1.1). This delineation falls closer to a map based solely on point observations or confirmed locations of the species.

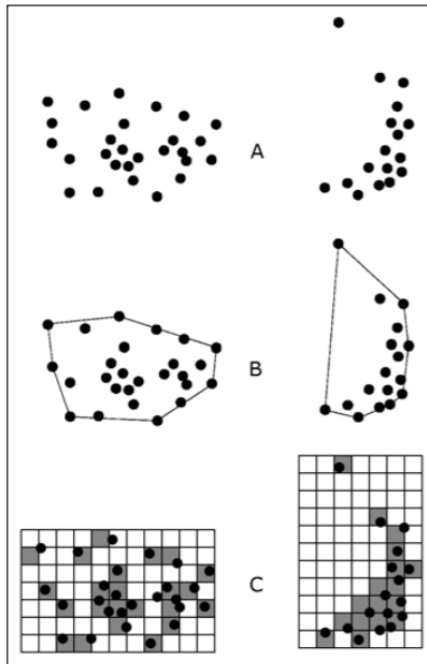


Figure 1.1: (A) Spatial distribution of presence points, (B) extent of occurrence and (C) area of occupancy delineated from the points (IUCN, 2001).

Despite the inaccuracy of range maps, they are one of the main tools used in IUCN's Red Listing process, which ranks species according to their threatened status. Two out of the five criteria for assessing species include variables related to geographic distributions (range size and fragmentation) (IUCN, 2016a), and range maps are often used to determine whether these criteria apply to the species being assessed. Another issue related to the Red List and species distributions, is that of species without enough information to form an assessment. These Data Deficient species represent nearly 20% of assessed species (IUCN, 2016b). Though sometimes this designation revolves around issues of taxonomic uncertainty, related to the Linnean shortfall mentioned above, often this assessment is reached when the species is so rare that it has been observed only a handful of times. Occasionally, the only piece of information about the distribution of a species is the type

locality at which it was discovered (as is the case with many amphibians) (Stuart et al., 2004).

Species distribution models (SDMs) are spatially explicit statistical tools used to model species-environment relationships (Guisan, Antoine, & Zimmermann, 2000). SDMs developed out of earlier disciplines, including ecological gradient analysis, biogeography, and geographic information science (GIScience) (Franklin, 2010a). SDMs have been referred to by many different names throughout the course of their development, including habitat suitability models (HSMs), ecological niche models (ENMs), predictive habitat distribution models, and resource selection functions (Elith & Leathwick, 2009), with species distribution models emerging as the most commonly used descriptor amongst current practitioners. SDMs experienced an explosion in popularity starting during the early 2000s (Vaz, Cunha, & Nabout, 2015) with the increase in availability of freely available data on both species occurrences (e.g. Global Biodiversity Information Facility - GBIF) and environmental predictors (e.g. WorldClim), as well as increasing computing power and availability of open-source software programs to prepare data and run models (e.g. MAXENT).

SDMs can be projected across geographic space to produce a spatial prediction of where the species occurs, often representing an improvement over expert range maps. However, SDMs cannot be implemented with insufficient data on the target species - ten to thirty records have been suggested as the minimum amount needed, depending on the modeling framework (Jiguet, Barbet-Massin, & Henry, 2010; Pearson et al., 2006; Stockwell & A. Townsend, 2002; Wisz et al., 2008). Deductive habitat models may be an option for species with too few observations to run an SDM (Rondinini et al. 2011). This challenge is sometimes referred to as the “Rare Species Paradox” in the SDM literature - the fact that those species which are most in need of conservation are sometimes the ones

for which SDMs cannot be effectively implemented (Lomba et al., 2010). Despite this challenge, when enough data do exist, SDMs represent a good compromise between range maps and point locations and are particularly appropriate for many conservation-related applications, such as projecting range expansions of invasive species, predicting species range shifts due to climate, and identifying suitable areas for translocations (Guisan et al., 2013). The ability to make these projections to new locations or time periods is one of the key advantages of SDMs over simple range maps or even deductive models.

1.2 METHODOLOGICAL CONSIDERATIONS

SDMs have been implemented across a variety of different methodological frameworks, including environmental envelopes, generalized linear models (GLMs), general additive models (GAMs), maximum entropy (i.e. MAXENT), decision trees, and other machine learning algorithms (Franklin, 2010a; Vaz et al., 2015). The choice of statistical framework is usually one of the most important decisions a researcher makes when implementing SDMs. In their seminal review, Guisan et al. (2000) pointed out the lack of studies comparing frameworks. Since then, this type of comparison has become commonplace, and most SDM studies now include more than one type of statistical framework (Elith et al., 2006). In fact, ensemble forecasting, a method which combines the results from many different frameworks (Araújo & New, 2007; Thuiller, Lafourcade, Engler, & Araújo, 2009), is a new and rapidly growing strategy in SDMs introduced to reduce the risk associated with focusing on one framework, and to quantify uncertainty (Vaz et al., 2015).

SDM research increasingly investigates the sensitivity of results to different modeling decisions, as these different user decisions (framework and settings) can lead to vastly different results (Elith & Graham, 2009). A major criticism levied against SDM

researchers in recent years is an unquestioning use of default settings, including a lack of awareness about model assumptions (Araújo & Peterson, 2012). Some of this misunderstanding revolves around the two main types of response data used in SDMs, presence-absence and presence-only (Jiménez-Valverde, Lobo, & Hortal, 2008; Soberón & Nakamura, 2009). For example, the most common type of data used in SDMs is presence-only, as it is readily available from digitized museum inventories or other sources. However, the most robust SDM methods are typically those that use presence-absence (PA) data. This has led to the ad hoc creation of pseudoabsences - points distributed (randomly or uniformly) across the study area that are used in place of absences in the model (Pearce & Boyce, 2006). In addition to the lack of agreement on the methodology for distributing these pseudoabsences (Barbet-Massin, Jiguet, Albert, & Thuiller, 2012; VanDerWal, Shoo, Graham, & Williams, 2009), their use has important implications for the interpretation of SDM outputs.

1.2.1 Presence-absence Data

Though presence-absence data are considered higher quality than presence-only data in many respects, there is a problem with absences related to their detectability. This issue is called imperfect detection (MacKenzie, Nichols, Hines, Knutson, & Franklin, 2003), and must do with the fact that absences are difficult to confirm (Lobo, Jiménez-Valverde, & Hortal, 2010). For example, just because a species was not observed at a time and place, does not mean the species does not occur in that area. It is possible the species was simply in a different area at that time (as with mobile animal species), or that it is hard to detect even when present (as with cryptic plant species). Occupancy models were developed to account for imperfect detection using a hierarchical framework (MacKenzie et al., 2006), also referred to as multilevel models, mixed models or random-effects models.

This hierarchical framework has separate levels for occupancy (the process of interest, i.e. the state variable) and probability of detection (the nuisance variable) (M. Kery & Royle, 2015; J. A. Royle & Dorazio, 2008). These levels are linked through conditional probability; in this case, the detection of the species is conditioned on its presence (i.e. the species cannot be observed if it is not present). This type of modeling also allows for nested, non-independent data, and is thus a good fit for data collected under the repeat observation design described below.

To estimate detectability, multiple observations at each sampling location are usually required. Camera traps are particularly useful for providing this type of data, as they can be set and left in the field for weeks or months at a time (O'Connell, Nichols, & Ullas Karanth, 2010). Whenever the motion sensor is tripped, a photo is taken. This data can then be separated into sampling occasions, usually periods that last multiple days, to avoid having too many non-detections and to ensure independence of detections within a site (Rovero and Zimmermann 2016). A 1 or a 0 is then assigned according to whether the species of interest was detected or not during that occasion. These sampling occasions combine to make a detection history at each sample location (i.e. a vector of 1's and 0's). For example, a camera may be deployed for a fifty-day period and split into sampling occasions of 10 days each. In this case, the detection history for a site might look like 00101 , if the species went undetected on sampling occasions 1,2, and 4, while it was detected on sampling occasions 3 and 5. In a simple detection model without covariates in the detection process, the probability of detection for the species at this site is $2/5$, or 40%. This implies that there is nearly an 8% chance that the species will go undetected across all five observations, even though it is present at the site. For this reason, SDMs that do not incorporate imperfect detection, often underestimate the true occupancy of the species across the study area. The length of the detection history is usually set to be long enough

to estimate detectability reliably, but also meet the assumption of site closure (no immigration or emigration from the site during sampling).

The fitted occupancy model can be used to make predictions at unsampled sites across the study area. However, data collected for occupancy models are often limited to small study areas, since this type of sampling is costlier and time consuming than collecting presence-only data (which might only require downloading museum records from GBIF). Due to this limited spatial coverage, predictions from occupancy models are typically not used to project species distributions across larger unsampled regions (Marc Kery, Gardner, & Monnerat, 2010). Though it has been shown that failing to account for imperfect detection can lead to biased parameter estimates (Dorazio, 2012; Lahoz-Monfort, Guillera-Aroita, & Wintle, 2014) among other issues (e.g. underestimation of occupancy), this type of framework is often not discussed in the SDM literature (Franklin, 2010a; Townsend Peterson et al., 2011), possibly due to their limited use in making spatial predictions, one of the main emphases of SDMs (Kéry, 2011). A Web of Science search (conducted February 6th, 2018) returned 5618 articles with “species distribution model*” or “ecological niche model*” in the topic, 1825 with “occupancy model*” or “imperfect detection”, and only 94 articles that are found in both searches.

1.2.2 Presence-only Data

Presence-only data have become one of the most commonly used response variable types, as they are easily collected or digitized from museum records. Multiple frameworks for modeling presence-only data in SDMs have been developed throughout the years. Early works found these models to be useful tools for conservation management, despite the limitations of PO methods in general (Pearce & Boyce, 2006). Some of these early presence-only frameworks (e.g. envelopes) are no longer used, following the introduction

of a presence-only algorithm based on maximum entropy (Phillips, Dudík, & Schapire, 2004) that has often been found to perform well in terms of predictions (Elith et al., 2006; Vaz et al., 2015). This framework is most commonly implemented through the software MAXENT, and can fit complex response curves, yet it has been criticized as a “black box”. Despite concerns voiced by (Yackulic et al., 2013a) about the widespread misuse of MAXENT, including lack of accounting for spatially biased samples and misinterpretation of model outputs, this framework remains one of the most popular used in SDM studies today (Vaz et al., 2015).

One issue with presence-only models is that they can only provide predictions about relative probability of presence, not absolute (i.e. actual) probability of presence (Hastie & Fithian, 2013). For example, in a presence-only model, a prediction of 0.5 does not mean that there is a 50% probability that the species occupies the site, it is instead analogous to the site being half as suitable as the most suitable site in the study area. Despite this challenge, some researchers have developed methods which claim to estimate absolute probability from models built on presence-only data (Christopher T. Rota et al., 2013). However, other authors have shown that these methods fail to adequately estimate true probability of presence, stating that there is “no panacea for lack of data on species prevalence” (p. 1409) (Phillips & Elith, 2013). Estimates of absolute probability of presence are more meaningful, therefore, researchers will continue to investigate ways to estimate this property with limited data. This is especially true of SDMs used for conservation, as estimates of absolute probability (or even density) are necessary for certain applications, including identification of sites for translocations and reintroductions, conservation planning, reserve system design, and integration with Population Viability Assessments (Townsend Peterson, 2006).

| Method | Advantages | Disadvantages |
|--------------------------|---|--|
| Maximum Entropy (MAXENT) | -Often performs well in model comparisons that assess accuracy of predictions | -“Black box” -Assumptions of the model are often ignored, and defaults setting used where inappropriate -No straightforward method to account for spatially biased sampling data |
| Point Process Model | -Provides well-developed diagnostic tools to assess models -Clear methods to follow for selection of quadrature points (i.e. pseudoabsences) | -Does not account for imperfect detection |
| Occupancy Models | -Accounts for imperfect detection | -Requires data collected from planned surveys (i.e. detection histories or point counts) that often cover limited geographic areas |
| Integrated SDM | -Combines all the advantages of PPMs and occupancy models, so that presence-only data and presence-absence data can be analyzed together | -Untested on real-world applications |

Table 1.1: Advantages and disadvantages of four SDM frameworks. The most common PO framework (i.e. Maxent), and the models which combine to create the Integrated SDM.

1.2.3 Spatially Biased Sampling of Presence Data

Another common issue in SDMs is spatial bias in the sampling of presence data, which violates the assumption of a random sample. This is not to be confused with spatial autocorrelation, which is a spatial pattern that can result from species interactions or spatially structured environmental data (J. Miller, 2012). Sampling for species can be difficult and costly, whether due to steep terrain, impassable swamps, or the inability to access private land. Often sampling occurs close to roads, due to the ease of access, or in protected areas sampling might be biased towards navigable rivers (the main method of

traveling in these environments). If sampling effort is well documented, most methods provide means to correct for the uneven sampling. However, this information is rarely recorded, often leading modelers to use ad hoc approaches to incorporate sampling bias in the model (Beck, Böller, Erhardt, & Schwanghart, 2014; Clements et al., 2012).

Taking a global perspective, there is a spatial bias of data collection towards temperate zones and Western countries, and less effort in the Tropics and developing countries (Collen, Ram, Zamin, & Mc Rae, 2008; C. Meyer, 2015). This is particularly concerning because most of the World's biodiversity and endangered species are found in Tropical regions. This exacerbates some of the challenges to SDM research which have already been mentioned, including working with limited presence data, and leads to a lack of SDM studies where they are needed the most (Cayuela et al., 2009; Feeley & Silman, 2011).

1.2.4 Model Thresholding

SDMs typically generate a probability or suitability score which ranges from 0 to 1. It is common practice to threshold this result to produce a binary presence-absence map for the species. This might be for practical reasons, such as creating a refined range map, or it might be to assess the accuracy of the model by comparing the prediction to presence-absence test data (either withheld from the data set, or independently collected). When assessing the accuracy of a thresholded SDM, there are two types of errors which are possible: predicting the species as absent in an area when in fact it is present (i.e. false negative), and conversely, predicting the species as present when it is absent (i.e. false positive). Modelers try to balance these two types of errors, since focusing on lowering the incidence of one type of error usually causes the other to increase.

1.2.5 Accuracy Assessment

When using SDMs for conservation, some papers suggest general rules to follow to prioritize the types of errors that are generated (Araújo & Peterson, 2012). For example, one could claim that false positives are costlier, to avoid wasting limited conservation resources on areas where the species is not found. On the other hand, false negatives could be costlier, if the species is so rare you cannot afford to lose any possible habitat. The general rule to follow, is that the costlier error depends on the context (Loiselle et al., 2003), possibly considering the rarity of the species and the availability of resources that can be used for conservation actions.

Interestingly, if enough presence data are available, rare species can produce more accurate models than widespread species, possibly because they have narrow environmental niches (Hernandez, Graham, Master, & Albert, 2006; M. McPherson & Jetz, 2007; Tsoar, Allouche, Steinitz, Rotem, & Kadmon, 2007). However, the increased accuracy from these studies may also be due to the properties of the data and the accuracy metric used, particularly the area under the receiver operating characteristic curve (AUC) (Jiménez-Valverde et al., 2008; Lobo, Jiménez-Valverde, & Real, 2008). The use of this measure has become common in SDM accuracy assessment because it is “threshold independent”, and thus does not require the user to make difficult decisions about which threshold to use (Liu, Berry, Dawson, & Pearson, 2005). Implicitly, rare species have lower prevalence (ratio of presences to absences) and relative area of occurrence (ratio of extent of occurrence to extent of study area), which leads to inflated AUC. Despite this and other problems (i.e. using pseudoabsences as test data), the frequent use of AUC in SDM studies has led to an unquestioning reliance on it to measure accuracy, when other metrics might be more appropriate. In fact, it is now recommended practice to include more than one measure of accuracy, as they can provide different perspectives (Lobo et al., 2008).

1.2.6 Virtual Species Approach

Simulated data are becoming more popular in SDM studies to investigate the effects of modelling decisions and to test model assumptions (J. Miller, 2014; Zurell et al., 2010). This is especially true of models that implement new and novel methods to demonstrate their advantages over more traditional methods (Crane, Liedloff, & Wintle, 2012; Dorazio, 2014a). Real species data present a challenge due to the multiple confounding effects of unknown sampling effort, complex species-environment relationships, unknown or unmodelable historical biogeographic contexts, and species interactions. Virtual species are useful in that the researcher knows all of the necessary information about the species and can change these settings to investigate the effect of issues that exist with real species data (e.g. spatially biased sampling effort, sparse data, or low species detectability). There are multiple considerations when generating virtual species for SDMs, including the complexity of virtual species responses to the environment, whether to include stochasticity (or randomness in the species response), and whether to use real environmental data (as opposed to simulated layers). The use of virtual species is now common in both SDM (Elith & Graham, 2009) and occupancy studies (Guillera-Aroita, Ridout, & Morgan, 2010).

1.3 THEORETICAL CONSIDERATIONS

Methodological decisions made by the SDM user have important implications for the interpretation of model results. These interpretations have a statistical basis, but also a theoretical one. For example, is the mapped prediction from the model the true distribution of the species? Below I discuss the different theoretical frameworks that impact the understanding and implementation of SDMs. This includes the BAM framework, niche theory, potential vs. actual distributions, and indirect vs. direct predictors.

1.3.1 Distributions and Niches

Species occupy geographic space through the movements and locations of individuals, and these individuals combine to make up the species' distribution, an abstract representation of where the species can be found. This spatial configuration is generally determined by the ecological requirements of the species, also referred to as the species' niche. Much has been written about ecological niches in general (Chase & Leibold, 2003; Elton, 1927; Grinnell, 1917; Hutchinson, 1957), as well as their relation to species distributions (Soberón, 2007; Townsend Peterson et al., 2011). The niche is determined by a variety of factors, including climate, resource availability (i.e. food, etc.), predators, competitor or facilitator species, and even parasites and diseases. Considering the many different definitions of the ecological niche, from simple to complex, it is convenient to focus on two definitions when thinking about SDMs: the Grinnellian and the Eltonian. The Grinnellian niche is determined by “non-interactive”, abiotic variables (e.g. climate) at coarse environmental resolution (also sometimes referred to as scenopoetic variables), while the Eltonian niche includes biotic interactions which act at local scales (Soberón, 2007). These two niche definitions form the basis of the BAM framework (for Biotic, Abiotic and Movement).

The BAM framework was developed to better understand the theory behind SDMs (Soberón & Peterson, 2005). This conceptual model shows that species distributions are determined by three main categories of factors: Biotic, Abiotic, and Movement. Abiotic variables include climate, topography, and other physical environmental factors which determine the potential distribution of a species. Biotic variables include competitive or symbiotic relationships with other species, while movement variables determine if a region is accessible to the species in question, either due to distance or dispersal barriers. These

last two factors, biotic and movement, in combination with abiotic factors, determine the actual distribution of a species.

It is important to understand niche theory when conducting SDMs, as the species' niche determines the environmental space occupied by the species, which in turn is manifested in geographic space. It is this back and forth interplay between geographic and environmental space (also known as "Hutchinson's Duality") which allows us to model species distributions using SDMs (Colwell & Rangel, 2009). Despite the importance of niches, SDM researchers have generally focused on species distributions, as opposed to niches, because the latter are harder to define and there is ongoing debate about what a "niche" means (McInerny & Etienne, 2012).

1.3.2 Potential vs Realized Niche

Niches have three nested realizations which help elucidate the relationship between environmental and geographic space. The fundamental niche represents the entire environmental space that a species could utilize (Hutchinson, 1957), while the potential niche includes those parts of the fundamental niche which exist in the study area at any given time (Jackson & Overpeck, 2000). Finally, the realized niche is the environmental space occupied by the species at that time (Hutchinson, 1957). As with the BAM diagram (Figure 1.2) there are several reasons why a species may not be able to access or utilize the full extent of its potential niche. With the fundamental niche, though more abstract, it is possible to imagine a scenario where a species could utilize environmental conditions that do not currently occur in the study area. For example, a species may have evolved to tolerate environmental conditions that occurred in the past, and due to niche conservatism (Peterson, Soberón, & Sanchez-Cordero, 1999), retained the ability to occupy that portion of environmental space for some later amount of time during which those conditions no

longer occurred (see Figure 3.2 in Colwell & Rangel (2009)). Understanding niche conservatism, and its converse niche evolution (Holt & Gaines, 1992), will have important implications for species' ability to adapt or persist in the face of climate change.

Though there is no analog with distributions for the fundamental niche, the potential and actual (i.e. realized) distribution are two concepts often used in SDMs. The duality between environmental space and geographic space is one-to-many, such that a single point in environmental space may correspond to several locations in geographic space (Colwell & Rangel, 2009). For this reason, the potential and actual niche are not exactly equivalent to the potential and actual distributions. For example, a species might occupy its entire potential niche, but not the complete potential distribution (if similar conditions occur in a far away, inaccessible location). Despite this inequality, potential and actual are very important distinctions to remember when using SDMs.

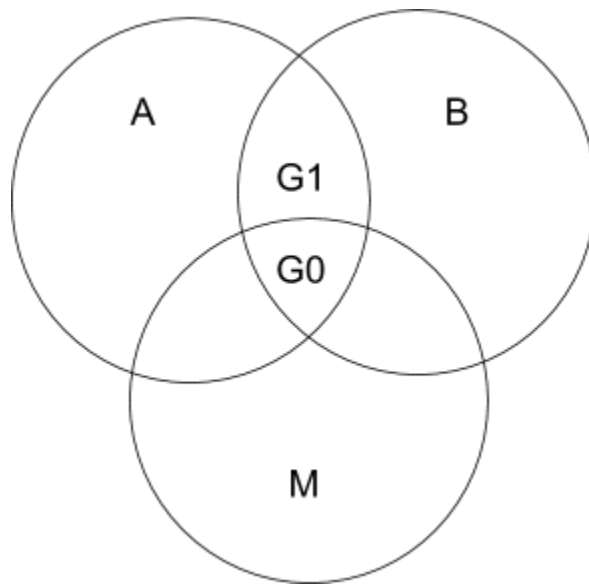


Figure 1.2: BAM Diagram (Soberón & Nakamura, 2009). *A* represents the regions in geographic space that have abiotically suitable conditions for the species, corresponding to the potential, or Grinnellian, niche. *B* represents the regions that have biotically suitable conditions (e.g. competitors, predators, diseases, etc.), corresponding to the Eltonian niche. *M* represents the regions that the species can access. *G0* is the actual distribution of the species, while *G1* is the potential distribution which could be occupied if the structure of *M* changes.

Whether the SDM predicts the potential or the actual distribution is generally determined by the type of presence data and environmental layers used in the analysis. Presence-only (PO) data can only provide information about the potential distribution, because they indicate only where the species can be, not where it cannot exist. Meanwhile, presence-absence (PA) data can be used to model the actual distribution of the species, because they provide information about the areas not occupied by the species.

For environmental layers, the Grinnellian and Eltonian definitions provide a useful framework to explain the difference between potential and actual niches. The Grinnellian niche is most analogous to the potential niche described above, and is thus determined by coarse grained abiotic predictors, while the Eltonian niche is closest to the actual or realized niche, and thus is modeled using fine-grained biotic variables. This generalization fits well for most examples, though the effect of micro-climates on plant species is an important exception, as abiotic variables (e.g. light, water, temperature) can act at very fine scales to determine plant distributions. In truth, the results from SDMs fall somewhere along a gradient between actual and potential distribution (Jiménez-Valverde et al., 2008). The terms indirect (e.g. latitude, elevation), direct (e.g. temperature), and resource (e.g. food source, soil nutrients) are also used to describe the predictors used in SDMs (M. P. Austin, 2002).

When developing SDMs it is important to keep these distinctions and their frameworks in mind. The most commonly available response data are presence-only, and often the only available predictor variables (especially in the tropics), are coarse grained abiotic variables determined using remotely sensed products (e.g. MODIS-derived land cover) or interpolated from sparsely located weather stations (e.g. BioClim). In this situation, the SDM is most likely producing a distribution close to the potential, as opposed to the actual. This has important implications for how the results of the SDM are applied for conservation. For example, if the potential distribution is modeled, not all areas predicted as suitable will contain the species, and the areas that are occupied might receive higher priority for some conservation action. Or, the unoccupied potential areas could be targeted as possible sites for species relocations. One difficulty is determining which parts of the modeled potential distribution are occupied, which can be determined using expert opinion or the available data (including that used to create the model).

Another complication is how to identify species observations that are from vagrant or dispersing individuals. If these observations are included in the model, the model results may be biased and unable to identify the true core habitat and niche of the species. One strategy to deal with this issue is to use expert opinion to remove these records from the data. Another option is to determine if the individuals are breeding. Theoretically, if the species is not breeding in that area, whether due to low suitability or the allee effect (inability to find mates due to low density (Stephens, Sutherland, & Freckleton, 1999)), then the area represents a sink that should not be considered part of the niche. This type of information could be gleaned from camera trap photos or direct observations, or more reliably from captured individuals. Other detailed health data from captured individuals may also help determine if the species has a healthy population in that area. If observations could be tagged as breeding/non-breeding, dispersing, and sink/source populations, this

information could be used in modeling context to model the niche of the species more accurately, as well as its dispersal capability. Occupancy models have already been developed to include breeding versus nonbreeding observations (L. L. Bailey, MacKenzie, & Nichols, 2014; MacKenzie, Nichols, Seamans, & Gutiérrez, 2009).

1.4 POINT PROCESS MODELS

New and novel methods are frequently introduced to address the challenges outlined above. Many of these will be passed over as too complex or not general enough, but some have the potential to be picked up and help move the field of SDM forward. One with particularly interesting potential is the framework introduced as a solution to the “presence-only problem” in SDMs (Warton & Shepherd, 2010). These spatial point process models (PPMs) have been used in several different applications, including astronomy, forestry, medicine, and ecology (Illian, Penttinen, Stoyan, & Stoyan, 2008). They were more recently used in spatial capture-recapture (SCR) models used to estimate species abundance (Borchers & Efford, 2008; Chandler & Andrew Royle, 2013; Chandler & Clark, 2014; M. Efford, 2004; Gardner, Reppucci, Lucherini, & Royle, 2010; J. A. Royle, Chandler, Sollmann, & Gardner, 2013), and even occupancy models based on point counts (J. Andrew Royle, 2004).

The property estimated by the Poisson point process is intensity, or a count of the object of interest per unit area. As applied to SDMs, the object of interest is individuals of a species, and thus PPMs can be used to estimate abundance (i.e. intensity), or converted to probability of occurrence (Dorazio, 2014a). PPMs are statistically equivalent to the maximum entropy algorithm implemented through MAXENT (Renner & Warton, 2013) and logistic regression that utilizes pseudoabsences (Warton & Shepherd, 2010), yet there are several advantages of PPMs: a clear framework for choosing quadrature points

(equivalent to pseudoabsences), well developed tools for assessing model fit, ability to handle spatial dependence, consistency of estimates in the face of changing units or spatial resolution, and clear avenues for improvements, i.e. marked point processes to incorporate species interactions, etc. (Renner et al., 2015). These advantages suggest PPMs may play an important role in the improvement of SDMs methods.

However, PPMs by themselves are insufficient as SDMs because they do not account for imperfect detectability (Kéry, 2011). For this reason, what is modeled is the intensity of species observations, not the actual intensity of the species distribution. A solution to this issue is to implement an occupancy-like framework with a PPM, as developed in Integrated SDMS by Dorazio (2014a) and Koshkina et al. (2017). Using this Integrated framework, presence-only data and presence-absence data can also be analyzed together. In this way, even a small amount of presence-absence data can be used to leverage information in presence-only data, and thus provide models with broad spatial coverage and unbiased coefficients. Since this model is based on a PPM, it can be used to estimate population size, as well as absolute probability of presence.

1.5 FRONTIERS IN SDMS

The research around SDMs is extremely active, and so it can be difficult to pinpoint specific avenues of development that will persist moving forward. Up to this point, I have covered the background of SDMs, their methodological and theoretical underpinnings, along with some challenges that apply to both realms (methods and theory), and some current solutions. Inevitably, the field will continue to address some of the challenges and shortcomings already mentioned above, as there is still much room for improvement. In general though, the future of SDMs will be driven by four main efforts: 1) creating more realistic models that incorporate ecological knowledge (M. Austin, 2007; Higgins, O'Hara,

& Römermann, 2012), 2) unifying disparate theories and multiple methods that address the same (or related) questions (McDonald, 2013; McGill, 2010), 3) working with big data to develop not only the tools that can handle large quantities of data, but also how to use these new datasets to ask questions that simply were not possible before (Stephens-Davidowitz, 2017), and 4) open access and democratization of research (e.g. citizen science, Map of Life SDM interface Jetz, McPherson, & Guralnick (2012)).

These developments are not unique to SDMs, as they apply across ecology, and even scientific research in general. Within each of these broader efforts fall more specific research agendas, some of which I have yet to discuss in detail, including dynamic models (i.e. spatio-temporally explicit), the incorporation of movement data and theory (J. Miller & Holloway, 2015), formalizing connections between SDMs and other relevant ecological models which are often used in an iterative fashion (e.g. population viability analysis, connectivity studies) (Keith et al., 2008; Zeller & Rabinowitz, 2011), and using models which can combine different data types (Dorazio, 2014a; Koshkina et al., 2017).

In the previous section, I introduced PPMs as a solution to many of the current challenges facing SDMs, yet they are also a useful framework to contribute to the frontiers in SDM research. For example, point process models can be extended to incorporate spatio-temporal patterns (Diggle, 2013). Thus far, some studies have sought to include the temporal dimension in their models through more ad hoc methods, including using weather variables matched to the time of species observation, as opposed to climate averaged over decades (Reside, Vanderwal, Kutt, & Perkins, 2010), creating separate SDMs for different months (Gschweng, Kalko, Berthold, Fiedler, & Fahr, 2012), and averaging models trained on different spatial and temporal regions (Fink et al., 2010). Each of these three studies uses interesting ideas which may be useful to include in the development of spatio-temporal models, yet PPMs, and hierarchical models as well, appear to provide more model-based

approaches to tackling this challenge (Cressie & Wikle, 2015; Hefley & Hooten, 2016; Snow, Jarzyna, & VerCauteren, 2017).

Another advantage of PPMs is the ability to incorporate spatial dependence (i.e. autocorrelation) using several types of point or area interaction processes (Renner et al., 2015). Thus far in SDMs, the standard approach to account for spatial autocorrelation is through an autoregressive model, which uses a spatial lag of the response in a specified neighborhood around each observation as an additional variable in the model (Cruse et al., 2012; F. Dormann et al., 2007; J. Miller & Franklin, 2002; J. Miller, Franklin, & Aspinall, 2007), or a bias grid when using Maxent (Clements et al., 2012; Elith, Jane, Michael, & Steven, 2010). PPMs offer multiple interaction processes that allow the models to incorporate attraction or repulsion between points, and the tools to determine the appropriate settings (Renner et al., 2015; Renner & Warton, 2013). These interaction processes can be used to account for both intra- and inter-species interactions. Thus far, very few studies exist which attempt to include species interactions in an SDM, and those that do simply add the presence or absence of a competitor as an additional variable in the model (Anderson, Peterson, & Gómez-Laverde, 2002; Leathwick & Austin, 2001). Though these studies showed that the predictive power of SDMs are significantly increased by incorporating species interactions in this way, there is no way to prove that this is due to actual competition, and not just the additional information about physical conditions that may be correlated with the presence-absence of the competitor.

Hierarchical models (HMs) appear to be a flexible way to incorporate both spatial dependence (Gelfand et al., 2006; Latimer, Wu, Gelfand, & Silander, 2006) and a spatially biased sample (Dorazio, 2014a; Koshkina et al., 2017). Though HMs are also the basis of occupancy models, which have become quite common, it is interesting to notice the lack of studies that use HMs to address the issues of spatial dependence and biased sampling. It

is possible these methods have been underutilized because they are relatively more difficult to use (than Maxent), or because they are typically based on a Bayesian approach (less than 6% of SDM or ENM papers returned from a Web of Science search used Bayesian methods). A new package introduced to R called hSDM (i.e. hierarchical SDMs) may help advance the use of these models. It is also important to note that the two frameworks can be used together, PPMs and HMs. In fact, there appears to be some support for SDMs to move in this direction (Hefley & Hooten, 2016). However, there is healthy skepticism about the excitement behind PPMs (M. Kery & Royle, 2015), and HMs appear to be the more general solution.

Species conservation is a high-stakes endeavor. When a species is lost, it is lost forever. That is why it is so important for SDMs used in conservation applications to use the most appropriate methods whenever possible. The frameworks proposed above are designed to address the issues with real world data and answer flexible questions. The early surge in SDMs was related to growth in computing and access to data. Now the use of HMs and a Bayesian approach (which enables more complex models) could be viewed as a second stage of growth in SDMs. Where widely-recognized deficiencies are addressed, producing more realistic models and predictions, that allow conservation managers to make better decisions.

1.6 RESEARCH OBJECTIVES

The main research objective for this dissertation is to apply a new SDM model, the Integrated SDM (Dorazio 2014a, Koshkina et al., 2017), to a real species, Baird's tapir (*Tapirus bairdii*). Thus far, this model has been applied to real species and environments in a very limited number of cases (Fletcher et al. 2016; Koshkina et al., 2017). More specifically, this research will address the following objectives in the research chapters that

follow: 1) identify how the ISDM performs compared to more traditional SDM methods and expert opinion, 2) investigate the sensitivity of the ISDM to different user decisions, and 3) use the results from the ISDM to inform a Range-Wide Conservation Plan (RWCP) for Baird's tapir.

1.6.1 Baird's tapir as a case study species

Baird's tapir is a species of significant conservation interest due to its globally endangered status, genetic uniqueness, role in ecosystem function, and increasing threat to its habitat (Brooks, Bodmer, & Matola, 1997; Garcia et al., 2016; Isaac, Turvey, Collen, Waterman, & Baillie, 2007). Tapir play an important ecological role as ecosystem engineers, primarily through seed dispersal and extensive selective browsing. Baird's tapir are the largest terrestrial mammals in the Neotropics and are therefore the sole effective dispersers of most large seeds, especially over long distances (O'Farrill, Galetti, & Campos-Arceiz, 2013). The loss of megaherbivores like the Tapir can lead to cascading detrimental effects on other species, ecological processes, and overall ecosystem function (Ripple et al., 2015). For this reason, Baird's tapir could be considered a keystone species in the environments in which it occurs (Mills, Soulé, & Doak, 1993). In addition, as a wide-ranging, large and charismatic megafauna, Tapir have the potential to act as both an umbrella and a flagship species for conservation (Caro & O'Doherty, 1999; Cove et al., 2014).

Baird's tapir (*Tapirus bairdii*) is an ideal candidate to test new SDM methods applied to conservation. Though the species is relatively rare with a low detectability rate, it is easily identifiable from camera trap photos and other tracks or signs left by the species. Also, Tapir have relatively small and well-defined home ranges, limited dispersal movements, and distinct habitat preferences. Other large mammal species, such as top

predators like the Jaguar (*Panthera onca*), have much larger home ranges and dispersal abilities, leading to observations in a wide range of habitats, which can cause poor model fit (Franklin, Wejnert, Hathaway, Rochester, & Fisher, 2009). Tapirs also live mostly solitary lives, as opposed to another endangered species which occupies similar habitats, the White-lipped Peccary (*Tayassu pecari*) (Brooks, Bodmer, & Matola, 1997; Oliver, 1993). Species that form large groups, such as peccaries, may require models which account for this grouping behavior (i.e. the ability to account for spatial autocorrelation).

Spatial mark-resight (SMR) and spatial capture-recapture (SCR) models which require individual identification can be applied to camera trap data for some species with unique markings (e.g. stripes or spots), including Jaguar, Tigers, Ocelots, and some other cat species (McClintock, White, Antolin, & Tripp, 2009; Royle, Chandler, Sollmann, & Gardner, 2013). Though these models have been applied to species without such markings, including Tapir and Puma, many researchers question the ability to accurately identify individuals (Oliveira-Santos, Zucco, Antunes, & Crawshaw, 2010/7). This makes the ISDM a good choice for Baird's tapir, as SMR or SCR have the potential to produce more robust results, yet are difficult to apply to this species. It is possible to ID individuals by collecting DNA or capturing and tagging individuals, but collecting this data is costly. Finally, there are large coverage gaps in high quality (i.e. site-occupancy) data for Baird's tapir in Honduras and Guatemala, making the ISDM a good choice to fill in these gaps with lower quality (i.e. presence-only) data.

1.6.2 Developing an international collaboration

The International Union for the Conservation of Nature (IUCN) is a membership union composed of government and non-governmental institutions working together to provide the knowledge and tools necessary to promote human progress, economic

development and nature conservation (www.iucn.org). The Species Survival Commission, one of the six commissions that make up the IUCN, is a science-based network of volunteer experts, specialist groups, Red List Authorities, and Task Forces (<https://www.iucn.org/ssc-groups>). This network provides an excellent infrastructure to enable the collaboration which is necessary to facilitate research and conservation planning for a species like Baird's tapir.

In the early stages of this research, I contacted the Tapir Specialist Group (TSG) to introduce myself and get connected with Tapir experts. This specialist group is especially collaborative and welcoming to new members (including non-PhD holders). I was also fortunate to contact them at a time when they were already planning to conduct the first range-wide SDM for Baird's tapir. I volunteered to help with this effort, and it was not long before I was leading the analysis. Within a year, this research was published in the journal of the TSG, *Tapir Conservation* (Schank et al., 2015). I grew the collaboration that started with this project to include additional researchers and data so that we could test the Integrated SDM on Baird's tapir, assess the results, and conduct a Range-wide Conservation Plan for the species.

Large collaborations which involve data sharing between many different researchers and organizations are necessary to conduct range-wide assessments for widely-distributed species like Tapir. To facilitate this type of collaboration, it is necessary to give proper attribution to data providers, and extend invitations for co-authorship in any resulting publications. This was enforced by a written data agreement developed for the initial research published in *Tapir Conservation*. Multi-author publications are now the dominant paradigm in science (Barlow et al., 2018), but there is still some concern for how to handle these situations when such collaboration is used in a dissertation. I hope this dissertation can serve as an example for future researchers to follow on similar projects.

CHAPTER 2: Using a novel model approach to assess the distribution and conservation status of the endangered Baird's tapir¹

2.1 INTRODUCTION

Determining current species distributions and assessing change over time are necessary for conservation management aimed at reducing species loss in ecosystems. Species Distribution Models (SDMs) provide continuous spatial predictions about species habitat preferences, and thus are useful tools for conservation management (Franklin, 2010a). However, building useful models for rare species can be challenging for many reasons, including that data on rare species are typically scarce and vary in quality and type (Aitken, Roberts, & Shultz, 2007; Wisz et al., 2008). The fact that species most in need of conservation attention are often the most difficult to model has been labeled the “rare species modelling paradox” (Lomba et al., 2010).

In general, SDMs use one of two types of data as the response (i.e. dependent) variable: presence-only (PO) or presence-absence (PA)². PO data contain only information about where a species has been observed, while PA data also provide information about where the species has not been found despite survey efforts.

Presence-only data are generally more available than presence-absence data, but the modelling options appropriate for use with them are limited and can lead to overestimates of distribution. Despite this, modelling methods for presence-only data have

¹ Version of Record (VoR) of this manuscript has been published and available in *Diversity and Distributions* (2017) <http://onlinelibrary.wiley.com/doi/10.1111/ddi.12631/abstract>. Some amendments to the VoR have been made. CJS was the lead author on this article, developing the research question, methodology, analysis, and results. As such there is no conflict for the content of this manuscript to constitute part of this dissertation.

² Throughout this chapter I refer to presence-absence (PA) data. In later chapters, this data type is referred to as site-occupancy (SO) data. For this research, these data consist of detection histories from camera traps.

advanced, albeit slowly. For over a decade, the most common method across all SDMs has been Maxent (Phillips et al., 2004; Vaz et al., 2015), likely due to its more accurate predictions compared to other frameworks (Elith et al., 2006). Despite its heavy use, Maxent has been criticized for several reasons. These criticisms include its use of vague indices to determine probability of occurrence (Royle, Chandler, Yackulic, & Nichols, 2012), lack of accounting for spatial sampling bias, which is common in presence-only data (Yackulic et al., 2013b), and difficulty in interpreting results (Renner & Warton, 2013).

Other methods developed to address deficiencies in SDMs include occupancy models (MacKenzie et al., 2002) and point process models (PPMs) (Warton & Shepherd, 2010). Occupancy models account for imperfect detection, or the fact that a species is not always detected despite its presence. Failing to account for imperfect detection can lead to biased parameter estimates (Dorazio, 2012; Lahoz-Monfort et al., 2014). In occupancy modeling, if detection probability is affected by the same set of covariates that affect occurrence probability, bias in the model is inevitable (Royle et al., 2012). To solve this issue, these models require repeated observations of either presence-absence or point-count (PC) data to estimate and account for detectability (Kéry, Guillera-Arroita, & Lahoz-Monfort, 2013).

Some researchers have attempted to work around the weaknesses of using PO data in SDMs by including pseudo-absences, or artificially compiled absences, in models. Pseudo-absences are used with methods that require PA data when only PO data are available (Zaniewski, A. Elizabeth, Anthony, & Overton, 2002). Maxent uses a similar construct called background points. Regardless of the term, there are no clear recommendations on the number of points to use or how to distribute those points across the area of inference (Barbet-Massin et al., 2012). PPMs solve the pseudo-absence problem

through a reformulation of the basic model structure for SDMs. In PPMs, pseudo-absences are viewed as equivalent to quadrature points used to estimate integrals. Following this conceptualization, the number of background points is increased until the likelihood stabilizes (Warton & Shepherd, 2010).

An additional advantage of PPMs is that the response variable modeled is ‘intensity’, or individuals per unit area. This is different from Maxent, which models occurrence on a per-pixel basis, and thus is dependent on the scale chosen for the analysis (Renner & Warton, 2013). The model-estimated intensity can also be summed across spatial units (e.g. a specific protected area) to obtain an estimate of population size in that area, a useful tool for conservation management.

For most rare species, PO data remain the most readily available because they are easiest to collect (Engler, Guisan, & Rechsteiner, 2004a). However, over the past decade PA data have become increasingly common due to analytical developments that have promoted their collection (MacKenzie et al., 2002) and the rapid expansion of camera-trapping, which enables this data collection (Lesmeister, Nielsen, Schaubert, & Hellgren, 2015). Given that data on rare species are often scarce by nature, methods to integrate both PO and PA data on rare species into the same model have the potential to improve the capacity to model rare species effectively. A new model combines the two frameworks of occupancy models and point process models, allowing for the use of both PO and PA data (Dorazio, 2014b). Thus far, there have been no real-world applications of this model (Fletcher, McCleery, Greene, & Tye, 2016), leaving conservation biologists unable to assess the full potential of this integrative approach of combining PO and PA data.

2.2 METHODS

2.2.1 Study Area

The study area for this research ranges from Veracruz and Oaxaca states in Southern Mexico to the Department of Chocó in Northern Colombia (approximate coordinates: 22°N to 3.6°S, 98° to 76°W), excluding islands. All analysis was conducted at a spatial resolution of one kilometer, in a global equal area projection.

2.2.2 Model Formulation

Four modeling frameworks were tested in this analysis, including 1) Maxent (Phillips et al., 2004), 2) a presence-only (PO) SDM based on a point process model (PPM), 3) a presence-absence SDM (PA), also based on a PPM, and 4) the Integrated SDM framework that combines both 2 and 3 (Dorazio, 2014b). Imperfect detectability is incorporated in both the PO (2) and PA (3) SDMs. In the PO model, probability of detection is modeled as a thinning of the point process describing the species distribution, using a Bernoulli distribution (Dorazio, 2014b). In the PA model, detectability is estimated as in an occupancy model using a zero-inflated formulation (J. A. Royle & Dorazio, 2008). Models were fit using the ‘dismo’ package (Hijmans, Phillips, Leathwick, & Elith, 2016) and custom code developed in R (R Development Core Team, 2016).

Nineteen candidate models were formulated to test different hypotheses about the factors that affect the distribution of Baird’s tapir (Table 2.1). All four model frameworks were fit using these candidate models (except the null model, model 0, for Maxent, as there is no way to run the software without predictor variables). To facilitate the creation of the candidate models, predictor variables were grouped into categories related to influential processes including climate, land cover and protected status, human pressure, slope, and sampling variables (Table 2.2). All continuous variables were scaled to have a mean of

zero and standard deviation of one, thus the magnitudes of the coefficients are comparable (except distance to/within protected areas, which was not centered, to preserve negative values as within protected areas, and positive values outside of these areas).

| Model | Climate | Landcover | Human Pressure | Slope | Sampling |
|--------------|----------------|------------------|-----------------------|--------------|-----------------|
| Model0 | | | | | |
| Model1 | y | | | | |
| Model2 | | y | | | |
| Model3 | | | y | | |
| Model4 | y | y | | | |
| Model5 | y | | y | | |
| Model6 | | y | y | | |
| Model7 | y | y | y | | |
| Model8 | | y | y | y | |
| Model9 | y | y | y | y | |
| Model10 | y | | | | y |
| Model11 | | y | | | y |
| Model12 | | | y | | y |
| Model13 | y | y | | | y |
| Model14 | y | | y | | y |
| Model15 | | y | y | | y |
| Model16 | y | y | y | | y |
| Model17 | | y | y | y | y |
| Model18 | y | y | y | y | y |

Table 2.1: Categories of variables included in the candidate models.

2.2.3 Description of Modeling Frameworks

Maxent is a popular SDM software that minimizes the entropy between two probability distributions based on a vector, z , of environmental covariates used in modeling (Elith et al., 2011). The first distribution, $f_1(z)$, is based on presence-only (PO) locations, while the second, $f(z)$, is based on a random sample of the background environment. The goal of most SDMs is to estimate the probability of presence, or $\Pr(y = 1 | z)$, where $y = 1$

denotes presence ($y = 0$ denotes absence). However, without absence data, this property can only be estimated up to a constant. The missing piece of information is $\Pr(y = 1)$, or the prevalence of the species in the study area. By default, the Maxent software assumes prevalence is 0.5 to create a logistic output which ranges from 0 to 1. I adjusted this setting to 0.2 to match the prevalence in the presence-absence test data. Maxent settings were also adjusted to include only linear and quadratic terms (when included in other models). Default settings in Maxent include more complex parameters, which can lead to model overfitting, and would not provide a reasonable comparison to the other frameworks used in this research.

| Category | Variables | Citation |
|-----------------|---|---|
| Climate | temp_seasonality*, precip_seasonality*, max_temp_warmest_month*, annual_precip* | (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) |
| Landcover | forestcover2000, distancePA, EVI* | (Google Earth Engine Team, 2015; Hansen et al., 2013; UNEP-WCMC, 2014) |
| Human Pressure | forestloss_focal, road_length_focal | (Eugster & Schlesinger, 2010; Hansen et al., 2013) |
| Terrain | Slope | (Hijmans et al., 2005) |
| PO Sampling | forest, protected, road_distance* | (Arino et al., 2012; Eugster & Schlesinger, 2010; UNEP-WCMC, 2014) |
| PA Sampling | forestcover2000, distancePA, road_distance* , slope | (Eugster & Schlesinger, 2010; Hansen et al., 2013; Hijmans et al., 2005; UNEP-WCMC, 2014) |

Table 2.2: Predictor variables in each category (* indicates quadratic term used in the model).

The Poisson point process model (PPM) is the foundation of the SDM framework (Dorazio, 2014b). In this model, $\lambda(s)$ is the expected density (number of individuals per unit area) of individuals at location s , for a Poisson point process. In the context of the SDM, $\lambda(s)$ is formulated as a log-linear function of unknown parameters and location-specific regressors $x(s)$ (i.e. $\log(\lambda(s)) = \beta_0 + \beta'x(s)$). The presence-only (PO) model formulated incorporates spatial bias in PO data through an independent thinning of the point process. This thinned point process, is the product of the original point process and $p_{po}(s)$, the probability that the site is surveyed, and the species is detected. $p_{po}(s)$ is formulated as a logistic function of unknown parameters and location-specific regressors $w_{po}(s)$:

$$\text{logit}(p_{po}(s)) = \alpha_{0,po} + \alpha'_{po} \cdot w_{po}(s)$$

I adjusted the planned survey model, which used point count data, to accommodate detection/non-detection) data (Dorazio, 2014b; Koshkina et al., 2017). The model used here follows an occupancy framework based on a zero-inflated binomial distribution. The presence or absence (i.e. occupancy) of the species at a site, i , follows a Bernoulli distribution. In this case, the detection histories at each site, y_i , have additional non-detections (i.e. zeros) due to imperfect detectability, the fact that an individual may go undetected even when present. This relationship is modeled as a Binomial distribution with J trials, and the probability of success (i.e. species detection) equal to the product of z_i (the occupancy state, $z_i = I(N_i > 0)$) and p_{pa} , the probability of detection at the site. Occupancy (ϕ) and intensity (λ) are linked following the equation below (Koshkina et al., 2017),

$$\phi = \Pr(N(C_i) > 0) = 1 - \exp\left(-\int_{C_i} \lambda(s) \cdot ds\right)$$

As with detectability in the PO model, $p_{pa}(s)$ is formulated as a logistic function of unknown parameters and location-specific regressors $w_{pa}(s)$:

$$\text{logit}(p_{pa}(s)) = \alpha_{0,pa} + \alpha'_{pa} \cdot w_{pa}(s)$$

Though their models of detectability differ, the PO and PA frameworks share the same SDM model based on a Poisson point process. In the Integrated SDM, the PO and PA models are estimated simultaneously, such that one set of parameters for the SDM is created (i.e. the β 's), while separate detectability parameters are estimated (i.e. the α 's) for the PO and PA models.

2.2.4 Presence Data

Presence data for Baird's tapir were compiled from several sources (Appendix 1), including planned surveys conducted using camera traps (presence-absence data), and collections of opportunistic observations or published sightings (presence-only data), which may originate from cameras, track, or signs of the animal, or direct observations of the animal. Though PA data are generally more clustered across the study area than the PO data (Figure 2.2), they do provide similar coverage of environmental gradients (min-mean-max percent overlap between background/PO: 0.36-0.63-0.80, background/PA:0.46-0.60-0.75, PO/PA:0.50-0.67-0.90, (Pastore, 2017)).

I used presence-absence data collected by camera traps to construct 100-day long detection histories at each site, establishing ten sampling occasions that span ten days each (tapir appear to cycle through their home range every 10-12 days; (Jordan, 2015)). Cameras at some sites malfunctioned or were removed before the 100-day sampling period was complete. In this case, I used data only from complete sampling 10-day occasions.

Due to the nature of modeling the entire distribution of a wide-ranging species like Baird's tapir, data must be combined from several sources which are often collected during different years. In this case, the data spans primarily the period from 2000-2015 (with few exceptions), thus the detection histories across sites are not from the same 100-day span.

For this reason, occupancy and density estimates represent the species presence during that time span (i.e. over 15 years).

2.2.5 Spatial Subsampling

A random spatial subsampling procedure was performed to thin the presence data prior to model fitting. The algorithm starts by randomly choosing one observation point and removing any other observation points within a given radius. The chosen observation is added to the subset, and the steps repeated until no observations are left in the original data. The effect of this procedure is to enforce a minimum distance between sampling points. A similar type of subsampling is sometimes used to remove survey bias in observation data (Beck et al., 2014), however, this grid-based approach can lead to samples that remain close in space, if they fall just across a boundary in an adjacent grid cell.

This subsampling process was used to maintain independence of observations. In occupancy models, the population at each site is assumed to be constant, with no immigration or emigration (MacKenzie et al., 2006). This assumption is known as closure, violations of which can lead to biased parameter estimates. For both types of data (PO and PA), a minimum distance of five kilometers was selected to avoid duplicate observations of the same individual. Data on home range sizes vary from 1.25 km² reported in Costa Rica (Foerster & Vaughan, 2002) to 8-10 km² in Nicaragua (C. Jordan personal observation, 2016), and 23.9 km² in Mexico (Reyna-Hurtado, Sanvicente-López, Pérez-Flores, Carrillo-Reyna, & Calmé, 2016a). Thus the 5-km spacing should ensure that detections are from different individuals given the tapir's home range size.

2.2.6 Model averaging

Model fit was repeated 100 times for each combination of candidate model and model framework to capture the variation in the random spatial subsampling. Parameters

were averaged across candidate models for each framework and point sample using Akaike weights (Symonds & Moussalli, 2011),

$$\bar{\beta} = \frac{\sum_{i=1}^R w_i \cdot \hat{\beta}_i}{\sum_{i=1}^R w_i}$$

where β is the parameter estimate for a given model i , and w is the Akaike weight of the model. The parameters were then averaged across these replicates to create one prediction and set of parameter estimates for each model framework (Table 2.2). In Maxent, calculation of AIC is not straightforward, thus for this framework I averaged across model iterations for candidate Model 9 (full model without sampling covariates). There are questions about the validity of model averaging using Akaike weights (Cade, 2015), however, for this research only the PO and PA models are affected, since this method is not used with the Maxent results, and model 18 has all the support in the Integrated framework (Table 2.2).

2.2.7 Identification of habitat cores

Using the predicted intensity from the Integrated model, I calculated the Getis-Ord G_i^* (i.e. local G) spatial statistic (Getis & Ord, 1992) to identify clustered areas both above (i.e. hot spots) and below (i.e. cold spots) the global mean (Nelson & Boots, 2008). The estimated intensity was then summed within each statistically significant ($\alpha = 0.05$) hotspot, and those with values above 200 were selected as habitat cores. It is estimated that 200 adult individuals of *Tapirus terrestris* were required to constitute a minimum viable population in the Atlantic Forest of Brazil (Medici & Desbiez, 2012). I assumed the same to be true for *Tapirus bairdii* although if inbreeding and genetic variability were considered, this number would likely need to be considerably higher, especially given that Baird's tapirs are known to have very low levels of heterozygosity and allelic diversity (Norton & Ashley, 2004).

2.3 RESULTS

2.3.1 Coefficient Estimates

The reciprocal of the condition number, the ratio of the smallest to the largest eigenvalues in the Fisher information matrix, can be used to determine if the parameters of the SDM are identifiable (Dorazio, 2014b). Following these criteria, coefficient standard errors were almost never estimated for the PO models, and thus this framework is not included in most of the discussion. This trouble with identifiability could be due to very low detectability in the PO data. For comparison, PO models were estimated with standard errors in less than 1% of models, while PA models were able to estimate these measures 100% of the time, and the Integrated model 80% of the time.

Slope was the only variable in the PA model that had a 95% confidence interval that did not overlap zero (Table 2.3). It is likely that establishing important relationships with environmental variables was difficult in the PA model because of reduced sample size (the PA model does not include any background information, while the other three models do). This was also evident in the Akaike weights (Table 2.4), as there was no clear top/best model. The different random samples that resulted from the spatial subsampling of the presence data appear to have a strong effect on which model has the best fit. In contrast, for the Integrated framework, the full model (model 18) had nearly all the support in all iterations.

In the Integrated framework, the only climate variables with 95% confidence intervals that did not overlap zero were annual precipitation (positive), and temperature seasonality (negative), both also had non-linear relationships (i.e. quadratic terms important). Meanwhile, nearly all non-climate variables had 95% confidence intervals that did not overlap zero. The modelled relationships for these variables were mostly as expected, showing a positive relationship between Baird's tapir occurrence and forest

cover, and negative relationships with roads and slope (Table 2.3). Among all of these, the effect of forest cover had the largest magnitude, followed by roads. This is not surprising, as deforestation and hunting are widely recognized as the two largest threats to the species (García et al., 2016; Medici et al., 2005). Interestingly, Enhanced Vegetation Index (EVI), had a negative and linear relationship with Baird's tapir occurrence in the model.

There were also several significant relationships among the sampling variables in the Integrated framework. With these, a positive relationship existed with protected status in the PO data, possibly indicating the positive sampling bias towards these areas. Also, with the PO data, there was a significant negative and non-linear effect with distance to roads, which could indicate the increased difficulty in sampling far from road networks. In the PA data, none of the sampling variables had a 95% confidence interval that did not overlap zero. It is important to note that the sampling variables in the PO data are meant mostly to account for sampling bias, while those in the PA model should reveal more about what affects the detectability of the species. For this reason, I used slightly different sampling variables for the two data sets.

It is also interesting to note that distance to/within a protected area was significant in the Integrated models without sampling variables ($\beta = -0.51$, $SE = 0.0806$, $p < 0.001$), but was not significant in the Integrated models with sampling variables ($\beta = -0.0379$, $SE = 0.118$, $p = 0.761$). Thus, accounting for survey bias by including variables in the detection process reduced the effect of a variable that was expected to have a significant influence on Baird's tapir distribution.

Maxent estimates parameters that are not directly comparable to the coefficients discussed above, but that can provide some context for comparison (Table 2.5). Forest cover, distance to/within protected areas, and presence of roads were the three variables with the most importance and contribution to Maxent (in model 9). This was like the results

| Coefficient | PO | PA | Integrated |
|-----------------------------|--------------------|-----------------|-------------------|
| beta0 | 0.102 (0.55) | -1.99 (1.6) | -1.22 (0.27)* |
| x.temp_seasonality | 0.0781 (0.00586)* | 0.0225 (0.534) | 0.106 (0.112) |
| x.precip_seasonality | -0.0541 (0.00616)* | -0.241 (0.553) | -0.0556 (0.118) |
| x.max_temp_warmest_month | 0.191 (0.00639)* | 0.362 (0.73) | 0.129 (0.131) |
| x.annual_precip | 0.404 (0.0078)* | -0.0488 (1.02) | 0.497 (0.152)* |
| x.temp_seasonality_sq | -0.254 (0.00472)* | -0.18 (0.343) | -0.284 (0.0905)* |
| x.precip_seasonality_sq | -0.144 (0.00531)* | 0.375 (0.349) | -0.115 (0.0957) |
| x.max_temp_warmest_month_sq | 0.0337 (0.0018)* | 0.436 (0.292) | 0.0548 (0.0278)* |
| x.annual_precip_sq | -0.198 (0.00442)* | -0.381 (0.452) | -0.297 (0.0881)* |
| x.forestcover2000 | 1.39 (0.0126)* | 1.63 (1.23) | 1.45 (0.18)* |
| x.distancePA | -0.147 (0.00841)* | 0.11 (0.359) | -0.0379 (0.118) |
| x.EVI | -0.348 (0.00955)* | -0.942 (0.729) | -0.426 (0.138)* |
| x.EVI_sq | 0.0423 (0.00448)* | 0.285 (0.334) | 0.0206 (0.0629) |
| x.forestloss_focal | -0.155 (0.00414)* | -0.0492 (0.156) | -0.167 (0.0561)* |
| x.road_length_focal | -0.837 (0.0128)* | -0.499 (0.329) | -0.691 (0.142)* |
| x.slope | -0.172 (0.000956)* | -0.834 (0.405)* | -0.358 (0.0904)* |
| alpha0.po | -1.03 (0.55) | 0 (0) | -7.58 (0.301)* |
| alpha0.pa | 0 (0) | -1.45 (0.622)* | -1.38 (0.644)* |
| w.po.forest | -1.79 (0.0147)* | 0 (0) | 0.0117 (0.208) |
| w.po.protected | 6.71 (0.0145)* | 0 (0) | 1.47 (0.203)* |
| w.po.road_distance | -3.79 (0.00934)* | 0 (0) | -0.752 (0.128)* |
| w.po.road_distance_sq | 0.673 (0.00208)* | 0 (0) | 0.123 (0.0301)* |
| w.pa.forestcover2000 | 0 (0) | 0.392 (0.453) | 0.24 (0.57) |
| w.pa.distancePA | 0 (0) | -0.0539 (0.182) | -0.0141 (0.273) |
| w.pa.road_distance | 0 (0) | -0.383 (0.248) | -0.665 (0.373) |
| w.pa.road_distance_sq | 0 (0) | 0.0558 (0.109) | 0.0768 (0.162) |
| w.pa.slope | 0 (0) | 0.146 (0.0762) | 0.204 (0.12) |

Table 2.3: Mean of model-averaged coefficient estimates for each model framework, with mean standard errors in parentheses. Symbols indicate estimates greater than 1.96 standard errors from zero (i.e. 95% confidence intervals). Standard error estimates for PO model are not provided.

with the Integrated model discussed above. Though it would have been possible to include the sampling variables in the Maxent model, there was no ability to differentiate between

variables that affect the sampling process versus those that affect the species distribution directly.

| Model | PO | PA | Integrated |
|--------------|----------------|-----------------|-------------------|
| Model0 | | 0.0166 (0.0687) | |
| Model1 | | 0.0178 (0.053) | |
| Model2 | | 0.017 (0.0542) | |
| Model3 | | | |
| Model4 | | 0.0292 (0.0738) | |
| Model5 | | 0.0179 (0.0478) | |
| Model6 | | | |
| Model7 | | 0.022 (0.0802) | |
| Model8 | | | |
| Model9 | 0.0354 (0.174) | 0.233 (0.306) | |
| Model10 | | 0.093 (0.172) | |
| Model11 | | 0.0361 (0.0768) | |
| Model12 | | 0.0444 (0.112) | |
| Model13 | | 0.0453 (0.0871) | |
| Model14 | | 0.0537 (0.0945) | |
| Model15 | 0.124 (0.331) | 0.021 (0.0453) | |
| Model16 | 0.359 (0.478) | 0.0294 (0.0515) | |
| Model17 | 0.0625 (0.242) | 0.0679 (0.17) | |
| Model18 | 0.472 (0.493) | 0.238 (0.295) | 0.998 (0.00421) |

Table 2.4: Mean of Akaike weights for each candidate model and model framework. Standard deviation of estimates in parentheses. Calculation of AIC is not straightforward for Maxent, thus we focus here on the other three frameworks. Cells left blank have no support (i.e. Akaike weight = 0).

2.3.2 Spatial Predictions

Spatial predictions for probability of occurrence were strikingly similar for two of the model frameworks: Maxent and Integrated (Figure 2.1). These predictions match well with the expected distribution for the species, though the probabilities may be too high across the landscape overall. Both the PO and the PA models had even higher probabilities

across the study area, and the PA model had probabilities close to one in regions that have not maintained Baird's tapir presence in recent history (e.g. Southern Guatemala and Northwestern Nicaragua). The PO model may have a similar pattern to the Integrated model, but high estimated probabilities covered too much of the distribution for the species.

| Variable | Contribution | Importance | Lambda |
|---------------------------|---------------------|-------------------|---------------|
| temp_seasonality | 0.5 | 1.38 | 0.727 |
| precip_seasonality | 0.0158 | 0.0268 | -0.0628 |
| max_temp_warmest_month | 0.0367 | 0.249 | 0.428 |
| annual_precip | 1.06 | 6.86 | 3.42 |
| temp_seasonality_sq | 4.45 | 3.67 | -1.45 |
| precip_seasonality_sq | 3.13 | 0.352 | -0.759 |
| max_temp_warmest_month_sq | 0.141 | 0.122 | 0.941 |
| annual_precip_sq | 1.2 | 6.75 | -6.16 |
| forestcover2000 | 59.7 | 53.3 | 4.05 |
| distancePA | 19.8 | 10.1 | -4.36 |
| EVI | 0.531 | 2.78 | -2.38 |
| EVI_sq | 0.00703 | 0.031 | 0.205 |
| forestloss_focal | 1.68 | 1.8 | -1.23 |
| road_length_focal | 4.74 | 7.96 | -15.1 |
| slope | 3.03 | 4.61 | -2.32 |

Table 2.5: Mean contribution, importance, and lambda values for each environmental variable. used in Maxent Model 9.

Inspecting the coefficient estimates (Table 2.3), the difference between the PO and Integrated model was largely due to the intercept (β_0), while the difference between the PA and Integrated models could be attributed to the difference in coefficient estimates for climate variables. The advantage that the Integrated model has over Maxent is the ability to estimate intensity (i.e. abundance), which I utilize below to identify viable habitat cores.

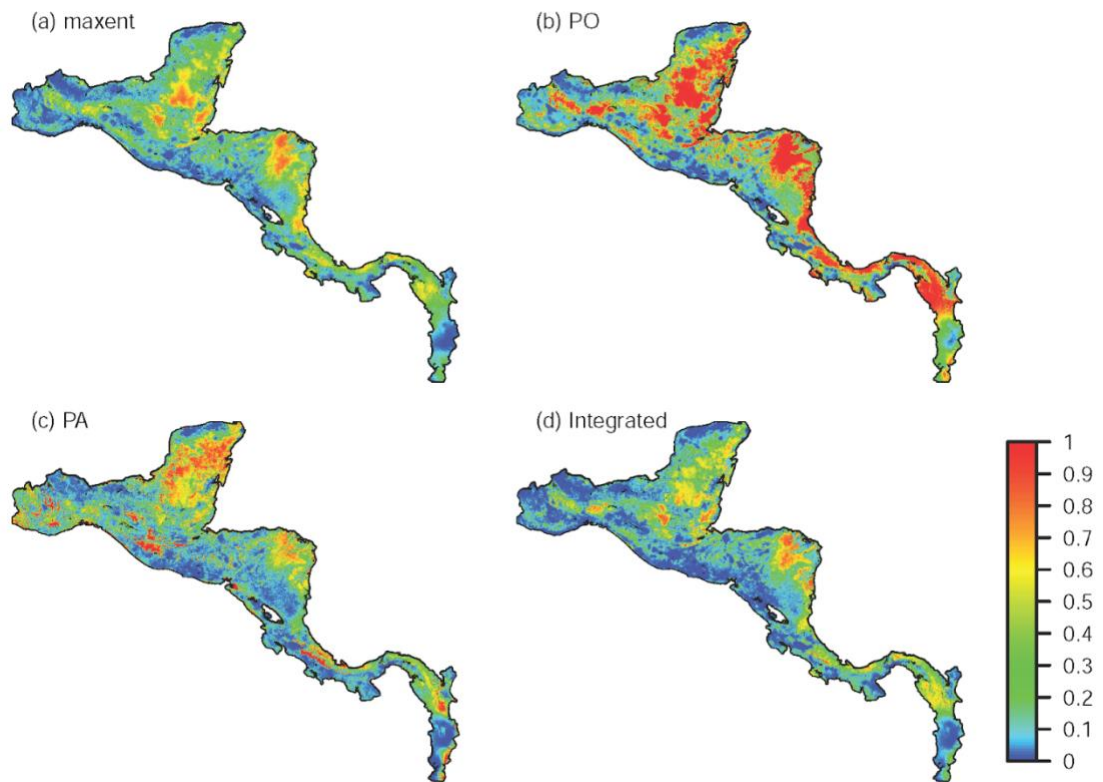


Figure 2.1: Predicted probability of presence for each model framework, (a) Maxent, (b) PO, (c) PA, and (d) Integrated.

2.3.3 Assessment of habitat cores and population estimates

A hotspot analysis using G_i^* and a circular neighborhood with radius of 10 km (Getis & Ord, 1992) identified 25 habitat cores in the Integrated model framework (Figure 2.2), with the two largest cores (the Yucatán Peninsula and the Honduras-Nicaragua border) comprising roughly 61% of core habitat for Baird's tapir. In general, the cores matched well with the expected distribution of the species. However, some small cores were predicted that have no tapir observations and likely do not harbor any individuals (e.g. La Encrucijada, México; Laguna de Términos, Mexico; North and West of Lacandona, Mexico; Pixvae, Panama). It is probable that tapirs historically occurred in these areas yet went locally extinct due to small core size and distance from larger cores

that act as population sources. There were also areas with confirmed tapir populations that did not appear in the prediction of viable cores (e.g. La Sepultura Biosphere Reserve and Frailescana Natural Protected Area, Mexico).

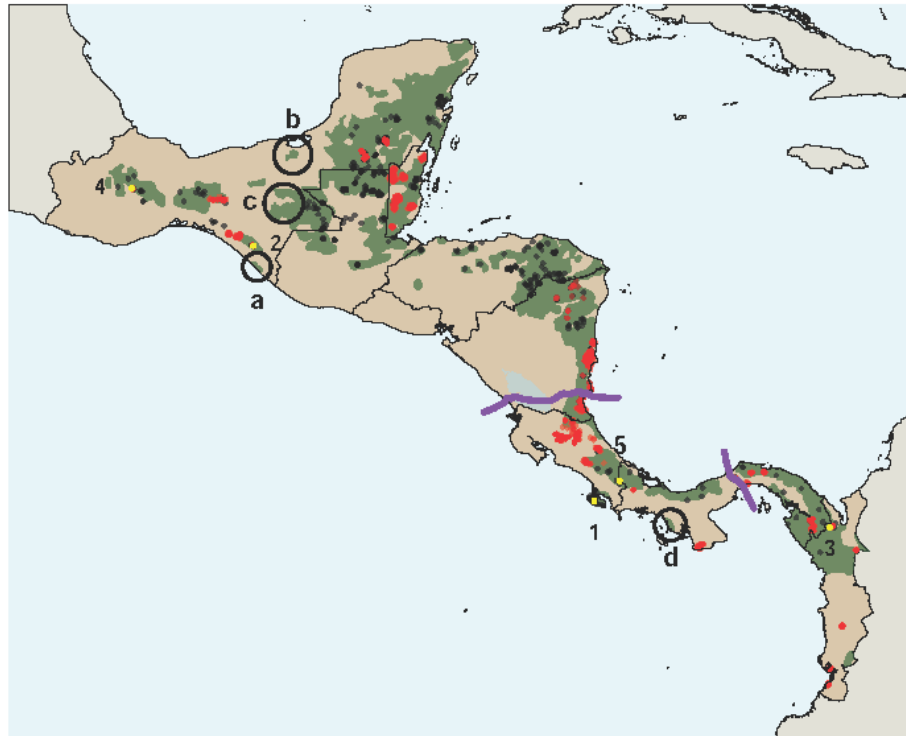


Figure 2.2: Map of habitat cores for the Integrated model. Purple lines represent the proposed Nicaragua canal path and the Panama Canal, black points are presence-only sample locations, red points are detection history sample locations, and the area shaded light brown is the study area. Yellow points are study areas for independent abundance estimates (1. Corcovado, 2. Triunfo, 3. Los Katíos, 4. Oaxaca, and 5. Talamanca). Black circles are cores overestimated by the model (a. Encrucijada, b. Laguna de Términos, c. N and W of Selva Lacandona and d. Pixvae).

According to model estimates, there are approximately 175,000 individual tapirs spread across these 25 core areas. This is more than an order of magnitude higher than

expert estimates for the species that suggest there may be fewer than 6,000 individuals left in the wild (García et al., 2016).

It is possible that model results would more closely approximate global experts expected state of the population with the addition of more proximate or more detailed environmental variables, such as differing management regimes among protected areas, more detailed habitat types, distance to villages, and improved hunting/poaching variables. Hunting has been reported as a widespread and substantial threat to the survival of Baird's tapirs due to their slow reproduction (Dunn, Estrada, & Smith, 2012; García et al., 2016; Jordan, Galeano, & Alonzo, 2014; Koster, 2006; McCann, Wheeler, Coles, & Bruford, 2012). Although hunting does not seem to drive tapirs to local extinction in all sites where the phenomenon has been studied, hunting could maintain tapir populations at extremely low densities. The human invasion of protected areas throughout significant portions of the Baird's tapir range has increased drastically in past decades, which has likely resulted in increased access for hunters in arguably all of Baird's tapir range countries both inside and outside of protected areas. Thus, substantially lower densities than would be expected in the absence of hunting might exist throughout the species' range. However, hunting data in a format compatible with PA and PO data on tapirs is not available on a range-wide basis for the species. Thus, to evaluate this, hunting data would need to be collected using a uniform methodology across the range of Baird's tapir, which would entail substantial time and economic investment.

Another factor that might explain the apparent overestimation of population from the ISDM is the detectability of the species in fragmented habitat. Tapir detection can be negatively related to increasing forest cover, such that tapirs were more detectable in small forest patches (Cove et al., 2014). The interpretation of intensity (from camera trap detections) correlating with true abundance could be misleading if tapir detection increases

in small forest patches. This detection phenomenon is another factor that might explain the disparity in the population estimate from the ISDM and expert opinions.

To verify the intensity estimates from this model, abundance estimations derived from other methods should be used as a comparison (i.e. via capture-recapture). Ideally, these abundance estimates should be summed up across intact cores to see how they compare to the summed intensity from this research. Interestingly, intensity values at a location do not appear to be outside the estimates of experts, it is only when summing the intensity across a wider region, including those areas not surveyed, that the model tends to disagree with expert opinion.

2.3.4 Comparison with other studies

The maximum intensity from the Integrated SDM was 4.4 individuals/ km² (Talamanca Mountains, Costa Rica) which is about 50% more than the highest reported estimates for the species (2.92 individuals/ km²) using capture-recapture methods (González-Maya et al., 2012). Though some researchers question the ability to use capture-recapture methods with camera traps on a species such as Baird's tapir due to lack of natural markings (Oliveira-Santos, Zucco, Antunes, & Crawshaw, 2010), others disagree with the validity of the methodology for different reasons (Tobler, Hibert, Debeir, & Richard-Hansen, 2014). Mark-resight models (McClintock, White, Antolin, & Tripp, 2009) or spatially-explicit models for presence-absence data (Ramsey, Caley, & Robley, 2015) could be good alternatives for estimating density but, have not yet been applied to tapir data.

Estimates from the ISDM are also relatively consistent with other studies that estimated Baird's tapir density in other parts of Central America. One study estimated 0.6 individuals/ km² in Corcovado National Park in Costa Rica using transects (Naranjo, 1995),

while the ISDM estimated a mean of 0.81 individuals/ km² in the same protected area. Another study found 0.12 individuals/ km² in the core area of El Triunfo Biosphere Reserve (though the estimates range up to 0.49 using different assumptions) (Carbajal-Borges, Godínez-Gómez, & Mendoza, 2014), while the ISDM estimated a mean of 0.40 individuals/ km² across the core area of the reserve. It was also estimated that 1.02 individuals/ km² occurred in Los Katíos National Park of Colombia (Mejía-Correa, Diaz-Martinez, & Molina, 2014), while the ISDM estimated a mean of 0.65 individuals/ km² in the same protected area. Finally, 0.32 individuals/ km² were estimated using capture-recapture methods with camera traps in cloud forest in Oaxaca state Mexico (Botello et al., 2017), while the ISDM estimated a mean of 0.63 individuals/ km² in the same study area. None of these differences are on the order of the difference between expert estimates of the population for the entire species, compared to the total population estimate across viable cores using the ISDM.

2.3.5 Habitat connectivity

Regardless of the estimated population size, there are clearly areas with potential obstructions to connectivity. This was partly a result of the shape of Central America, as the land mass narrows between Nicaragua and Colombia, limiting the amount of land available for Baird's tapirs, but also due to widespread forest loss in several countries, particularly in central Nicaragua and northern Honduras/Southern Guatemala. The Panama Canal and the proposed Nicaragua Canal (Fig. 2) also pose significant potential genetic barriers. The Panama Canal already presents a substantial barrier to connectivity between habitat cores for Baird's tapir (N. Meyer et al., 2015) although there are anecdotal data of tapirs crossing the canal zone (N. Meyer, Moreno, & Jansen, 2013). The proposed construction of the larger Nicaragua Canal would further fragment the species distribution

and almost certainly pose a permanent genetic barrier because the areas with the highest density of canal related infrastructure coincide with the only remaining areas of viable Baird's tapir habitat in the vicinity of the proposed canal zone (Jordan et al., 2016). Therefore, the proposed Nicaragua Canal would isolate populations in the two major cores to the north, from several substantial cores to the south, including the Talamanca as discussed above, and a large habitat core in the Darién of Panama, which extends into the Chocó of Colombia. These two areas represent two of the largest cores for Baird's tapir, as the fifth and third largest cores, respectively. It is unclear to what extent this could affect the long-term survival of Baird's tapirs given their previously reported low genetic diversity (Norton & Ashley, 2004).

2.4 CONCLUSION

The main advantage of the integrated SDM framework tested here is the ability to use both presence-only and presence-absence data in estimating an SDM. This is especially advantageous for wide-ranging, but relatively rare, species such as the Baird's tapir. Modeling the entire distribution of these species requires the combination of several independently collected data sets, some of which may not include absence data.

One of the most exciting possibilities of the Integrated SDM, and point process models in general, is the ability to estimate species population size by summing intensities. This can be done because the expected number of individuals in each area is equal to the sum of the estimated intensities across that area (Diggle, 2013). However, care should be taken when interpreting a fitted PPM and what is being modelled (Renner et al., 2015). The Integrated SDM used in this research includes the observation process as a latent variable, and thus it should estimate the true species abundance (not relative patterns in species abundance). However, this research highlights the need for further work to validate the

population estimates from the model due to discrepancies between estimated intensity and expert opinion on total population size for the species.

It is possible that the seemingly high population estimates from this model for Baird's tapir are due to missing proximate (i.e. causal) environmental variables, meaning the model is closer to the potential (i.e. carrying capacity) than the actual distribution of the species (M. P. Austin, 2002). In other words, areas identified as suitable may not actually be occupied by the species due to movement constraints (e.g. dispersal barriers), competition with other species, disease, hunting pressure, or other variables not included in the model.

It is also possible that duplicate records of single individuals exist in the data used to fit the model, despite efforts to reduce this possibility using spatial subsampling with a minimum allowable distance of 5 km. Future research using simulated data should investigate the impact of duplicate records on the modelled intensity. Though it is possible some duplicates remain in the data used to fit the model after subsampling, this number should be small, and the effect on estimated intensity minimal. Simulated data could be ideal for testing the magnitude of this effect.

A sensitivity analysis of the Integrated SDM framework using the same Baird's tapir data used in this research found that the spatial resolution of the analysis has a strong effect on the magnitude of intensity and total population estimates (Schank et al., in prep). It is suspected that this is due to the effective sample area being expanded by the individual movements of the species, and the way area is included in the model formulation. Thus, any simulation that seeks to understand the model better should incorporate animal movement.

CHAPTER 3: A sensitivity analysis of Integrated Species Distribution Models (ISDMs): Investigating the effect of multiple spatio-temporal scales³

3.1 INTRODUCTION

Species Distribution Models (SDMs) are a widely applied and rapidly developing statistical tool used in the study of wildlife, with new methods regularly proposed as solutions to various challenges encountered during modelling (Elith & Leathwick, 2009; Franklin, 2010a). A deficiency of most SDMs is the failure to account for imperfect detection, the possibility that a species may go undetected even when it is present (Lahoz-Monfort et al., 2014). Occupancy models, a similar but distinct field of research from SDMs, account for this scenario by separating the species-environment response from that of the detection process using a hierarchical modeling framework (MacKenzie et al., 2003). Another challenge for most SDMs is how to appropriately use presence-only (PO) data, which are often the most common type of data used in SDMs, due to their ease of collection. This type of data is sometimes also referred to as presence-background (PB), for the class of models that combine PO data with the background environment to estimate species-environment responses. Recently, the challenge of using PO data in SDMs has been addressed using point process models (Renner et al., 2015; Warton & Shepherd, 2010). Integrated SDMS (ISDMs) represent a new development that uses both approaches, combining opportunistic (e.g. PO) and higher quality site-occupancy (SO) data in the same model (Dorazio, 2014a; Koshkina et al., 2017). ISDMs have potential as useful tools, but they require further investigation (i.e. sensitivity analyses), as there are few applied examples to follow.

³ The manuscript written for this chapter is under review with Ecological Applications.

Data simulation is a powerful tool used to answer questions about how models react to various user decisions (J. Miller, 2014; Zurell et al., 2010). However, the design of simulated studies sometimes assumes data conditions that are unrealistic for many rare or cryptic species. The assumptions of the simulations used in the two studies that introduced ISDMs (Dorazio, 2014a; Koshkina et al., 2017), include a larger and more even sample than is typically available for most species. Simulation studies that do not mirror reality are especially problematic for a species like Baird's tapir (*Tapirus bairdii*), which is wide-ranging and relatively rare, leading to wide gaps in spatial coverage of high quality presence data. Schank et al. (2017) applied an ISDM to approximately 800 presence-only (PO) observations and 1600 camera trap detection histories (created from SO data) for Baird's tapir. This research estimated a total population size of approximately 200,000 individuals for the species, more than an order of magnitude higher than expert estimates (which range from 3000 mature adults, to 16500 total individuals). There are a variety of reasons that could explain this discrepancy, including violations of model assumptions and sensitivity of the model to various modelling decisions. I focused on two assumptions: independence between sites and population closure. As with most statistical models, occupancy models require independent observations. In this case, observations would not be independent if the same individual was detected at more than one site during the same observation period. To avoid this possibility, Schank et al. (2017) used a spatial subsampling procedure to enforce a minimum distance between sites, as many of the sites were so close together that independence would be violated. Occupancy models also assume population closure, which states that no immigration or emigration of individuals from the site occurs during the sampling period (MacKenzie et al., 2006). Violation of the closure assumption can originate from a sampling period that is too long (C. T. Rota, Fletcher, & Dorazio, 2009).

This research investigated the effect of user decisions on model outputs and population estimates when using ISDMs, focusing on how issues of spatial and temporal scale relate to the model assumptions above. Specifically, I investigated the effect of different settings for site area, subsampling radius, and sampling period length, using data from prior Baird's tapir analysis (Schank et al. 2017). ISDMs have great potential as useful tools for conservation, however, researchers using these tools need clear recommendations for how to apply them. The results of this research sheds light on how these models can be applied appropriately to species of conservation concern.

3.2 METHODS

The complete sensitivity analysis covered three model formulations (PB, SO, and Integrated), four site area sizes, and three sampling period lengths, with 100 spatially subsampled iterations -- a total of 3600 models. Custom R code was adapted from Dorazio (2014) and Royle and Dorazio (2008) to run the models.

3.2.1 Model Descriptions

With ISDMs, two separate models are formulated to accommodate the two types of data used (PB and SO), both based on a Poisson point process model. In these models, $\lambda(s)$ is the expected intensity (number of individuals per unit area) at location s . In the context of the SDM, $\lambda(s)$ is formulated as a log-linear function of unknown parameters and location-specific regressors $x(s)$,

$$\log(\lambda(s)) = \beta_0 + \beta'x(s)$$

The general class of models used here are hierarchical models, which have separate levels for abundance (the process of interest), and detection (the nuisance process). Though they share the same SDM based on a point process model, the two types of data use different model formulations to account for imperfect detection, including that which

results from spatially biased survey effort. With opportunistic data, spatial bias and imperfect detection are incorporated through an independent thinning of the point process. This thinned point process is the product of the original point process and $p_{pb}(s)$, the probability that the site is surveyed, and the species is detected. $p_{pb}(s)$ is formulated as a logistic function of unknown parameters and location-specific regressors $w_{pb}(s)$,

$$\text{logit}\left(p_{pb}(s)\right) = \alpha_{0,pb} + \alpha'_{pb} \cdot w_{pb}(s)$$

With data from planned surveys (SO), imperfect detection is modeled following a zero-inflated binomial distribution (Koshkina et al., 2017). Under this model, the presence or absence of the species at a site, i , follows a Bernoulli distribution. In this case, the detection histories at each site, y_i , have non-detections (i.e. zeros) due to both species absence and imperfect detectability, the fact that an individual may go undetected even when present (MacKenzie et al., 2003). This relationship is modeled as a Binomial distribution with J trials, and the probability of success (i.e. species detection) equal to the product of z_i (the occupancy state, $z_i = I(N_i > 0)$) and p_{so} , the probability of detection at the site. As with detectability in the PB model, $p_{so}(s)$ is formulated as a logistic function of unknown parameters and location-specific regressors $w_{so}(s)$,

$$\text{logit}\left(p_{so}(s)\right) = \alpha_{0,so} + \alpha'_{so} \cdot w_{so}(s)$$

In ISDMs, the PB and SO models are estimated simultaneously, such that one set of parameters for the SDM is created (i.e. the β 's), while separate detectability parameters are estimated (i.e. the α 's) for the two models.

3.2.2 Model Convergence and Parameter Identifiability

There are two levels of model convergence. First, the *optim* function in R was used to estimate model parameters from the model likelihood. Occasionally this function failed to return an optimized set of parameters. Next, if estimates were returned from this

function, I determined whether they were correctly identified using the reciprocal of the condition number. This number is the ratio of the smallest to the largest eigenvalues in the Fisher information matrix and can be used to determine whether the parameters of the SDM are identifiable (Dorazio, 2014a). The reciprocal of the condition number falls between 0 and 1, with values near zero indicating ill-conditioning (Golub & Van Loan, 2012). If this number fell below a certain threshold (in this case 1×10^{-6}), the results for that model were not used in the subsequent analysis.

3.2.3 Presence Data

The species presence data used here consisted of 784 presence-only observations compiled from eleven data sources, and 1595 camera trap detection histories from nineteen sources. These data came from a multinational collaboration examining Baird's tapir occurrence and distribution (Schank et al., 2017).

3.2.4 Spatial Subsampling

I processed the presence data prior to model fitting using a random spatial subsampling procedure to help preserve independence among sites. The algorithm began by randomly choosing one observation point and removing any other observation points within a given radius. I added the chosen observation to the subset and repeated the steps until no observations were left in the original data. A similar type of subsampling is sometimes used to remove survey bias in observation data (Beck et al., 2014), however, this grid-based approach can lead to samples that remain close in space, if they fall just across a boundary in an adjacent grid cell.

The effect of this procedure was to enforce a minimum distance between sampling points. This minimum distance was matched to the site area to ensure that no site contained more than one data point. For example, I used a subsampling radius of 5657m for a site

area of 16km² (the diagonal length of a square that size). This process was repeated one hundred times for each radius to capture the variability introduced by the randomness of the sampling. Model parameters were then averaged across iterations. A set of models also were fit on the complete (non-subsampled) presence data to investigate the effect of violating the independence assumption.

3.2.5 Predictor Variables

Environmental variables were grouped into five classes: climate, land cover, anthropogenic, topographic, and sampling variables (i.e. the variables used in the detection process). Climate variables at 1-kilometer resolution were downloaded from CHELSA (Karger et al., 2017), and included annual precipitation, maximum temperature of the warmest month, temperature seasonality, and precipitation seasonality. Land cover variables consisted of percent forest cover for the year 2000 at 30-meter resolution (Hansen et al., 2013), distance to/within protected areas (IUCN & UNEP-WCMC 2014) and mean enhanced vegetation index (EVI) from MODIS for years 2000 to 2015 downloaded using Google Earth Engine. Anthropogenic variables included forest loss between 2000 and 2014 (Hansen et al., 2013) road density (Eugster & Schlesinger, 2010), and density of fires between 2001 and 2014 (NASA, 2017). Anthropogenic variables were then converted to focal averages using a moving circular window and a ten-kilometer radius (centered on each one-kilometer pixel in the study area), to account for the fact that humans are mobile, and presence in one area means access is likely within a reasonable distance (Barber, Cochrane, Souza, & Laurance, 2014). Slope was calculated from 90-meter resolution elevation data downloaded from the ‘raster’ package in R (Hijmans, van Etten, et al., 2016).

Sampling variables (i.e. those used in the detectability process) for the PB data included binary indicators for forest (Arino et al., 2012) and protected status (IUCN &

UNEP-WCMC 2014), and distance to roads (Eugster & Schlesinger, 2010), while sampling variables for SO data were the same forest cover and distance to/within protected areas used in the land cover group, as well as distance to roads, and maximum slope. With the PB data, these variables were meant to capture sampling bias in presence-only data, which I believe heavily favors forested and protected areas that are reasonably accessible by road. I included a quadratic term for distance to roads, as there could be optimal locations that are far enough from roads to minimize anthropogenic factors, but close enough to facilitate sampling. With SO data, the sampling variables were chosen as variables which might influence the detectability of the species. For example, tapir detectability might decrease as distance from protected areas increases, due to likely increased levels of hunting outside of protected areas, and thus increased response by the species to avoid humans (de la Torre, Rivero, Camacho, & Álvarez-Márquez, 2017; Ferreguetti, Tomás, & Bergallo, 2017).

All variables were scaled to have a mean of zero and standard deviation of one, except distance to/within protected areas (zero represents the border of the protected area). The models also incorporated quadratic terms for all climate variables, EVI, and distance to road, to account for their suspected non-monotonic relationships with tapir occurrence (aided by single-variable response curves created in early stages of the modelling process).

I resampled all environmental variables to four spatial resolutions: 1, 2, 4, and 8 kilometers. These resolutions correspond to site areas of 1, 4, 16, and 64 km². Estimates of home range size for Baird's tapir vary from 1.25 km² reported in Costa Rica (Foerster & Vaughan, 2002) to 8-10 km² in Nicaragua (C. Jordan personal observation, 2016), while estimates of maximum distance traveled range up to 10.5 km in Mexico, from camera trap data on a marked individual over four years (Reyna-Hurtado, Sanvicente-López, Pérez-Flores, Carrillo-Reyna, & Calmé, 2016b). From this information, it is possible that an

individual could be detected at more than one site, but the likelihood of this is probably small, especially for the larger site areas used in this analysis.

3.2.6 Sampling Period Length

When creating camera trap detection histories, researchers can adjust their data structure by defining: 1) the length of each sampling occasion, and 2) the number of sampling occasions to use in a discretized season. Since camera traps operate continuously, there is some flexibility in determining the sampling occasion and sampling period length. These decisions can be made (and adjusted) after the data is collected and will determine the balance of detections and non-detections. It is also important to consider the behavior of the target species. For sampling occasion length, one suggestion is to select a length of time during which an individual will visit all or most of its home range, and GPS telemetry data suggest Baird's tapir cycle through their home ranges about once every 10-12 days (Jordan, 2015). For this reason, I used a sampling occasion length of 10 days.

With sampling period length (i.e. number of sampling occasions), it is important to consider the assumption of closure, and choose a length of time during which immigration/emigration at a site is unlikely. As the sampling period length length increases, it becomes more likely that the assumption of closure will be violated. On the other hand, it is crucial to include enough sampling occasions to estimate detectability reliably. Some recommendations suggest three as the minimum number of occasions to use, though this number should be higher for species with low detectability (Mackenzie & Royle, 2005). Baird's tapir is unsurprisingly a species with low detection probability (range of 0.2-0.3) (Cove et al., 2014; Jordan, 2015). Considering a detectability in this range, there is about a 4-13% chance a present individual will go undetected after 9 sampling occasions. With the SO data, I tested sampling period lengths of 30, 60 and 90 days (three, six, and

nine samples). The PB data contain only one sample, as there are not repeat observations at each site for this data set.

3.2.7 Accuracy Assessment

I used two presence-only (PO) accuracy measures to assess the spatial predictions of each model: the Boyce Index (Boyce, Vernier, Nielsen, & Schmiegelow, 2002) and the minimum predicted area (Engler, Guisan, & Rechsteiner, 2004b). I did not use detection/non-detection data with accuracy measures that require presence-absence data given the difficulty of properly defining absences (Lobo et al., 2010), and given the bias of these measures when test data are missing from large portions of the study area (Bean, Stafford, & Brashares, 2012). After the spatial subsampling step, the retained PO data were randomly split following a 75/25 training/testing ratio (Fielding & Bell, 1997). For both accuracy measures, intensity was converted to occupancy, ψ , which ranges from 0 to 1, using the formula from Dorazio (2014), where N is the number of individuals in the spatial unit, C :

$$\Pr(N(C) > 0) = \psi = 1 - \exp(-\mu(C))$$

$$\mu(C) = \int_C \lambda(s) \cdot ds$$

To calculate the Boyce Index, I partitioned the occupancy surface into bins (i.e. $0.0 < \psi < 0.1, \dots, 0.9 < \psi < 1.0$), and calculated the percentage of test data occurring in each bin (P_i). I then compared the proportion of the area covered by the bin with respect to the study area (E_i). Finally, I converted these two measures to a ratio $F_i = P_i / E_i$. If the model correctly predicts low suitability areas, the low suitability classes should contain fewer test points than expected by chance (i.e. $F_i < 1$), and the graph of F_i versus average suitability of each bin should be monotonically increasing. The Boyce Index is the correlation between the average suitability of each bin and its respective F_i , with values greater than 0

indicating a model whose predictions are consistent with the test data, and negative values indicating an incorrect model. The continuous version of this measure uses overlapping bins (Hirzel, Le Lay, Helfer, Randin, & Guisan, 2006).

The minimum predicted area (MPA) is the smallest possible area covered by a thresholded prediction map which contains at least 90% of the test PO points. The smaller the MPA, the more parsimonious the model, and the less likely there are errors of commission in the predictions (Rupprecht, Oldeland, & Finckh, 2011).

3.3 RESULTS

The ISDM converged (with estimated standard errors) in more than 97% of model iterations, while the PB model had low convergence rates across site areas, and the SO model exhibited a sharp drop-off in convergence at 16km² and above (Table 3.1). Both measures of model accuracy showed that the ISDM was the most accurate framework, a relationship that was consistent across site area and number of samples (Figure 3.1). Focusing on the ISDM, estimates of total population decreased exponentially as site area increased (Figure 3.2). The number of samples used, and whether the data had been subsampled, had a much smaller effect on population estimates.

The decrease in population estimates across site areas was driven by estimates of model intercepts, primarily β_0 , while the coefficients representing species-environment relationships remained relatively stable (Table 3.2). Annual precipitation (+), forest cover (+), and road density (-) were the three most important environmental variables in the model. Temperature seasonality (-), precipitation seasonality (+), maximum temperature of the warmest month (+), EVI (-), forest loss (-), and maximum slope (-) also appeared as significant environmental predictors. Annual precipitation was the only environmental variable with a clearly significant quadratic term. In the PB detectability process, presence

in a protected area (+) was the most important variable. Distance to roads (-) was significant in both detectability components (PB and SO). Maximum slope (+) was also a significant variable in the SO detectability process.

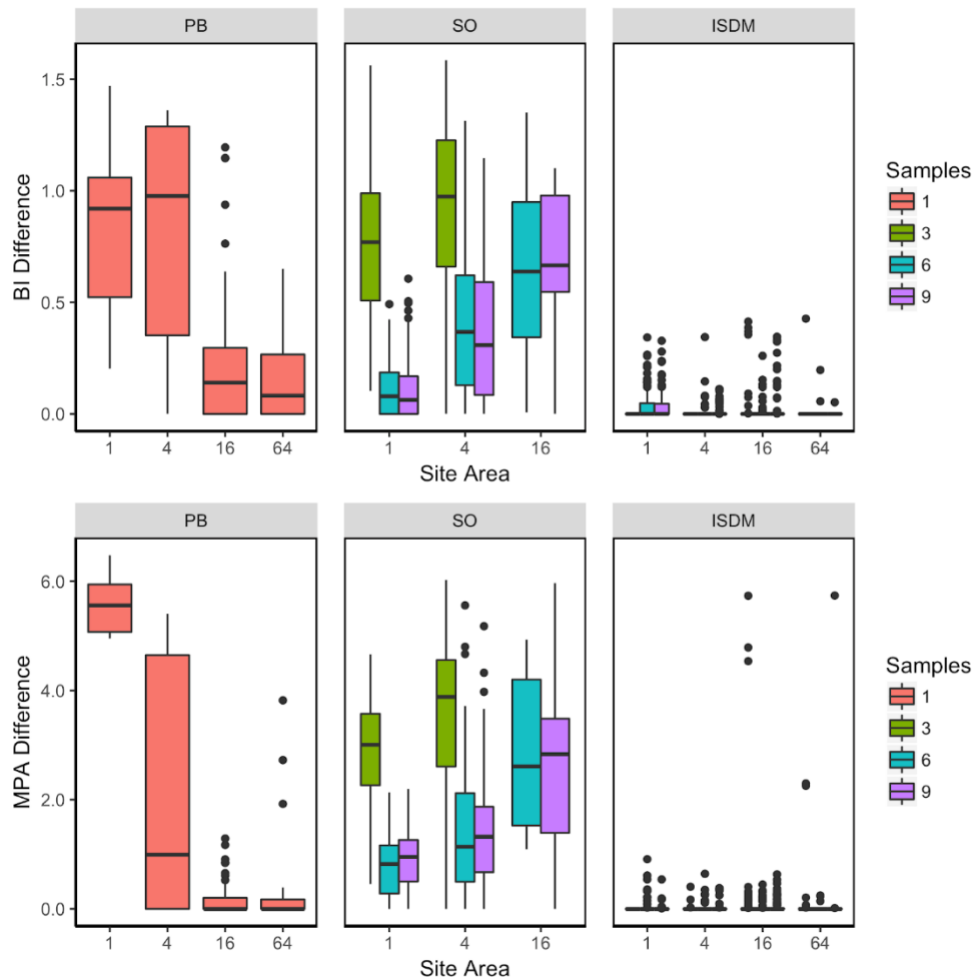


Figure 3.1: Model accuracy using Boyce Index (BI) and Minimum Predicted Area (MPA). The difference calculated from the maximum (i.e. most accurate) Boyce Index, and from the minimum (i.e. most accurate) Minimum Predicted Area, within each combination of model settings. The greater the difference (from zero), the less accurate the result. MPA differences have been rescaled to positive numbers and are represented in units of 100,000 square kilometers.

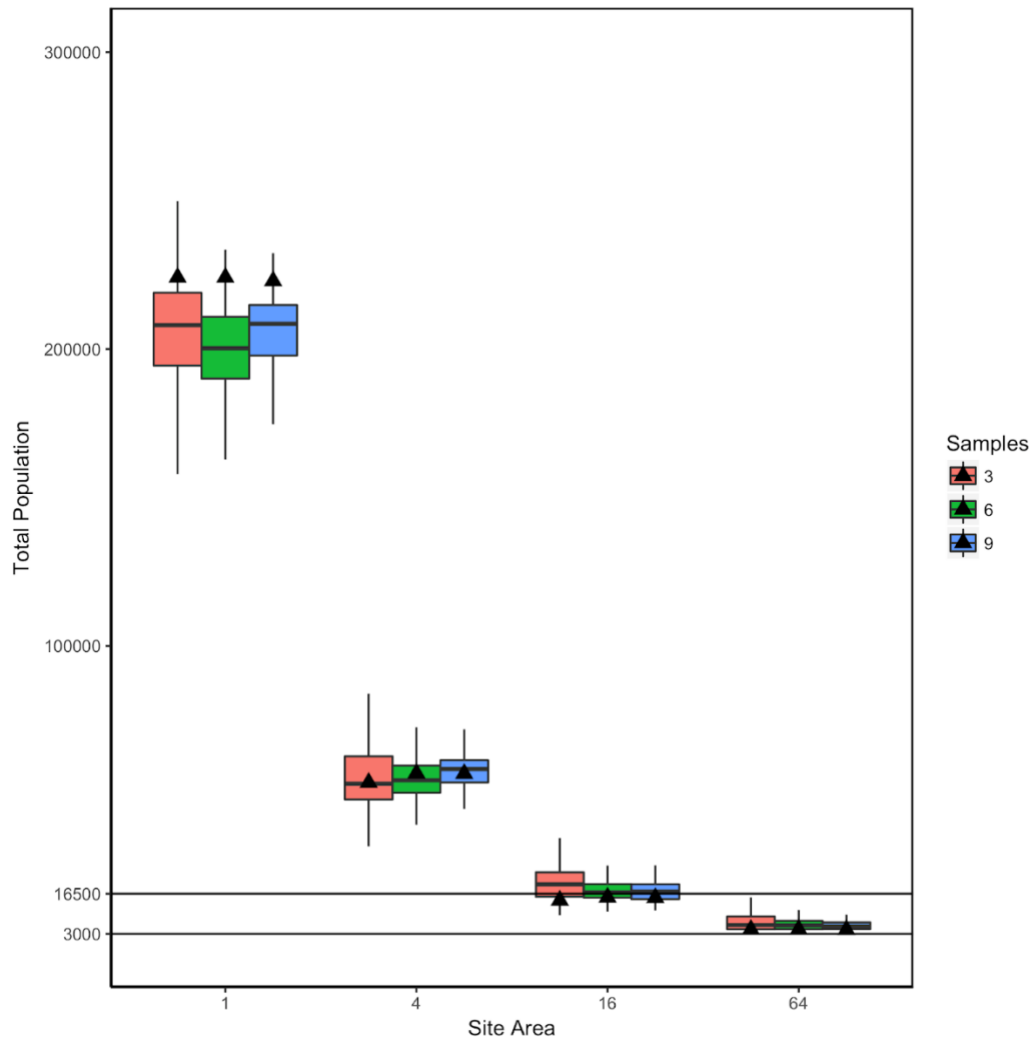


Figure 3.2: Population estimates from the Integrated SDM framework. Horizontal lines are placed at expert population estimates for the species, 3000 (current Red List assessment: Garcia et al 2017) and 16500 (Population Viability Assessment report: Medici et al 2005). Black triangles are the estimates for models run on the complete (not sub-sampled) set of presence data.

| Model | Site Area | Samples | Rate |
|--------------|------------------|----------------|-------------|
| PB | 1 | 1 | 0.03 |
| PB | 4 | 1 | 0.03 |
| PB | 16 | 1 | 0.41 |
| PB | 64 | 1 | 0.08 |
| SO | 1 | 3 | 1 |
| SO | 1 | 6 | 1 |
| SO | 1 | 9 | 1 |
| SO | 4 | 3 | 0.99 |
| SO | 4 | 6 | 1 |
| SO | 4 | 9 | 0.98 |
| SO | 16 | 3 | 0 |
| SO | 16 | 6 | 0.06 |
| SO | 16 | 9 | 0.1 |
| SO | 64 | 3 | 0 |
| SO | 64 | 6 | 0 |
| SO | 64 | 9 | 0 |
| ISDM | 1 | 3 | 1 |
| ISDM | 1 | 6 | 1 |
| ISDM | 1 | 9 | 1 |
| ISDM | 4 | 3 | 1 |
| ISDM | 4 | 6 | 1 |
| ISDM | 4 | 9 | 1 |
| ISDM | 16 | 3 | 0.9 |
| ISDM | 16 | 6 | 0.99 |
| ISDM | 16 | 9 | 1 |
| ISDM | 64 | 3 | 0.87 |
| ISDM | 64 | 6 | 0.95 |
| ISDM | 64 | 9 | 0.97 |

Table 3.1: Model Convergence and Standard Estimation Rates for each model framework and combination of site area and number of samples.

| Coefficient | 1 | 4 | 16 | 64 |
|---------------------------|----------------|----------------|----------------|----------------|
| beta0 | -1.403 (0.214) | -2.66 (0.258) | -3.8 (0.416) | -4.654 (0.812) |
| temp_seasonality | -0.527 (0.092) | -0.458 (0.109) | -0.438 (0.131) | -0.338 (0.163) |
| precip_seasonality | 0.448 (0.134) | 0.382 (0.16) | 0.394 (0.195) | 0.302 (0.24) |
| max_temp_warmest_month | 0.291 (0.129) | 0.236 (0.156) | 0.204 (0.191) | 0.115 (0.244) |
| annual_precip | 1.448 (0.252) | 1.398 (0.298) | 1.428 (0.356) | 1.41 (0.43) |
| temp_seasonality_sq | -0.651 (0.09) | -0.557 (0.105) | -0.41 (0.123) | -0.285 (0.148) |
| precip_seasonality_sq | -0.18 (0.089) | -0.164 (0.108) | -0.147 (0.13) | -0.129 (0.16) |
| max_temp_warmest_month_sq | 0.066 (0.026) | 0.05 (0.032) | 0.023 (0.041) | -0.006 (0.055) |
| annual_precip_sq | -1.045 (0.202) | -1.1 (0.247) | -1.213 (0.312) | -1.225 (0.379) |
| forestcover2000 | 1.692 (0.166) | 1.568 (0.193) | 1.496 (0.223) | 1.309 (0.27) |
| distancePA | -0.005 (0.103) | -0.041 (0.119) | -0.099 (0.141) | -0.095 (0.172) |
| EVI | -0.554 (0.111) | -0.567 (0.14) | -0.649 (0.168) | -0.6 (0.229) |
| EVI_sq | -0.025 (0.051) | -0.09 (0.071) | -0.09 (0.078) | -0.176 (0.119) |
| forestloss_focal | -0.217 (0.056) | -0.179 (0.066) | -0.104 (0.078) | -0.04 (0.095) |
| road_length_focal | -1.147 (0.201) | -1.185 (0.244) | -1.314 (0.322) | -1.426 (0.411) |
| fire_density_focal | 0.061 (0.091) | 0.035 (0.112) | -0.032 (0.142) | -0.031 (0.163) |
| max_slope | -0.395 (0.08) | -0.357 (0.095) | -0.312 (0.117) | -0.215 (0.146) |
| alpha0.pb | -7.353 (0.243) | -6.31 (0.297) | -5.405 (0.465) | -5.05 (0.884) |
| alpha0.so | -1.326 (0.335) | -1.839 (0.526) | -2.316 (0.751) | -1.913 (0.858) |
| pb.forest | 0.043 (0.183) | 0.164 (0.224) | 0.177 (0.271) | 0.44 (0.35) |
| pb.protected | 1.614 (0.18) | 1.462 (0.203) | 1.28 (0.235) | 1.186 (0.281) |
| pb.road_distance | -0.792 (0.107) | -0.751 (0.127) | -0.721 (0.157) | -0.744 (0.207) |
| pb.road_distance_sq | 0.146 (0.024) | 0.137 (0.03) | 0.125 (0.042) | 0.116 (0.056) |
| so.forestcover2000 | 0.301 (0.298) | 0.751 (0.469) | 0.954 (0.652) | 0.348 (0.726) |
| so.distancePA | -0.043 (0.178) | -0.01 (0.232) | -0.162 (0.354) | -0.193 (0.468) |
| so.road_distance | -0.565 (0.222) | -0.811 (0.298) | -1.309 (0.469) | -1.59 (0.798) |
| so.road_distance_sq | 0.103 (0.082) | 0.166 (0.108) | 0.211 (0.201) | -0.08 (0.607) |
| so.max_slope | 0.316 (0.087) | 0.307 (0.11) | 0.344 (0.168) | 0.412 (0.224) |

Table 3.2: Coefficient estimates and standard errors for the Integrated SDM (samples = 6) averaged across 100 model iterations fit on randomly subsampled presence data.

3.4 DISCUSSION

In the SO model, convergence decreased for larger site areas possibly due to reduced sample sizes following the subsampling step (mean sample sizes, 1 km²: 663, 4km²: 370, 16km²: 182, 64km²: 93). The ISDM could maintain convergence at these larger site areas possibly because of the added information from the PB data. However, there was a detectable decline in convergence with shorter sampling period lengths at these larger site areas. Clearly sample size is affected by both number of sites, and number of repeated observations at those sites (Mackenzie & Royle, 2005). In addition to higher rates of convergence, the ISDM was consistently the most accurate model. Taken together, these results demonstrate the importance of this new modeling framework. The ability to combine two types of presence data in the same model leads to better results.

In the ISDM, any of the species-environment relationships exhibited the expected outcome for this species (e.g. a preference for forest and avoidance of humans) (Cove et al., 2014; Jordan et al., 2016). However, there are two results which contradicted expectations. As in Schank et al. (2017), EVI had a negative relationship with tapir intensity. However, tapir intensity should be positively associated with increasing vegetation (higher EVI) because vegetation is both a food source and it provides cover (Brooks et al., 1997; Pettorelli, Ryan, Mueller, & Bunnefeld, 2011). Also surprising was the positive relationship with maximum slope in the detectability of the SO data. This variable was included in this part of the model as it was suspected to either have a negative effect on sampling effort, because steep terrain is harder to sample, or actual detectability of tapirs due to the same constraints, as difficult terrain is an impediment to wildlife movement as well (D. W. Bailey, Gross, & Laca, 1996; Mair & Ruete, 2016).

Clearly, the most important factor driving estimated population (and underlying magnitude of intensity) was the assumption about the size of a sampling unit, or site area.

The number of sampling occasions and whether the presence data had been subsampled to preserve site independence, had a much smaller effect, with no discernible patterns. The models developed in Dorazio (2014) and Koshkina et al. (2017) explicitly incorporate site area in a way that is different than in traditional occupancy models. This likely explains the behavior of total population changing proportional to the area of a grid cell used in the analysis. In the traditional formulation of an occupancy model (see Panel 3.8 in Royle & Dorazio (2008)), site area is not included anywhere in the model likelihood. The important question that remains is what exactly constitutes a site? For sessile species, or species that have small home ranges relative to the survey method, the site is easily defined as the area covered during the survey (i.e. a quadrat). However, when the species is relatively mobile, with a home range much bigger than the area covered in the survey, the concept of the site is less straightforward (M. G. Efford & Dawson, 2012).

For example, when using camera traps, the cone of detection (i.e. the area in which a species can trigger the camera), is often very small in relation to the movements of the target species, which are typically large and mobile. The Effective Sampling Area (ESA) is the area that contains the activity centers of any individuals that could come in contact with this cone of detection (Figure 3.3) (White, 1982). This area should be approximately equal to the average home range of the species. Original estimates of the home range for Baird's tapir were about one square kilometer (Foerster & Vaughan, 2002), surprisingly small for a species of its size. More recent estimates put the home range size closer to 10 km² (C. Jordan personal observation, 2016). The differences in reported estimates of home range size could be due to differences in the methodology used, and differences in topography and the availability of resources, specifically regarding the availability of water. In mountainous sites with complex topography and permanent availability of quality water throughout the year, the home range could be much smaller than in flat sites with

very marked seasonality (Botello et al., 2017). Interestingly, using a site area of 16km², provides a total population estimate that is within the range of expert estimates for the species (Figure 3.2).

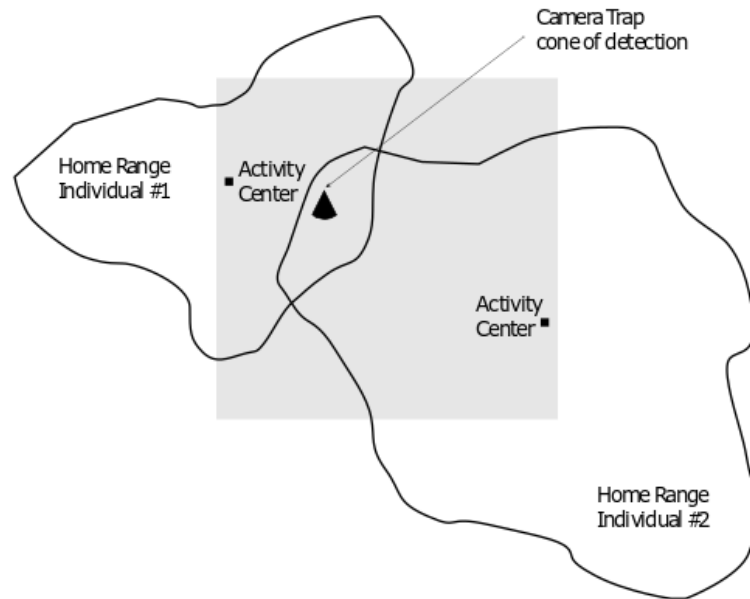


Figure 3.3: Effective Sampling Area. A simplified diagram of two individuals with overlapping home ranges, and a camera trap in the area of their intersection. The effective sampling area is equal to the area that incorporates the activity centers of all individuals detected at the camera.

In Schank et al. (2017), the model was implemented using a site area of 1km². To contextualize the population estimates from that model, the results were compared to multiple independent studies that focused on estimated abundance of Baird's tapir. The estimates from those studies were similar to the estimates using the ISDM, which provided a conflicting story, as the total population estimates were thought to be an overestimate by at least an order of magnitude. Most of these independent studies used Capture-Recapture

methods, and did control for the ESA, however, they used the old estimate of home range size from Foerster and Vaughan (2002). Clearly, accounting for the size of a species' home range, and the variability in those estimates, has a huge effect on abundance and population estimates using the ISDM (Figure 3.2).

In fact, some authors have called into question the ability to use Capture-Recapture on a species like Baird's tapir, which does not have obvious and distinct markings by which individuals can be identified (Foster & Harmsen, 2012). In the case of Baird's tapirs, sexually immature subadults lose their juvenile pelage before one year of age and develop very quickly. In addition to making individual adults difficult to identify, this makes older juveniles effectively indistinguishable from mature adults in camera traps. This means that sexually immature individuals have likely been included into prior population estimates using Capture-Recapture methods and puts into doubt whether those studies accurately estimate the effective population size.

In addition to problems with misidentification of individuals, Capture-Recapture can overestimate species abundance due to the ad hoc correction of ESA (Noss et al., 2012), which can also lead to large errors in population estimates due to extrapolation (Foster & Harmsen, 2012). These critiques recommend the use of the newer Spatial Capture-Recapture (Royle et al., 2013), which explicitly accounts for species movement using an additional scaling parameter in the model.

3.5 CONCLUSION

This research has demonstrated the potential connection between ISDMs and the Effective Sampling Area (ESA). Yet, the methods used in this research include ad hoc procedures that should be replaced by formal incorporation into the statistical model. In the models, accounting for the ESA is done in a way that matches site area with the best

information about average movements for the species. Spatial Capture-Recapture provides an example for properly scaling the model to incorporate animal movement (Royle et al., 2013). Rather than approximating this effect through the selection of an appropriate site area, it would be better to combine concepts from Spatial Capture-Recapture with the ISDMs used here.

Second, sometimes additional ad hoc steps must be taken to “fix” the data. In this case, a spatial subsampling approach was used to avoid duplicate observations of the same individual at more than one site. Here Spatial Capture-Recapture can provide some guidance as well. These models require that sites be close enough to ensure individuals are observed at more than one site and uses this information to help estimate the spatial scalar of movement for the species. Thus, combining concepts from Spatial Capture-Recapture with ISDMs may allow for the use of all data possible, though some alterations may be necessary for sites that have data covering more than one sampling period (as the tapir data used here does).

A significant contribution of this research is the linkage between the ESA and estimating abundance using ISDMs (or any other SDM/occupancy model). It is unclear why the discussion of ESA is almost entirely tied to Capture-Recapture models that use marked individuals. However, there is at least one study which addresses this issue as it relates to occupancy models (M. G. Efford & Dawson, 2012).

The issue created by incorrectly accounting for the ESA only becomes apparent in a small number of situations: when studying mobile species, estimating their abundance, and extrapolating these estimates to produce population estimates. With Baird’s tapir, these steps made it clear that something could be incorrect in the model. While it is possible to hypothesize multiple reasons for this disparity (see Conclusion Schank et al. 2017), the most straightforward answer is that ESA was not properly accounted for.

The incorporation of ESA into SDMs and occupancy models could use additional research. Failure to account for this properly could lead to inaccurate estimation of occupancy or abundance. Although, as seen in this research, species-environment relationships might remain the same. To answer these questions, a future study using a detailed simulation (incorporating the movement of individuals) is needed. Future modeling efforts for this species should also explore unexpected species-environment relationships in more detail (e.g. negative association between species presence and EVI, and positive association between species detectability and slope). It is possible that there is some interaction with other variables which caused these unexpected results. Finally, the reciprocal of the condition number used to determine parameter identifiability is new to species distribution modeling and should be investigated further.

CHAPTER 4: Population status, connectivity, and conservation action for Baird's Tapir

4.1 INTRODUCTION

Approximately a quarter of all mammal species are under threat, and in need of some action to decrease their risk of extinction (Schipper et al., 2008). This is especially true of large mammals, which require greater amounts of habitat, are often slow reproducers, and face threats from hunting or conflict with humans (Dirzo et al., 2014). Within this group, large herbivores play a vital role in seed dispersal and other ecosystem functions (O'Farrill, Calmé, Sengupta, & Gonzalez, 2012; Ripple et al., 2015). Therefore, the loss of these species across landscapes can have cascading effects on the environment and can even decrease carbon storage of forests (Bello et al., 2015; Osuri et al., 2016). In addition to habitat loss, increasing habitat fragmentation leads to isolated populations that face an increased risk of extinction, threatening the overall viability of a species (Reed, 2004). Decreased genetic diversity can also result in negative impacts on species' viability, due to susceptibility to inbreeding or disease (Frankham, 2005).

To facilitate species conservation in response to these threats, it is necessary to create range-wide conservation plans to prioritize and direct actions to specific areas (Sanderson et al., 2002). These plans consider both habitat patches and their connectivity, identifying the most important patches and corridors in need of conservation attention at both national and international levels. In most cases, this planning involves collection of expert knowledge on the species of concern (A. Rabinowitz & Zeller, 2010). However, many tools exist to make inferences about species distributions and connectivity from existing presence data, and these have been shown to be more effective than expert information alone (Zeller, McGarigal, & Whiteley, 2012). Regardless of the information

used in the planning process, a range-wide perspective is important if the overall conservation of the species is the goal (Sanderson et al., 2002).

Baird's tapir (*Tapirus bairdii*) is the largest terrestrial herbivore native to Mesoamerica and plays an important role in ecosystems as a browser and seed disperser (Brooks et al., 1997; O'Farrill et al., 2012). However, this species is highly threatened by habitat destruction, habitat fragmentation and poaching (Cove et al., 2014; Naranjo, 2009; Naranjo, Alcalá, Briones, & Hurtado, 2015). As a result, it is classified as Endangered by the IUCN Red List (García et al., 2016). Additionally, tapirs are relatively unique evolutionarily, thus their extinction would represent a greater loss of genetic lineage compared to a species with closer genetic relatives (Isaac et al., 2007). It is estimated that the distribution of Baird's tapir has been reduced drastically, possibly by as much as 50% in the past 30 years (García et al., 2016).

This research combined an Integrated Species Distribution Model (ISDM) with connectivity analyses to identify suitable habitat patches for Baird's tapirs and the critical areas required to maintain connectivity across its distributional range from Mexico to northern Colombia. Previous research modeled the range-wide distribution for Baird's tapir using an ISDM (Schank et al., 2017), yet no effort has been made to assess the connectivity of the habitat patches identified, or a prioritization plan for conservation action based on such an analysis. This research addresses these gaps, using best practices for ISDMs (Schank et al., 2018), then applying a network analysis to habitat patches to calculate the contribution of each patch to overall network integrity (i.e. connectedness), and isolation from other patches. These measures are combined with others, including population size, protection status, and forest loss, to develop separate indices of ecological importance and vulnerability. This research balances these two aspects of extinction risk when developing a conservation plan, similar to others that previously conducted range-

wide assessments (Altrichter et al., 2012; Zeller & Rabinowitz, 2011). Common SDM methods have been combined with graph networks in several instances (Dilts et al., 2016; Foltete, Clauzel, Vuidel, & Tournant, 2012), yet this work represents a first step in combining ISDMs with networks and using the results to develop basic indices for conservation prioritization. The results are used to develop a set of spatially explicit conservation actions for Baird's tapirs, in addition to suggestions for improvements to the models and methods used. This approach may be useful for many species that have not yet been assessed in a similar manner.

4.2 DATA AND METHODS

4.2.1 Presence Data and Environmental Predictors

The presence data used in the ISDM were from an international collaboration among more than 30 researchers across Mexico, Central America, and Colombia. These data included approximately 800 presence-only (PO) points and 1600 sites with site-occupancy (SO) data from camera traps. Details and sources of this data can be found in previous publications (Schank et al., 2017, 2018).

Due to the clustered nature of the presence data, a random spatial subsampling (pre-modeling) step was used to enforce a minimum distance between the presence locations used in the model (Schank et al., 2017). The distance selected for this step was based on an estimate of the effective sampling area (ESA) around each site, using knowledge about the average home range for the species, which ranges from 1 to 24 km² (Foerster & Vaughan, 2002; Reyna-Hurtado et al., 2016a). To maintain a consistent spatial scale, the pixel size of environmental predictors was matched to the estimated ESA (i.e. 4 km resolution for an ESA of 16 km²), and a subsampling radius of 5.7 km was used to ensure that no two presence points could fall in the same pixel. The ISDM was trained separately

on 1000 random samples generated by this process to capture the variability introduced by this step. Parameter estimates were then averaged across these iterations.

Environmental predictors used in the model included climate, land cover, anthropogenic disturbance, and terrain. Temperature and precipitation seasonality, maximum temperature of the warmest month, and annual precipitation were used as climate variables in the model. These variables were selected as a subset of important predictors based on earlier modeling efforts for the species (Mendoza et al., 2013; Schank et al., 2015). Climate layers were downloaded from CHELSA (Karger et al., 2017) at 1 km resolution (all subsequent variables were resampled to match this resolution and extent). Percentage forest cover from the year 2000 (Hansen et al., 2013), distance to/within protected areas (IUCN & UNEP-WCMC 2014), mean Enhanced Vegetation Index (EVI) from years 2000-2015 (Google Earth Engine Team, 2015), and water occurrence (Pekel, Cottam, Gorelick, & Belward, 2016) were used as land cover variables. Forest loss between 2000 and 2016 (Hansen et al., 2013), fire frequency for approximately the same years (NASA, 2017), and road density (Eugster & Schlesinger, 2010) were used as indicators of anthropogenic disturbance. Finally, slope was used to incorporate the effect of terrain. This variable was calculated from 90 m resolution elevation data (i.e. SRTM) downloaded from the ‘raster’ package in R (Hijmans, van Etten, et al., 2016).

The ISDM also incorporated variables that could influence the detectability of the species, which allowed a separate set of detectability variables for the two different types of presence data used in the model (i.e. PO and SO). For the PO data, these variables included binary indicators for forest (Arino et al., 2012) and protected status (IUCN & UNEP-WCMC 2014), distance to roads (Eugster & Schlesinger, 2010), and slope. Variables for SO data incorporated land cover and anthropogenic disturbance, as well as distance to roads, slope, and dummy variables for different sampling methods (on- vs off-

trail and surveys targeting either tapirs, large cats, or general biodiversity inventories). These last variables were included to control for the different sampling regimes used by the various sources of presence data. Variables used with PO data were meant to capture sampling bias in presence-only data, which heavily favors forested and protected areas that are reasonably accessible by road. A quadratic term for distance to roads was also used, as there could be optimal locations that are far enough from roads to minimize anthropogenic factors, but close enough to facilitate sampling. With SO data, the sampling variables were chosen as variables that might influence the detectability of the species. For example, tapir detectability might decrease as distance from protected areas increases, due to avoidance of humans as result of heightened levels of hunting outside of protected areas (de la Torre et al., 2017; Ferregueti et al., 2017).

All variables were resampled to a resolution of four km, scaled ($SD=1$) and centered on zero, except for distance to/within protected areas, which was scaled but not centered. This was done to preserve zero as the boundary between inside (negative values) versus outside (positive values) of the protected area. Quadratic terms for all climate variables, EVI, water occurrence, and distance to road were used in the model to account for their suspected non-monotonic relationships with tapir presence and detectability (aided by single-variable response curves created in the early stages of the modelling process).

4.2.2 Integrated SDM

This research utilized an Integrated SDM (ISDM) to incorporate both opportunistically collected presence-only (PO) data and site-occupancy (SO) data from planned surveys. The ISDM uses a hierarchical framework to estimate separately species occurrence and detectability. Species occurrence is modeled as the intensity of a Poisson point process, $\lambda(s)$, which is the expected density (number of individuals per unit area) at

location s . The intensity is formulated as a log-linear function of unknown parameters and location-specific regressors $x(s)$,

$$\log(\lambda(s)) = \beta_0 + \beta'x(s)$$

In the ISDM, separate detectability formulations were used for the two types of data. For the PO data, spatial bias and imperfect detection are incorporated through an independent thinning of the point process. This thinned point process is the product of the original point process and $p_{po}(s)$, the probability that the site is surveyed, and the species is detected. $p_{po}(s)$ is formulated as a logistic function of unknown parameters and location-specific regressors $w_{po}(s)$,

$$\text{logit}(p_{po}(s)) = \alpha_{0,po} + \alpha'_{po} \cdot w_{po}(s)$$

With the SO data, imperfect detection is modeled following a zero-inflated binomial distribution (Koshkina et al., 2017). Under this model, the presence or absence of the species at a site, i , follows a Bernoulli distribution. In this case, the detection histories at each site, y_i , have non-detections (i.e. zeros) due to either species absence or imperfect detectability, when an individual may go undetected despite being present (MacKenzie et al., 2003). This relationship is modeled as a Binomial distribution with J trials (i.e. the number of repeated observations at a site), and the probability of success (i.e. species detection) equal to the product of z_i (the occupancy state, $z_i = \mathbf{I}(N_i > 0)$) and p_{so} , the probability of detection at the site. As with detectability in the PO model, $p_{so}(s)$ is formulated as a logistic function of unknown parameters and location-specific regressors $w_{so}(s)$,

$$\text{logit}(p_{so}(s)) = \alpha_{0,so} + \alpha'_{so} \cdot w_{so}(s)$$

In ISDMs, the PO and SO models are estimated simultaneously, such that one set of parameters for the SDM is created (i.e. the β 's), while separate detectability parameters are estimated (i.e. the α 's) for the two models. Starting values for both detectability

intercepts (i.e. $\alpha_{0,po}$ and $\alpha_{0,so}$) were set to the naive detectability for each respective data type to help solve issues with model convergence. All other parameters in the model used a starting value of 0.

The intensity estimates from the ISDM can be integrated across a spatial unit to estimate the expected number of individuals in that area. This property is used to estimate population sizes in the habitat patches identified by the analysis.

4.2.3 Quantifying Habitat Connectivity

The estimated intensity surface was converted to a map of habitat patches using a hotspot analysis. Specifically, the Getis-Ord G_i^* (i.e. local G) spatial statistic (Getis & Ord, 1992) was calculated using a 10 km radius, and statistically significant hotspots designated as habitat patches for the species. This threshold was selected as the maximum distance traveled by an individual tapir (Reyna-Hurtado et al., 2016a), thus accounting for a level of interactivity between individuals that could be representative of a habitat patch.

The estimated intensity was then summed within these patches, and those with an estimated population of 10 or more were retained for the next stages of the analysis. Researchers in the Atlantic Forest of Brazil estimated that 200 adult individuals of *Tapirus terrestris* are needed to constitute a minimum viable population (Medici & Desbiez, 2012). Rather than use 200 individuals as the cut off for these habitat patches, I used the minimum population from different test scenarios carried out in a Population Viability Assessment for Baird's tapir (Medici et al., 2005) in order to maximize the number of vertices in the network (within reason), while considering that some patches may act as population sinks (i.e. they harbor less than a viable population, and thus are reliant on immigration of individuals from adjacent patches to prevent a local extinction). In the PVA, in nearly all

scenarios, a population of ten individuals had a greater than 50% probability of going extinct within 100 years. However, these scenarios did not consider immigration.

An estimate of flow between patches was created by first calculating the least-cost distance between pixels using the intensity surface as a conductance layer (van Etten, 2017). Only those connections with distances less than an estimated maximum dispersal distance were retained, calculated as 40 times the linear dimension of the home range estimate (Bowman, Jaeger, & Fahrig, 2002). Due to the variability in home range estimates, three maximum dispersal distances were tested: 40, 80, and 160 km (corresponding to home ranges of 1, 4, and 16 km²). The least-cost distance was then used in a negative exponential dispersal kernel (Nathan, Klein, Robledo-Arnuncio, & Revilla, 2012):

$$\frac{1}{2\pi a^2} \cdot \exp\left(-\frac{r}{a}\right)$$

Where a is two times the mean dispersal distance and r is the distance between pixels. The median dispersal distance was used in place of the mean, which was estimated as seven times the linear dimension of the home range (Bowman et al., 2002). Finally, this dispersal probability was multiplied by the square root of the estimated intensity in the source pixel (Estrada & Bodin, 2008) (see Figure 4.1 for a zoomed in map documenting this step). For each pair of adjacent patches, the corresponding flow between their respective pixels was summed. This flow matrix was used to create a weighted and directed graph network of vertices (i.e. patches) and edges (i.e. corridors) with the R package ‘igraph’ (Csardi & Nepusz, 2006).

Using this graph network, two measures of vertex centrality were calculated: farness and betweenness. Farness (i.e. the inverse of closeness) was defined as the average length of the shortest paths from all other vertices in the graph. This measure was then used

to represent the isolation of individual patches in the indices calculated. Vertex betweenness is roughly defined as the number of shortest paths going through a given vertex. This measure was used to represent the importance of a given patch to the overall connectivity of the network. Each vertex was then removed from the graph one at a time, and the number of components (i.e. isolated subgraphs) recorded. This was done to identify when new components were created by removing the vertex, indicating a new break in the graph.

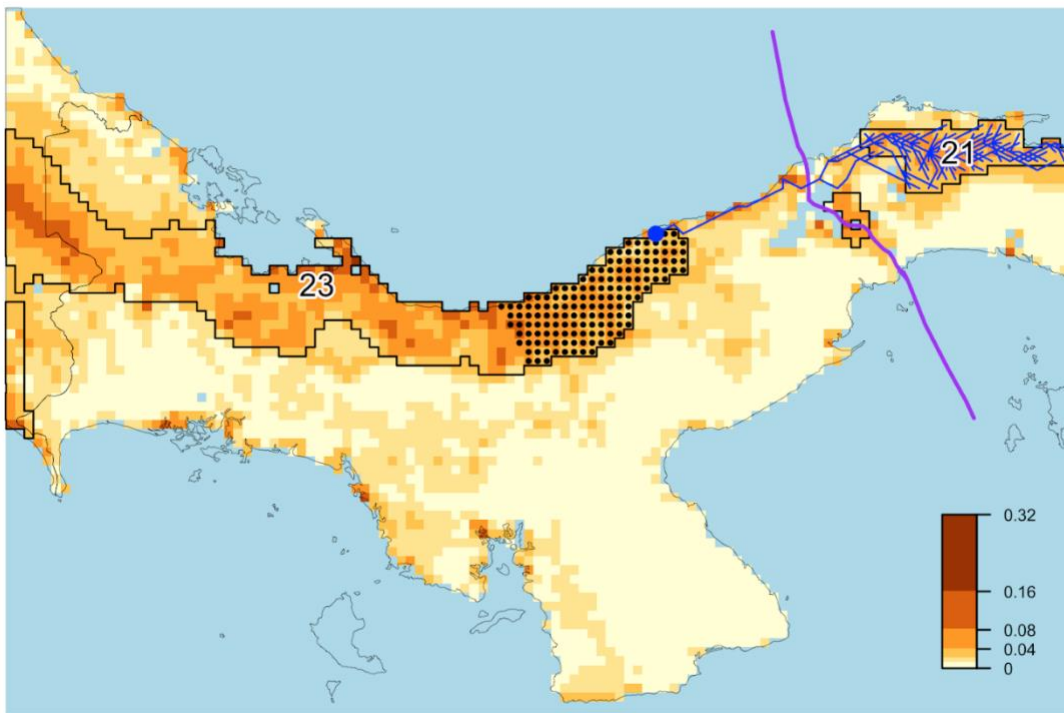


Figure 4.1: Example of flow calculation for one source pixel going from patch 23 to patch 21. The source pixel (blue dot) contains 0.758 individuals. Black dots are additional pixels that fall within range of the destination patch, and will subsequently have LCPs and flow calculated. Blue lines are LCPs. Purple line is the Panama Canal.

4.2.4 Prioritization Criteria

To prioritize the different patches for conservation, the goal was to balance vulnerability to local extinctions and importance to the survival of the species. For this reason, two separate indices were calculated for each patch and then combined for an overall index of conservation priority.

Before calculating these indices, some initial measures were created and used in the subsequent calculations: 1) the percentage of each patch that was not under protection (IUCN & UNEP-WCMC 2014), 2) a population index that converted the estimated population in each patch to a scale of 0 to 1, 3) evidence of occupation by the species using both PO and SO data (binary: 0 or 1), and 4) an index of forest loss occurring in the patch (Hansen et al., 2013). Population size provides information about both the vulnerability and ecological importance of a patch, while level of protection and amount of forest loss occurring in the patch are informative about vulnerability. Confirming the species presence in a patch provides more confidence in its ecological importance, as there are likely some patches (especially small ones) that are predicted to contain species, which do not. These patches do have value for species conservation (e.g. for translocations, or re-establishment of connectivity), but perhaps less than if they were already occupied.

The vulnerability index was calculated to represent the relative risk of a local extinction for a given habitat patch. Less protection, more deforestation, smaller populations, and greater distance from other patches in the network should lead to cores that are more vulnerable. Thus, this index was created as the sum of the following:

$$\% \text{ unprotected} + \text{deforestation} + (1 / \text{population index}) + \text{farness}$$

The importance index was created to represent the relative importance of a given habitat patch to the overall survival of the species. Larger estimated populations, confirmed presence of the species, and importance to the overall network connectivity should indicate

patches that had greater importance. Thus, this index was calculated as the sum of the following:

$$\text{population index} + \text{confirmed presence} + \text{betweenness} + \text{components created}$$

Both indices were rescaled to have a maximum value of 1, before they were summed in a combined index of conservation priority; the population measures used in these two indices will cancel each other out when combined. However, they are useful to include when considering vulnerability and importance separately. Similarly, the centrality measures will have some counteracting effect when combined, but the relationship between these is not exactly reciprocal.

4.3 RESULTS

Significant environmental variables to model the intensity of Baird's tapir (i.e. 95% CI not overlapping 0; Table 4.1) included water occurrence (1.44, SE=0.45), forest cover (1.33, SE=0.20), annual precipitation (1.20, SE=0.34), EVI (-0.56, SE=0.17), and temperature seasonality (-0.44, SE=0.13). Significant quadratic terms for water occurrence (-0.84, SE=0.34), temperature seasonality (-0.39, SE=0.12) and annual precipitation (-0.81, SE=0.26) indicated possible non-monotonic relationships for those variables. A 3-dimensional response curve for the interaction between forestcover and EVI demonstrated highest lambda values when high forest cover was combined with low EVI (Figure 4.2). Significant variables to model the detectability of tapir using presence-only (PO) data (Table 4.1) included presence in a protected area (1.35, SE=0.24) and distance to roads (-0.43, SE=0.15). With the SO data, the dummy variable for surveys targeting tapir was significant (1.03, SE=0.44), while forest loss was nearly significant (-0.30, SE=0.15).

| Coefficient | Estimate | SE | z-score | p-value |
|------------------------------------|-----------------|-----------|----------------|----------------|
| beta0 | -3.237 | 0.316 | 10.252 | 0.000 |
| x.chelsa_temp_seasonality | -0.444 | 0.127 | 3.500 | 0.000 |
| x.chelsa_precip_seasonality | 0.282 | 0.192 | 1.472 | 0.141 |
| x.chelsa_max_temp_warmest_month | 0.194 | 0.187 | 1.037 | 0.300 |
| x.chelsa_annual_precip | 1.202 | 0.344 | 3.490 | 0.000 |
| x.chelsa_temp_seasonality_sq | -0.392 | 0.119 | 3.287 | 0.001 |
| x.chelsa_precip_seasonality_sq | -0.208 | 0.127 | 1.637 | 0.102 |
| x.chelsa_max_temp_warmest_month_sq | 0.042 | 0.038 | 1.103 | 0.270 |
| x.chelsa_annual_precip_sq | -0.815 | 0.255 | 3.189 | 0.001 |
| x.forestcover2000 | 1.335 | 0.202 | 6.591 | 0.000 |
| x.distancePA | -0.088 | 0.155 | 0.570 | 0.569 |
| x.EVI | -0.564 | 0.168 | 3.357 | 0.001 |
| x.EVI_sq | -0.126 | 0.129 | 0.972 | 0.331 |
| x.water_occurence | 1.441 | 0.450 | 3.199 | 0.001 |
| x.water_occurence_sq | -0.842 | 0.338 | 2.495 | 0.013 |
| x.forestloss | -0.095 | 0.071 | 1.329 | 0.184 |
| x.road_length | -0.287 | 0.207 | 1.388 | 0.165 |
| x.fire_density | -0.211 | 0.224 | 0.941 | 0.347 |
| x.mean_slope | -0.181 | 0.229 | 0.794 | 0.427 |
| x.forestcover2000_EVI | 0.133 | 0.179 | 0.745 | 0.456 |
| alpha0.po | -5.410 | 0.367 | 14.748 | 0.000 |
| alpha0.so | -2.373 | 0.673 | 3.526 | 0.000 |
| w.po.forest | 0.276 | 0.262 | 1.056 | 0.291 |
| w.po.protected | 1.352 | 0.245 | 5.525 | 0.000 |
| w.po.road_distance | -0.434 | 0.153 | 2.833 | 0.005 |
| w.po.road_distance_sq | 0.057 | 0.049 | 1.166 | 0.243 |
| w.po.mean_slope | -0.106 | 0.216 | 0.490 | 0.624 |
| w.so.forestcover2000 | 0.469 | 0.518 | 0.905 | 0.366 |
| w.so.distancePA | -0.043 | 0.325 | 0.131 | 0.895 |
| w.so.road_distance | -0.667 | 0.475 | 1.406 | 0.160 |
| w.so.road_distance_sq | -0.017 | 0.207 | 0.083 | 0.934 |
| w.so.mean_slope | -0.002 | 0.171 | 0.011 | 0.991 |
| w.so.forestloss | -0.296 | 0.155 | 1.908 | 0.056 |
| w.so.road_length | 0.165 | 0.553 | 0.298 | 0.766 |
| w.so.fire_density | 0.453 | 0.852 | 0.532 | 0.595 |
| w.so.sampling.Tapir | 1.027 | 0.444 | 2.313 | 0.021 |
| w.so.sampling.Cat | 0.544 | 0.546 | 0.996 | 0.319 |
| w.so.sampling.OnTrail | 0.129 | 0.399 | 0.323 | 0.746 |

Table 4.1: Coefficient estimates, standard errors, z-scores, and p-values.

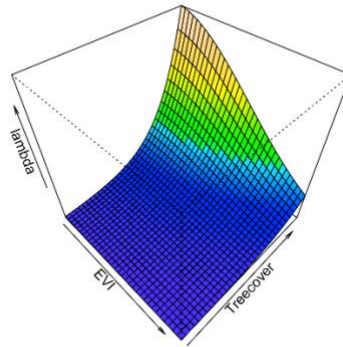


Figure 4.2: 3D response plot of intensity to EVI and forest cover.

Average estimated intensity from the model was 1.8 individuals/ 100 km², with a maximum of 46.4 individuals/100 km². The hotspot analysis identified 28 habitat patches with an estimated population of 10 or more individuals (Figure 4.3). These patches covered 23.1% of the land in the study area, and contained an estimated total of 10,343 individuals, roughly 61.3% of the total population estimated across the entire study area.

The three dispersal scenarios showed varying levels of connectivity (Figure 4.4). In the most restrictive case (max dispersal = 40 km), the network contains several disruptions in connectivity, with ten separated graph components. While the most generous dispersal scenario (max dispersal = 160 km), shows greater connectivity, though even it has breaks, as it contains three distinct components. Across all cases, a low degree of connectivity occurred at the extreme northwestern edge of the distribution, with completely isolated patches in Oaxaca (ID 3) and Chiapas (ID 4,5) Mexico. In the moderate scenario (max dispersal = 80 km), breaks in connectivity emerged across Honduras and at the Panama Canal.

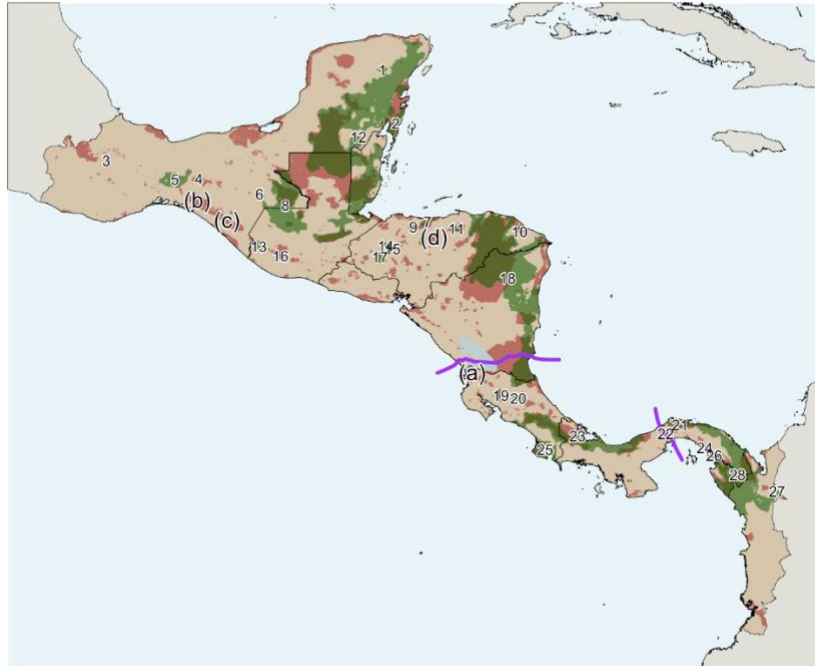


Figure 4.3: Map of habitat patches. Green areas are habitat patches. Brown areas are protected areas. Purple lines represent the Panama Canal and the planned path of the Nicaragua Canal. (a) - Santa Rosa, (b) - La Sepultura, (c) - El Triunfo, (d) - La Muralla.

Focusing on the indices calculated using the most generous dispersal scenario (to consider the impact of potential connections), the most vulnerable habitat patch was found in the northern Yucatan Peninsula (ID 1). This patch had a small estimated population (~11 individuals), is completely unprotected, and had a high level of isolation (i.e. farness) from other patches (Table 4.2). The top four most vulnerable patches also had low importance values, with small estimated populations (mean = 18.5 individuals), no confirmed tapir presence, and low contribution to connectivity. The two patches with the highest estimated

population also had the highest combined scores, Maya Forest (ID 12) and Nicaragua (ID 18). These patches also had an important role in connectivity.

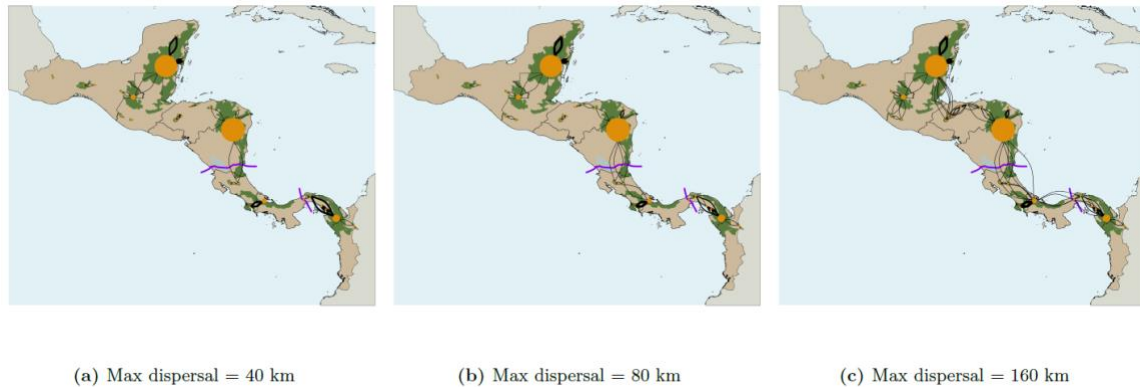


Figure 4.4: Maps of network connectivity. Thickness of edges indicates strength of flow between nodes. Size of nodes indicates estimated population size.

About half (~49.3%) of the total area covered by the habitat patches is under some form of protected status. Much of the remaining area occurs in indigenous territories with some level of authority over their lands. This includes Selva Zoque (ID 5), Nicaragua (ID 18), and between Chagres and Darien National Parks in Panama (ID 28). This last patch includes the Darien region of northern Colombia, a significant area that is unprotected and not under indigenous control. However, this area has remained largely undeveloped due to the difficulty of the environment and armed conflict between the government and paramilitary groups (Giroto, 2002; Grajales, 2013).

IUCN recently produced a map of indigenous territories in Central America (IUCN-ORMACC, 2016). Focusing on these countries (i.e. excluding Colombia and Mexico), 88.1% of tapir patches are either protected or in an indigenous area. Of the 41.0% of area that is not protected, 71.0% falls within indigenous areas. This includes the Qeqchi (Kekchi) in Guatemala, Miskito-Miskitu and Ulwa in northern Nicaragua (Wawashang),

Ngabe and Bugle in Golfo de los Mosquitos Panama, and Kuna and Embera-Wounaan near Chagres and Darien, Panama.

| ID | Occu | Pop | Bet | Comp | Impor | %Unp | F. loss | Far | Vuln | Comb |
|-----------|-------------|------------|------------|-------------|--------------|-------------|----------------|------------|-------------|-------------|
| 18 | 1 | 3515 | 0.93 | 0.67 | 0.97 | 0.36 | 0.87 | 0.12 | 0.41 | 1.37 |
| 12 | 1 | 3322 | 0.76 | 1.00 | 1.00 | 0.53 | 0.47 | 0.12 | 0.35 | 1.35 |
| 8 | 1 | 762 | 0.40 | 0.67 | 0.62 | 0.64 | 0.61 | 0.17 | 0.66 | 1.28 |
| 23 | 1 | 724 | 1.00 | 0.67 | 0.77 | 0.43 | 0.15 | 0.12 | 0.45 | 1.23 |
| 25 | 1 | 240 | 0.00 | 0.00 | 0.29 | 0.57 | 0.24 | 0.72 | 0.74 | 1.03 |
| 1 | 0 | 11 | 0.00 | 0.00 | 0.00 | 1.00 | 0.68 | 0.64 | 1.00 | 1.00 |
| 28 | 1 | 1098 | 0.14 | 0.33 | 0.48 | 0.68 | 0.12 | 0.12 | 0.48 | 0.97 |
| 5 | 1 | 147 | 0.00 | 0.00 | 0.28 | 1.00 | 0.19 | 0.00 | 0.65 | 0.93 |
| 7 | 1 | 34 | 0.93 | 0.00 | 0.52 | 0.06 | 0.12 | 0.12 | 0.39 | 0.91 |
| 26 | 0 | 28 | 0.00 | 0.00 | 0.00 | 0.99 | 0.87 | 0.12 | 0.90 | 0.90 |
| 22 | 1 | 11 | 0.27 | 0.00 | 0.34 | 0.62 | 0.08 | 0.12 | 0.55 | 0.89 |
| 24 | 0 | 16 | 0.00 | 0.00 | 0.00 | 0.74 | 1.00 | 0.12 | 0.86 | 0.86 |
| 11 | 1 | 28 | 0.00 | 0.00 | 0.27 | 0.09 | 0.67 | 0.12 | 0.56 | 0.83 |
| 9 | 1 | 13 | 0.18 | 0.00 | 0.32 | 0.22 | 0.30 | 0.12 | 0.49 | 0.81 |
| 6 | 0 | 19 | 0.00 | 0.00 | 0.00 | 1.00 | 0.33 | 0.31 | 0.79 | 0.80 |
| 21 | 1 | 60 | 0.27 | 0.00 | 0.35 | 0.11 | 0.10 | 0.12 | 0.40 | 0.74 |
| 2 | 0 | 81 | 0.00 | 0.00 | 0.01 | 0.32 | 0.14 | 1.00 | 0.73 | 0.74 |
| 20 | 1 | 25 | 0.00 | 0.00 | 0.27 | 0.29 | 0.04 | 0.13 | 0.44 | 0.71 |
| 17 | 0 | 34 | 0.09 | 0.00 | 0.03 | 0.64 | 0.45 | 0.12 | 0.66 | 0.69 |
| 27 | 0 | 21 | 0.00 | 0.00 | 0.00 | 0.94 | 0.08 | 0.23 | 0.68 | 0.68 |
| 13 | 0 | 28 | 0.00 | 0.00 | 0.00 | 0.87 | 0.19 | 0.17 | 0.67 | 0.67 |
| 10 | 0 | 16 | 0.00 | 0.00 | 0.00 | 0.70 | 0.06 | 0.44 | 0.67 | 0.67 |
| 15 | 0 | 11 | 0.64 | 0.00 | 0.17 | 0.07 | 0.23 | 0.12 | 0.43 | 0.60 |
| 4 | 1 | 10 | 0.00 | 0.00 | 0.27 | 0.00 | 0.06 | 0.00 | 0.32 | 0.59 |
| 3 | 0 | 16 | 0.00 | 0.00 | 0.00 | 0.85 | 0.10 | 0.00 | 0.59 | 0.59 |
| 14 | 0 | 12 | 0.20 | 0.00 | 0.06 | 0.26 | 0.26 | 0.12 | 0.49 | 0.55 |
| 16 | 0 | 16 | 0.00 | 0.00 | 0.00 | 0.18 | 0.38 | 0.17 | 0.52 | 0.52 |
| 19 | 0 | 55 | 0.00 | 0.00 | 0.00 | 0.53 | 0.08 | 0.12 | 0.52 | 0.52 |

Table 4.2: Vulnerability, importance, and combined indices for conservation prioritization in the maximum dispersal = 160 km scenario.

4.4 DISCUSSION

All the significant parameter estimates from the model (and even those that were close to significant), had the expected relationship with tapir intensity. In previous studies, the negative relationship with EVI was unexpected (Schank et al., 2017). However, showing the interaction between EVI and forest cover, it was clear that the model demonstrates evidence of tapir preference for secondary forest, as these are areas which would show up with high forest cover, but relatively lower biomass. The hump-shaped relationship with annual precipitation is likely what drives the low estimates of tapir presence in the Chocó, where rainfall is much higher than the rest of the study area (mean rainfall twice as high as any other portion of the study area). The Chocó is one of the wettest areas globally (Schwerdtfeger, 1976).

The results from this research almost certainly represent a best-case scenario in the distribution and extent of habitat patches for Baird's tapirs. Though there are a few underestimated areas that do not show up as suitable despite confirmed tapir occurrence (e.g. dry forests of Santa Rosa Costa Rica, and all of the Sierra Madre de Chiapas in S Mexico, between La Sepultura in the N and El Triunfo in the S), the overall estimated intensity and resulting habitat patches likely represent an overestimate of the species' true distribution. This can be attributed to multiple factors, including that forest cover from the year 2000 is used in the model (even though forest loss in the intervening years is also used), and land cover is changing rapidly across much of Mexico and Central America.

In addition, the connectivity between cores is likely overestimated in many cases. Very little is known about the long-distance dispersal capabilities for this species. The ability of Baird's tapir to disperse even 40 km is questionable, especially across the low

suitability habitat it would encounter moving between most patches. Even in the most optimistic scenario, there is an alarming lack of connectivity and suitable habitat, with tenuous connections across Honduras, and a clear break in Mexico. Genetic data from jaguars revealed limited connectivity between La Selva Maya and Honduras (Wultsch et al., 2016), and these large carnivores likely have much greater dispersal capacity compared to tapirs. In addition, this analysis does not incorporate potential barriers to dispersal, such as the Panama Canal Zone and the increasing numbers of highways across southern Mexico and Central America. The planned Nicaragua Canal could present an additional dispersal barrier for the species (Jordan et al., 2016). Assuming this canal is built, and both it and the Panama Canal act as barriers to dispersal, while also including the breaks in the network that already exist, there is a potential for five major areas of tapir habitat to become isolated from one another. This would have significant adverse impacts on the long-term survival of the species. Isolating populations makes local extinctions more likely, due to reduced genetic diversity and interruption of source-sink/meta-population dynamics (Pardini, Bueno, Gardner, Prado, & Metzger, 2010).

Considering these results, the following conservation actions for the species are recommended: 1) protect existing habitat, 2) re-establish and strengthen habitat connectivity, 3) conduct additional surveys on tapir occurrence, and 4) collect additional telemetry data.

1) Protect existing habitat

The results showed that about half of the remaining habitat for Baird's tapir can be found in protected areas. Though these areas are protected (at least on paper), many of them still experience substantial deforestation inside park boundaries (Bray et al., 2008; Joppa, Loarie, & Pimm, 2008; Watsa, 2014). In fact, even when deforestation is not evident from satellite imagery, these protected areas still experience significant levels of poaching

and smaller scale deforestation that can have a strong negative impact on endangered wildlife (Peres, Barlow, & Laurance, 2006). Of the nearly 312 protected areas that intersect tapir habitat, 37 have experienced 5% or more deforestation over the last 13 years, with almost half of these found in Guatemala (Table 4.3). This could help explain why distance to/within protected areas was not a significant variable in the model (Table 4.1).

The government agencies tasked with monitoring and enforcing laws within protected areas often have too little money or power to do their jobs effectively (Balmford & Whitten, 2003; Bruner, Gullison, & Balmford, 2004). Working with these agencies to provide the training, support, and resources to improve their effectiveness should be a priority. High-level lobbying by nationals to encourage governments to prioritize protected area management and the sentencing of environmental criminals is needed in most countries across Baird's tapir range. In Nicaragua, for example, anecdotal data suggests that multiple tapirs trafficked in the capital city of Managua in 2017 were not confiscated despite the Ministry of Natural Resources being made aware of this environmental crime. Likewise, environmental criminals destroying core protected areas have not been held legally accountable in Nicaragua (Salazar, 2018). The Spatial Monitoring and Report Tool (SMART) is an open-source software program designed to facilitate the collection, analysis, and communication of data from protected area monitoring and vigilance efforts, which has proven a useful tool in many countries globally (Critchlow et al., 2017).

In addition to protected areas, much of the remaining habitat for Baird's tapir also occurs in indigenous areas, where indigenous communities often are more effective at conserving wildlife habitat than State government agencies (Bray et al., 2008; Porter-Bolland et al., 2012). The exclusionary practices of protected areas are based on the "pristine myth" (Denevan, 1992), or the misconception that displacing people from "natural" areas is good for conservation (Schwartzman, Moreira, & Nepstad, 2000).

| ID | Name | Forest Loss | Area km² | Location |
|-----------|---|--------------------|----------------------------|-----------------|
| 18 | Cerro Bana Cruz | 0.16 | 270.30 | NI-AN |
| 12 | Santa Rosa | 0.16 | 5.08 | GT-IZ |
| 18 | Cerro Wawashang | 0.16 | 2244.64 | NI-AS |
| 12 | Mountain Pine Ridge | 0.13 | 432.74 | BZ-CY |
| 12 | Yaxhá | 0.13 | 1.98 | GT-PE |
| 18 | Punta Gorda | 0.12 | 636.94 | NI-AS |
| 12 | Candilejas | 0.12 | 0.70 | GT-IZ |
| 18 | Cerro Silva | 0.10 | 2921.97 | NI-AS |
| 18 | Laguna de Bacalar | 0.10 | 72.70 | Not Reported |
| 12 | Parque Nacional Yaxhá-Nakum-Naranja | 0.09 | 4874.45 | GT-PE |
| 8, 12 | Maya | 0.09 | 20382.32 | GT-PE |
| 12 | Mischner & Bowen Reserve | 0.08 | 6.97 | BZ-BZ |
| 12 | Cástulo | 0.08 | 3.59 | GT-IZ |
| 18 | Cerro Cola Blanca | 0.08 | 104.91 | NI-AN |
| 12 | Laguna del Tigre | 0.08 | 2895.07 | GT-PE |
| 12 | Cerro San Gil | 0.08 | 437.53 | GT-IZ |
| 8, 12 | Maya | 0.08 | 7549.04 | GT-PE |
| 12 | El Higuero | 0.08 | 6.78 | GT-IZ |
| 8 | Laguna Lachuá | 0.07 | 512.17 | GT-AV |
| 18 | Sierra de Río Tinto | 0.07 | 880.71 | HN-CL |
| 12 | Grants Works | 0.07 | 32.01 | BZ-SC |
| 18 | Patuca | 0.07 | 3755.98 | HN-OL |
| 18 | Cerro Saslaya | 0.07 | 631.09 | NI-JI |
| 28 | Canglón | 0.07 | 320.87 | PA-5 |
| 12 | Montañas Mayas Chiquibul | 0.07 | 615.07 | GT-PE |
| 18 | Tawahka Asangni | 0.07 | 2508.79 | HN-OL |
| 12 | Caracol | 0.06 | 103.19 | BZ-CY |
| 8 | Sierra del Lacandón | 0.06 | 2002.87 | GT-PE |
| 8 | Empalme Santa Felicitas | 0.06 | 42.72 | MX-CHP |
| 12 | Swasey Bladen | 0.06 | 59.81 | BZ-TOL |
| 18 | Sureste de Nicaragua | 0.06 | 18315.05 | NI-AS |
| 12 | Laguna Ik | 0.06 | 287.82 | MX-CAM |
| 12 | Río Sarstun | 0.06 | 324.13 | GT-IZ |
| 18 | Fortaleza la Inmaculada Concepción de María | 0.06 | 35.82 | NI-SJ |
| 12 | Otoch Ma'ax Yetel Kooch | 0.06 | 53.72 | MX-ROO |
| 13 | Buenos Aires | 0.05 | 8.75 | GT-RE |
| 12 | Santa Elena | 0.05 | 1.41 | GT-IZ |

Table 4.3: Forest loss rates (2000-2014), protected status, area, and location of protected areas that intersect tapir habitat patches.

Indigenous people are possibly the most important allies that conservationists have in protecting habitat for endangered species (Schwartzman, Nepstad, & Moreira, 2000).

In some high priority patches, State governments have also granted legal tenure and certain levels of legal autonomy to indigenous communities, including in the Honduran Moskitia where 11 indigenous territories were granted titles to approximately 1.4 million hectares in 2016 and the 23 indigenous territories that cover nearly the entirety of Nicaragua's Caribbean Coast. However, despite legal recognition of their rights to their ancestral lands, these communities are often under pressure from illegal colonists driving cattle ranching and agricultural frontiers, lack adequate support from State governments, and lack the necessary resources to monitor and enforce their land rights (Stocks, McMahan, & Taber, 2007). Thus, similar strategies that would enhance funds, staff, technology, and training for indigenous territories will be crucial to protecting these areas. In many cases, protected areas and indigenous communities overlap. In this case, some suggest that management responsibilities should be passed from federal institutions to local communities (Bonham, Sacayon, & Tzi, 2008). All strategies implemented in indigenous territories must abide by international instruments designed to protect the traditions, cultures, languages, and self-determination of indigenous peoples (i.e. ILO Convention 169).

Another strategy to maintain existing habitat for Baird's tapir is through Payments for Ecosystem Services (PES) in communities that maintain tapir habitat or critical areas for connectivity. PES is a conservation scheme that provides economic compensation to local communities for conserving their land with natural forest, and studies have shown that this conservation scheme plays a crucial role in protecting the habitat and corridors of other emblematic species of Central America, such as the jaguar (de la Torre et al., 2017).

2) Re-establish and strengthen connectivity

Areas with poor connectivity include southwestern Mexico and Honduras, where the distance between patches is too far, as well as areas with major dispersal barriers (e.g. Panama Canal). Strategies to promote connectivity could include the purchase of private land to preserve or restore habitat that can act as stepping stones for the species dispersing between the larger patches (Worboys, Francis, & Lockwood, 2010). Another strategy is to work with large agricultural or forestry operations that may lie along movement corridors (Meijaard et al., 2005). To plan for both of these actions, detailed information on land ownership and concessions is necessary (K. Miller, Chang, & Johnson, 2001).

The path of the proposed Nicaragua Canal could create another disruption in connectivity for the species and destroy a large portion of an important habitat patch (ID 26). Some strategies have been suggested to reduce the impact of this canal (Jordan et al., 2016). Even if such strategies were employed, the canal would present a major barrier to dispersal, as evidence suggests the Panama Canal already does for tapir (N. F. V. Meyer et al., 2015/7; N. Meyer et al., 2013; Moreno, 1993). Major roads also act as dispersal barriers for large mammals like the tapir, with potential negative genetic effects and influences on overall species health (Holderegger & Di Giulio, 2010). To improve the connectivity across the species range, it is necessary that major roads that interrupt the connectivity for Baird's tapir populations include mitigation measures to avoid wildlife collisions. Some projects have focused specifically on tapir-mortality when dispersing across roads (e.g. The Belize Tapir Project, Lowland Tapir Conservation Initiative). Wildlife under- and overpasses, installation of road signs and speed bumps, are potential strategies to promote connectivity for tapir populations that require further investigation, however even the best mitigation efforts will not overcome massive road development, meaning that conservationists must be engaged with development plans at the local level to discourage road construction in highly vulnerable or important tapir patches (Glista, DeVault, & DeWoody, 2009).

3) Conduct additional surveys to estimate tapir occupancy and abundance

Ten of the patches identified in this analysis have no confirmed tapir presence. Most of these patches are small (mean estimated population size = 30, max = 61), and several of them are very close to larger patches, suggesting a low level of importance. However, surveys in some small patches should be prioritized due to their location in highly fragmented landscapes where maintaining connectivity is critical. First, there are several small patches throughout Honduras that should be prioritized (IDs 7, 12, 13, 21, & 22). Although presence data exist in some of these patches, this area is so crucial to the connectivity of the species due to its central location in the network that all these patches could represent important stepping stones. There also are several nearby smaller patches (< 10 individuals) and protected areas that could serve this purpose, and that have confirmed tapir presence, including La Muralla, where a recent survey of limited effort detected the species (Sagastume-Espinoza & Romero, 2017).

The second area of focus for future tapir surveys should be the fragmented landscape of northern Costa Rica (e.g. IDs 24 & 27). As with Honduras, there are smaller patches (< 10 individuals) that were not identified in the hotspot analysis, but which could act as stepping stones for the species dispersing between larger patches. This includes Volcan Barva, which was the subject of an extensive camera trap survey over almost a decade and carried out by the Tropical Ecology Assessment and Monitoring (TEAM) Network (Ahumada, Hurtado, & Lizcano, 2013). Roughly 25% of the 400 camera trap histories from this study had tapir detections. These areas need surveys not just to confirm or deny presence of the species, but also to assess the size of their populations, potentially using Spatial Capture-Recapture for unmarked populations (Ramsey et al., 2015).

In addition to surveys to confirm presence in important patches, the distribution model would benefit from surveys conducted in habitats that are underrepresented in the

presence data. Most of the available presence data is from the fringe of habitat patches. There are few surveys from low suitability areas, because researchers targeting tapir or other endangered species often want to ensure they have some detections. On the other hand, there are also few surveys from deep within habitat cores, likely due to problems with accessibility. Targeting surveys to these two types of habitats, would likely improve model accuracy and precision.

4) Collect additional telemetry data

This research was based on limited knowledge about the dispersal capability of Baird's tapir. Small changes in home range size and long-distance dispersal would have a large impact on the results. Estimates of home range size for the species range anywhere from 1 to 24 km², though these are based on very small sample sizes (i.e. 1-5 individuals) and simple methods (Foerster & Vaughan, 2002; Reyna-Hurtado et al., 2016a). Changes to the home range estimate would impact estimates of species density, which would alter the patches that are included in the analyses. More accurate population estimates could also be used in population viability assessments. The estimate of maximum dispersal distance uses a method of extrapolation based on home range size and would also be affected by more accurate home range estimates. However, it would be best to incorporate high resolution GPS telemetry data to directly estimate maximum dispersal for the species.

4.5 CONCLUSION

The methods and approach used in this analysis can be applied to other endangered large mammals, with little adjustment, when sufficient data are available to run the ISDM, and there is some knowledge of species movements or home ranges. However, data limitations regarding predictors used in the ISDM or telemetry to make inferences about

ESA and dispersal, will lead to biased results. Despite these concerns, even relative rankings of habitat patches are useful for range-wide conservation planning.

There are several similar studies that combine knowledge of species distributions with connectivity analyses to develop a set of priority areas for conservation (A. Rabinowitz & Zeller, 2010). Yet, the approach developed in this analysis paves the way for more accurate and rigorous range-wide conservation plans. This can be achieved by building on existing methods, collecting better (i.e. more proximal and/or more accurate) predictors, and additional presence data from under-sampled habitats. Specifically, the ISDM used in this research would benefit from an additional parameter to account for species movement, corresponding with the sigma parameter in Spatial Capture-Recapture models (J. A. Royle et al., 2013; Schank et al., 2018). This would eliminate the need to determine the ESA through ad hoc methods. If the model is formulated in a Bayesian framework, telemetry data could also be used as prior information for the movement parameter. Ideally, this parameter should be allowed to vary across the landscape, perhaps through a linkage with the intensity estimated by the model, which would allow for more accurate estimates of density from the ISDM.

The methods used in this research would also benefit from estimation of actual dispersal probability, rather than using relative measures of flow between patches. However, this would require a spatiotemporal dispersal kernel that could provide an estimate of dispersal probability over time. Combining more accurate population estimates with dispersal probabilities would facilitate using these methods with population viability assessments. These methods could then be designed to create probabilities of extinction for each patch, and the resulting impact on the overall viability of the species. This would represent an improvement over this analysis, which creates only relative rankings between habitat patches.

CHAPTER 5: Next steps - Linking SDMs, Connectivity analyses, and PVAs

5.1 INTRODUCTION

By investigating new SDM methods and linking those outputs with a connectivity analysis in a range-wide conservation plan for Baird's Tapir, this research established a path forward, and outlines some future steps to take in developing the methods for these types of plans. In conducting this research, I uncovered multiple questions that could use more interrogation, and here I outline these issues in more detail, while also proposing some possible solutions and how to achieve them. I also reflect on the overall process, and what could be improved from that standpoint. This research and the future steps follow well from the frontier's in SDM research outlined in the conclusion of Chapter 1 (here I list them again and how they are addressed by this dissertation and final chapter):

- 1) Creating more realistic models that incorporate ecological knowledge (M. Austin, 2007; Higgins et al., 2012)
 - a. I propose strategies to test ISDMs on more realistic data simulations (i.e. by simulating individual movements of species)
 - b. I propose combining hotspot analysis with biologically relevant thresholds to determine habitat patches for species more realistically
- 2) Unifying disparate theories and multiple methods that address the same (or related) questions (McDonald, 2013; McGill, 2010)
 - a. ISDMs as a framework combine aspects of Species Distribution Models (SDMs), Occupancy Models, and Point Process Models (PPMs)
 - b. I propose further incorporating a spatial scale parameter in ISDMs (following the example of Spatial Capture-Recapture models)

- 3) Working with big data to develop not only the tools that can handle large quantities of data, but also how to use these new datasets to ask questions that simply were not possible before (Stephens-Davidowitz, 2017)
 - a. Data integration (through ISDM) is one way to utilize big data
- 4) Open access and democratization of research (Jetz et al., 2012)
 - a. I propose implementing these workflows in Map of Life

5.2 LINKING SDMS, CONNECTIVITY ANALYSES, AND PVAS

The end vision of developing this research is to create a chain of analyses where an SDM feeds into a connectivity analysis, which both then are used to provide inputs to a Population Viability Analysis (PVA)⁴. PVAs are models which estimate the probability of extinction for a population over a certain number of years, under various conditions, and require inputs on population size and immigration from outside the population, among other demographic parameters. Ideally, in a RWCP context, we would know the probability of extinction for each habitat patch, and how that affects the overall survivability of the species, which could then be tested under different scenarios (e.g. climate change predictions, forest loss projections, or potential reintroduction scenarios). However, before this can be done, we first need to be more confident in the estimates that feed into the PVA⁵.

⁴ Research combining both SDMs with graph theory, and SDMs with PVAs, has already been underway for over a decade (Foltete, Clauzel, Vuidel, & Tournant, 2012; Franklin, 2010b), but the combination of all three is less explored.

⁵ Would also be useful to have the ability to run PVA in R so that sensitivity analysis of the models could be more easily conducted.

5.2.1 Improving ISDM estimates

In Chapter 2, the ISDM produced total population estimates which were much higher than expected using expert opinion as a baseline. Though it is true that expert opinions can be inaccurate, the distance between these estimates (ISDM and expert) was more than an order of magnitude, which raised a red flag. The reviewers of this manuscript when it was submitted to Diversity & Distributions, and the developer of the ISDM framework, suggested that duplicated observations could be the reason for the inflated population estimate. Thus, I developed a spatial thinning process to reduce the possibility that an individual could be detected at more than one site. However, the population estimates did not change significantly after thinning the data. For this reason, I investigated the sensitivity of the ISDM to some of the modelling decisions which were easily controllable.

In the sensitivity analysis of Chapter 3, the relationship between site area and the total population estimate was striking. I found that the determination of the site area⁶ in the model is what drives the magnitude of the population estimate. Because the ISDM is based on a Poisson point process model (PPM), unlike typical occupancy models, the sampling area is explicitly included as an input in the model. Rereading the original paper that introduced the model, there is one sentence about how the model assumes that individuals are detected near their activity centers (Dorazio, 2014b):

“I assume that individuals are detected at their activity centers, recognizing that this assumption is likely to be violated if individuals are highly mobile or possess large territories (that is, large relative to the size of region B).” (p.1474)

⁶ In Chapter 2, I used a site area of 1 km² to match the resolution of the environmental predictors used in the model. There was always some anxiety about this decision, because it seemed arbitrary. However, my mind was somewhat at ease, since 1 km² was generally the accepted home range size for the species (though new data that contradicted this was starting to emerge).

When considering how this would affect the model estimates, the concept of Effective Sampling Area (ESA) from Capture-Recapture (CR) provided some insight. Because CR models explicitly estimate species density, they have focused on the issue of sample area. Using camera traps (or other point sample techniques) to study mobile species, it is recognized that the actual area sampled is unknown (M. Kery & Royle, 2015; J. A. Royle & Dorazio, 2008). There are two ways to deal with this: 1) consider that the estimates from the model are relative indices of abundance, 2) use an ad-hoc correction method by adding a buffer around the sites that is equivalent to the mean distance moved during sampling (or sometimes the radius of a home range, if data on the movement is not available). In the latter option, the modeler is correcting for the ESA.

Spatial Capture-Recapture (SCR), a variation of the CR model that utilizes a PPM, has an additional model parameter that accounts for individual movements of the species (i.e. known as the spatial scale parameter, σ). This parameter is estimated by having cameras placed close enough to get detections of the same individual at multiple sites (the same issue I was trying to avoid by using the thinning process⁷). In this way, the model uses the spatial autocorrelation of detections to estimate the spatial scale parameter. It is possible that adding this parameter to the ISDM model could allow for more accurate estimates of intensity, by directly estimating the effective sampling area.

Typically, in both SCR and CR, it is necessary to identify at least some individuals of the population (hence “capture-recapture” or “mark-resight”). For a species like Tapir, some have tried to use CR techniques (González-Maya et al., 2012), but the identifications are difficult due to the lack of markings which can help identify individuals in photos (which is possible with Tigers or Jaguars) (Oliveira-Santos et al., 2010; Tobler et al., 2014).

⁷ So, if this same structure could be implemented in an ISDM, we wouldn't need to throw out any data.

However, adaptations to SCR have been developed which do not require any individual identification (Ramsey et al., 2015), and some are currently being developed for partial identification (Augustine et al., 2017). These unmarked SCR models do have a problem with convergence, and are not yet developed to use environmental predictors, which would be needed to make predictions.

When thinking about implementing a similar structure in the ISDM, we must ask: Why not just use the SCR? In these previous chapters, it became obvious that there was some benefit of including PO data in the ISDM, not just in that it filled in geographic sampling gaps, but also that it provided additional information to estimate model parameters (when there was some problem with convergence after thinning and reducing the data). This was because additional information was added by using the PO data, and also including the background environment in the model. That's thousands and thousands of background points, which improved parameter precision. However, as with other models that use this kind of background or pseudo-absence framework, it is necessary to be careful how to interpret the model estimates. Ideally, the use of SO data reduces this issue, in that there is some idea of the species prevalence.

To implement this combination of ideas from the ISDM and SCR models, the first step is to convert the ISDM to a Bayesian framework. This will allow more flexibility and ease of coding additional structure to the model, as well as allowing for prior information (such as from telemetry data), and more easily giving estimates of model uncertainty.

It is also increasingly important to do a test of the ISDM, and any development thereof, with a simulated dataset. However, it is crucial to do the simulation in a way that uses a realistic sample (i.e. how many data points, and how biased), and realistically portrays individual movements. As was seen in this research, movement has a very strong effect on model estimates. Two R packages have functions that could be useful to simulate

this type of data: adehabitatLT and moveHMM (Calenge, 2011; Michelot, Langrock, & Patterson, 2016).

It became clear during this research that species do not exhibit consistent movement patterns across space (or even across sex and age, for that matter). A model which incorporates movement through a spatial scale parameter should allow this parameter to vary across space. Perhaps this could be done by linking the parameter to the relative suitability based on the environmental predictors (irrespective of the intensity, and the intercept, which will rely on the ESA)⁸.

5.2.2 Improving connectivity analyses

Similar to the challenges with the ISDM, there were issues with the connectivity analysis implemented in Chapter 4. In that analysis, a dispersal kernel was used to transform distances between patches into probability of dispersal. This was done because it was assumed that an individual is more likely to disperse closer to the origin, as distance increases it becomes less likely that the individual will disperse that far (up to some maximum dispersal distance⁹). However, this kernel was purely spatial, and did not have any information about time. If spatio-temporal kernels exist, we could estimate a rate of dispersal over time. As it was done in this research, the flow is interpreted simply as a relative estimate, so it could only be used to calculate relative indices of flow between patches, not as an input to PVA.

⁸ As seen in Chapter 3, estimates of coefficients that describe the species response to the environment do not vary across changes in the site area, it is only the intercepts of the model which change.

⁹ This is not necessary but does speed up the calculations by eliminating the need to do the flow calculation for some pixels. Same could be said for taking a patch approach. We ignore all those pixels outside of the patches.

Another issue with the connectivity analysis is that I did not have good data for the dispersal capabilities of the species. At first, I used a method of extrapolating from the home range size (Bowman et al., 2002), but this seemed to vastly overestimate what experts expected for the species, and what limited data showed. For this reason, I tried three different dispersal scenarios, and even the most generous scenario showed isolated habitat patches. Dispersal data for juveniles is especially lacking, and these individuals will likely disperse greater distances as they seek to establish a new home range. This data is harder to collect, because juveniles are still growing, and thus the collars could cause health problems (including death) unless they are designed to be expandable.

Finally, rather than using least cost distances, it would be better to use an accumulated cost distance. This would account for variation in dispersal with habitat quality or a movement cost surface (including dispersal barriers), as some landscapes will be more or less stressful to move through. In between many of the patches identified in this research, the matrix habitat appears to be of very poor quality. It is unlikely to expect a tapir to disperse dozens of kilometers across highly disturbed habitat, with multiple road crossings, etc. However, there are challenges with how to relate this accumulated distance to maximum dispersal or use it in a dispersal kernel (i.e. what are the units of the accumulated cost?). Related to these challenges with cost distance, it is necessary to do more research on the best way to convert the ISDM intensity surface into a conductance or resistance surface used in connectivity analysis (in Chapter 4, untransformed intensity was used as a conductance surface).

It is possible that some methods already exist to solve these problems. So, the first step is a literature review to identify those solutions if they exist. If not, it will be necessary to develop new methods. In addition, additional telemetry data will provide valuable prior

information on how to parameterize the dispersal kernel¹⁰. Genetic data would also be incredibly useful to verify connectivity analyses, and see which patches are connected by individuals dispersing between them. The most recent range-wide genetic analysis for Baird's tapir was completed almost 15 years ago (Norton & Ashley, 2004) - an update would be helpful. Collecting both types of data is costly, which partially explains the lack of data available so far. However, there are good projects to follow in Brazil and Costa Rica that are focused on studying the road ecology of tapirs and collecting DNA samples from tapir that are killed in collisions with vehicles. As well as studies in Brazil and Nicaragua which are focused on collecting both DNA and telemetry data of tapir tagged with GPS collars. The results of these studies should become available soon, which will greatly inform the RWCP effort for Baird's tapir.

5.2.3 Converting ISDM outputs to habitat patches

One requirement of the methods used in this research, is to convert the resulting intensity surface to a map of habitat patches. Normally this would be done by converting the continuous prediction to a binary map using some threshold that is determined from an accuracy metric (Liu et al., 2005). Often accuracy is calculated on some test data withheld from fitting the model. However, if the data we start with is spatially biased (as we admit when designing models that can account for this), then how can we expect this very same data to adequately address model accuracy? In fact, I saw this issue when trying to assess the ISDM accuracy in Chapter 2. In the end, I decided to average the models and simply assess their accuracy visually. In Chapter 3, I discovered that some accuracy metrics designed for PO data might be less susceptible to sampling bias. However, even if we have the right data to assess accuracy, is there a more biologically meaningful threshold we

¹⁰ In addition to being useful prior information for the spatial scale parameter in an SCR model.

could select? If we can accurately estimate intensity by making some adjustments suggested earlier, then we can simulate a point pattern from that intensity surface. This would essentially be a map of all individuals on the landscape. From this, and some information about species movements and biology, we could use a clustering analysis like hotspots (i.e. Getis-Ord G_i^* , or local G , statistic) to come up with biologically meaningful habitat patches.

5.3 CONCLUSION

This dissertation was heavily focused on the methods, and so this chapter largely covers how the methods could be improved. However, it is important to not lose sight of the process-based goals, and how they could be also be enhanced. In the sections below, I outline a few important considerations to improve the process of conducting this type of research.

5.3.1 Including Stakeholders

This research was meant to inform conservation. Yet all too often, SDM exercises such as this one can fall short if they do not incorporate the appropriate stakeholders (Guisan et al., 2013). This research did include input and guidance from the Tapir Specialist Group, who could be seen as the conservation practitioners for which this analysis was intended. Indeed, the outputs from the process will go on to inform the Baird's Tapir Survival Initiative¹¹. However, I could have done a better job reaching out to a wider network of government officials, local communities, and other stakeholders. Admittedly,

¹¹ This project will create a regional team of researchers and conservationists working together to ensure the survival of Baird's tapir through capacity building, research, direct conservation, and a coordinated awareness-building campaign in Panama, Costa Rica, Nicaragua, Honduras, and Guatemala.

this would have increased the scope of the dissertation, and the continuation of this work provides more opportunity to fulfill this goal.

5.3.2 Public outreach, engagements, and sci-comm

In a similar sense, I could have done better reaching out to the public at large. We are entering a new era in science communication, with an important focus on achieving broader dissemination of research findings. Engaging the public keeps our research relevant, and can even provide additional funding, as it did for me with a Kickstarter Campaign I used to raise funds for field work during the summer of 2014 in Nicaragua. I regret that I did not do a better job of keeping these funders informed of the ongoing research, and even growing that community¹². Focusing on the methods is great, they need improvement, but we can't lose sight of why the work is being done, and what is necessary to achieve the desired outcomes - the long-term survival of endangered species. In all honesty, this goal is unachievable without broad support and continued funding.

5.3.3 Expanding this research – applying these methods to other species

The Map of Life is a project which seeks to bring together disparate data sets on species distributions so that they can be analyzed together. One of the goals of this project is to provide modeling tools and biodiversity products for a wide variety of uses, including academic research and conservation planning (Jetz et al., 2012). The model used in this dissertation is well-suited for this goal, as it combines two common data types¹³, and the methods used in this dissertation could be applied to many species. Even though tapirs are

¹² For example, by using Twitter more frequently to share my research (see #SciComm). Twitter is also a great way to find collaborators, build a research network, and keep up to date with current research trends.

¹³ Another model combines PO data with expert range maps (Merow, Wilson, & Jetz, 2017).

large megafauna, they are relatively understudied. The pieces of information needed to run this model include:

- 1) PO data: which are widely available.
- 2) SO data: the costs of collecting this data are decreasing (due to sampling methods like camera traps and citizen science sites like eBird).
- 3) Some idea of habitat preferences: which can be garnered from databases like the IUCN Red List, so that we know what environmental predictors to use in the model.
- 4) Some idea of home range size and dispersal abilities: which, if no data exist, could be inferred from allometric relationships or sister species.
- 5) Delineation of a study area¹⁴: arguably includes some of the most difficult decisions but could be automated based on presence data.

For all of this to happen, researchers need to feel more comfortable sharing data. The proper systems need to be in place to adequately recognize and motivate researchers to share data when it makes sense to do so (i.e. after they have published their analysis on it). There are already several options for publishing data sets (e.g. Zenodo, ESA's Ecological Archives). As well, Map of Life is not the only system seeking to democratize SDMs: see the R packages 'zoon' for reproducible and shareable SDMs (Golding et al., 2017) and 'sdm' (Naimi & Araújo, 2016), and the Biodiversity and Climate Change Virtual Library (<http://www.bccvl.org.au/>). This is an exciting time to be a modeler, but we must be careful and disciplined to extract the signal from the noise in the era of big data (Silver, 2012).

¹⁴ The delineation of the study area matters due to the background environment included in the model.

Appendix

| Type | Provider | Locations | Country | Citation |
|------|-------------------------------------|-----------|---------|---|
| PO | Francisco Botello | 15 | MX | 2015-2016: (Vazquez-Camacho et al., in prep); 1991-2006: (Colección de Fotocolectas Biologicas, IBUNAM http://unibio.unam.mx/minero/.) |
| PO | Chris Jordan | 186 | NU | Jordan et al., 2015 |
| PO | Eduardo Mendoza | 159 | MX | Mendoza et al., 2013 |
| PO | Joel | 96 | CS | |
| PO | Juan Carlos Cruz Diaz | 41 | CS | |
| PO | Diego Lizcano | 3 | CO | Jordan et al., 2015 |
| PO | Manolo Garcia | 173 | GT | Jordan et al., 2015 |
| PO | Nereyda Estrada | 125 | HO | |
| PO | Ninon Meyer | 26 | PM | Meyer et al., 2013 |
| PO | Celso Poot | 28 | BH | Jordan et al., 2015 |
| PO | Jan Schipper | 7 | CS | Gonzalez-Maya et al., 2012 |
| PO | Victor Montalvo | 66 | CS | |
| PA | Angelica Diaz-Pulido | 77 | CO | Diaz-Pulido, A. 2016 (Unpublished) |
| PA | Antonio de la Torre & Marina Rivero | 39 | MX | |
| PA | Carolina Saenz Bolanos | 21 | CS | |
| PA | LaRoy Brandt | 8 | CS | |
| PA | Chris Jordan | 195 | NU | Jordan et al., 2016 |
| PA | Cody Schank | 24 | NU | |
| PA | Mike Cove | 38 | CS | Cove et al., 2014 |
| PA | Michael Dobbins | 12 | BH | |

| | | | | |
|----|-------------------------------------|-----|----|-------------------------------------|
| PA | Eduardo Mendoza | 17 | MX | Carbajal-Borges et al., 2014 |
| PA | Esteban Brenes-Mora | 27 | CS | |
| PA | Francisco Botello | 8 | MX | Botello et al., in review |
| PA | Jess Fort | 55 | PM | Fort, 2016 |
| PA | Jose Gonzalez-Maya | 42 | CS | González-Maya et al., 2012 |
| PA | Marcella Kelly | 269 | BH | Kelly & Rowe, 2014 |
| PA | Ninon Meyer | 70 | PM | Meyer et al., 2015, 2016 |
| PA | Ninon Meyer, Clayton, Andrew Carver | 45 | PM | Meyer et al 2015,2016 |
| PA | G. Pozo-Montuy | 37 | MX | Pozo-Montuy et al., 2014,2015, 2016 |
| PA | Rafael Reyna | 87 | MX | |
| PA | Sebastian Mejia | 27 | CO | Mejía-Correa et al., 2014 |
| PA | TEAM | 436 | CS | Hurtado, 2014 |
| PA | Margot Wood and Jessica Gilbert | 113 | CS | Wood, Gilbert et al. in prep |

Bibliography

- Ahumada, J. A., Hurtado, J., & Lizcano, D. (2013). Monitoring the status and trends of tropical forest terrestrial vertebrate communities from camera trap data: a tool for conservation. *PloS One*, 8(9), e73707.
- Aitken, M., Roberts, D. W., & Shultz, L. M. (2007). Modeling distributions of rare plants in the Great Basin, western North America. *Western North American Naturalist / Brigham Young University*, 67(1), 26–38.
- Altrichter, M., Taber, A., Beck, H., Reyna-Hurtado, R., Lizarraga, L., Keuroghlian, A., & Sanderson, E. W. (2012). Range-wide declines of a key Neotropical ecosystem architect, the Near Threatened white-lipped peccary *Tayassu pecari*. *Oryx: The Journal of the Fauna Preservation Society*, 46(1), 87–98.
- Anderson, R. P., Peterson, A. T., & Gómez-Laverde, M. (2002). Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos*, 98(1), 3–16.
- Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, 22(1), 42–47.
- Araújo, M. B., & Peterson, A. T. (2012). Uses and misuses of bioclimatic envelope modeling. *Ecology*, 93(7), 1527–1539.
- Arino, O., Perez, J. J. R., Kalogirou, V., Bontemps, S., Defourny, P., & Van Bogaert, E. (2012). Global land cover map for 2009 (GlobCover 2009) [Data set]. PANGAEA. <https://doi.org/https://doi.pangaea.de/10.1594/PANGAEA.787668>
- Augustine, B., Andrew Royle, J., Kelly, M., Satter, C., Alonso, R., Boydston, E., & Crooks, K. (2017, June 8). Spatial capture-recapture with partial identity: an application to camera traps. *bioRxiv*. <https://doi.org/10.1101/056804>
- Austin, M. (2007). Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, 200(1–2), 1–19.
- Austin, M. P. (2002). Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, 157(2-3), 101–118.
- Bailey, D. W., Gross, J. E., & Laca, E. A. (1996). Mechanisms that result in large herbivore grazing distribution patterns. *Journal of Range Management*. Retrieved from http://www.academia.edu/download/42804673/Mechanisms_That_Result_in_Large_Herbivor20160218-24197-1de5fki.pdf
- Bailey, L. L., MacKenzie, D. I., & Nichols, J. D. (2014). Advances and applications of occupancy models. *Methods in Ecology and Evolution / British Ecological Society*, 5(12), 1269–1279.

- Balmford, A., & Whitten, T. (2003). Who should pay for tropical conservation, and how could the costs be met? *Oryx: The Journal of the Fauna Preservation Society*, 37(2), 238–250.
- Barber, C. P., Cochrane, M. A., Souza, C. M., Jr., & Laurance, W. F. (2014). Roads, deforestation, and the mitigating effect of protected areas in the Amazon. *Biological Conservation*, 177, 203–209.
- Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution / British Ecological Society*, 3(2), 327–338.
- Barlow, J., Stephens, P. A., Bode, M., Cadotte, M. W., Lucas, K., Newton, E., ... Pettoirelli, N. (2018). On the extinction of the single-authored paper: The causes and consequences of increasingly collaborative applied ecological research. *The Journal of Applied Ecology*, 55(1), 1–4.
- Bean, W. T., Stafford, R., & Brashares, J. S. (2012). The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. *Ecography*, 35(3), 250–258.
- Beck, J., Böller, M., Erhardt, A., & Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics*, 19, 10–15.
- Bello, C., Galetti, M., Pizo, M. A., Magnago, L. F. S., Rocha, M. F., Lima, R. A. F., ... Jordano, P. (2015). Defaunation affects carbon storage in tropical forests. *Science Advances*, 1(11), e1501105.
- Bonham, C. A., Sacayon, E., & Tzi, E. (2008). Protecting imperiled “paper parks”: potential lessons from the Sierra Chinajá, Guatemala. *Biodiversity and Conservation*, 17(7), 1581–1593.
- Borchers, D. L., & Efford, M. G. (2008). Spatially explicit maximum likelihood methods for capture-recapture studies. *Biometrics*, 64(2), 377–385.
- Botello, F., Romero-Calderón, A. G., Sánchez-Hernández, J., Hernández, O., López-Villegas, G., & Sánchez-Cordero, V. (2017). Densidad poblacional del tapir centroamericano (*Tapirella bairdii*) en bosque mesófilo de montaña en Totontepec Villa de Morelos, Oaxaca, México. *Revista Mexicana de Biodiversidad*, 88(4), 918–923.
- Bowman, J., Jaeger, J. A. G., & Fahrig, L. (2002). Dispersal Distance of Mammals is Proportional To Home Range Size. *Ecology*, 83(7), 2049–2055.
- Boyce, M. S., Vernier, P. R., Nielsen, S. E., & Schmiegelow, F. K. A. (2002). Evaluating resource selection functions. *Ecological Modelling*, 157(2-3), 281–300.

- Bray, D. B., Duran, E., Ramos, V. H., Mas, J.-F., Velazquez, A., McNab, R. B., ... Radachowsky, J. (2008). Tropical Deforestation, Community Forests, and Protected Areas in the Maya Forest. *Ecology and Society*, 13(2).
- Brooks, D. M., Bodmer, R. E., & Matola, S. (1997). Tapir Action Plan. IUCN/SSC Tapir Specialist Group.
- Bruner, A. G., Gullison, R. E., & Balmford, A. (2004). Financial Costs and Shortfalls of Managing and Expanding Protected-Area Systems in Developing Countries. *Bioscience*, 54(12), 1119–1126.
- Cade, B. S. (2015). Model averaging and muddled multimodel inferences. *Ecology*, 96(9), 2370–2382.
- Calenge, C. (2011). Analysis of animal movements in R: the adehabitatLT Package. R Foundation for Statistical Computing, Vienna. Retrieved from <ftp://ftp.br.debian.org/CRAN/web/packages/adehabitatLT/vignettes/adehabitatLT.pdf>
- Carbajal-Borges, J. P., Godínez-Gómez, O., & Mendoza, E. (2014). Density, abundance and activity patterns of the endangered *Tapirus bairdii* in one of its last strongholds in southern Mexico. *Tropical Conservation Science*, 7(1), 100–114.
- Caro, T. M., & O’Doherty, G. (1999). On the Use of Surrogate Species in Conservation Biology. *Conservation Biology: The Journal of the Society for Conservation Biology*, 13(4), 805–814.
- Carrillo Reyna, N. L., Weissenberger, H., & Reyna-Hurtado, R. (2015). Distribución potencial del Tapir Centroamericano en la península de Yucatán. *Therya*, 6(3), 575–596.
- Cayuela, L., Golicher, D. J., Newton, A. C., Kolb, M., de Albuquerque, F. S., Alkemade, J. R. M., & Pérez, A. M. (2009). Species distribution modeling in the tropics: problems, potentialities, and the role of biological data for effective species conservation. *Mongabay.com Open Access Journal - Tropical Conservation Science* Vol, 2(3), 319–352.
- Chandler, R. B., & Andrew Royle, J. (2013). Spatially explicit models for inference about density in unmarked or partially marked populations. *The Annals of Applied Statistics*, 7(2), 936–954.
- Chandler, R. B., & Clark, J. (2014). Spatially explicit integrated population models. *Methods in Ecology and Evolution / British Ecological Society*, 5(12), 1351–1360.
- Chase, J. M., & Leibold, M. A. (2003). *Ecological Niches: Linking Classical and Contemporary Approaches*. University of Chicago Press.
- Clements, G. R., Mark RAYAN, D., Aziz, S. A., Kawanishi, K., Traeholt, C., Magintan, D., ... Tingley, R. (2012). Predicting the distribution of the Asian tapir in

- Peninsular Malaysia using maximum entropy modeling. *Integrative Zoology*, 8(1), 400–406.
- Collen, B., Ram, M., Zamin, T., & Mc Rae, L. (2008). The tropical biodiversity data gap: addressing disparity in global monitoring. *Mongabay.com Open Access Journal - Tropical Conservation Science* Vol, 1(2), 75–88.
- Colwell, R. K., & Rangel, T. F. (2009). Hutchinson's duality: the once and future niche. *Proceedings of the National Academy of Sciences of the United States of America*, 106 Suppl 2, 19651–19658.
- Cove, M. V., Pardo Vargas, L. E., de la Cruz, J. C., Spínola, R. M., Jackson, V. L., Saénz, J. C., & Chassot, O. (2014). Factors influencing the occurrence of the Endangered Baird's tapir *Tapirus bairdii*: potential flagship species for a Costa Rican biological corridor. *Oryx: The Journal of the Fauna Preservation Society*, 48(03), 402–409.
- Cruse, B., Liedloff, A. C., & Wintle, B. A. (2012). A new method for dealing with residual spatial autocorrelation in species distribution models. *Ecography*, 35(10), 879–888.
- Cressie, N., & Wikle, C. K. (2015). *Statistics for Spatio-Temporal Data*. John Wiley & Sons.
- Critchlow, R., Plumptre, A. J., Alidria, B., Nsubuga, M., Driciru, M., Rwetsiba, A., ... Beale, C. M. (2017). Improving Law-Enforcement Effectiveness and Efficiency in Protected Areas Using Ranger-collected Monitoring Data. *Conservation Letters*, 10(5), 572–580.
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695(5), 1–9.
- de la Torre, J. A., Rivero, M., Camacho, G., & Álvarez-Márquez, L. A. (2017). Assessing occupancy and habitat connectivity for Baird's tapir to establish conservation priorities in the Sierra Madre de Chiapas, México. *Journal for Nature Conservation*. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1617138117303680>
- Denevan, W. M. (1992). The Pristine Myth: The Landscape of the Americas in 1492. *Annals of the Association of American Geographers*. Association of American Geographers, 82(3), 369–385.
- Diggle, P. J. (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*, Third Edition. Taylor & Francis.
- Dilts, T. E., Weisberg, P. J., Leitner, P., Matocq, M. D., Inman, R. D., Nussear, K. E., & Esque, T. C. (2016). Multiscale connectivity and graph theory highlight critical areas for conservation under climate change. *Ecological Applications: A Publication of the Ecological Society of America*, 26(4), 1223–1237.
- Dirzo, R., Young, H. S., Galetti, M., Ceballos, G., Isaac, N. J. B., & Collen, B. (2014). Defaunation in the Anthropocene. *Science*, 345(6195), 401–406.

- Dorazio, R. M. (2012). Predicting the geographic distribution of a species from presence-only data subject to detection errors. *Biometrics*, 68(4), 1303–1312.
- Dorazio, R. M. (2014a). Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. *Global Ecology and Biogeography: A Journal of Macroecology*, 23(12), 1472–1484.
- Dorazio, R. M. (2014b). Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. *Global Ecology and Biogeography: A Journal of Macroecology*, 23(12), 1472–1484.
- Dunn, M., Estrada, N., & Smith, D. A. (2012). The coexistence of Baird's tapir (*Tapirus bairdii*) and indigenous hunters in northeastern Honduras. *Integrative Zoology*, 7(4), 429–438.
- Efford, M. (2004). Density estimation in live-trapping studies. *Oikos*, 106(3), 598–610.
- Efford, M. G., & Dawson, D. K. (2012). Occupancy in continuous habitat. *Ecosphere*, 3(4), 1–15.
- Elith, J., & Graham, C. H. (2009). Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, 32(1), 66–77.
- Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., ... Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2), 129–151.
- Elith, J., Jane, E., Michael, K., & Steven, P. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution / British Ecological Society*, 1(4), 330–342.
- Elith, J., & Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), 677–697.
- Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17(1), 43–57.
- Elton, C. (1927). *Animal ecology*. 207 pp. Sidgwick & Jackson, LTD. London.
- Engler, R., Guisan, A., & Rechsteiner, L. (2004a). An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *The Journal of Applied Ecology*, 41(2), 263–274.
- Engler, R., Guisan, A., & Rechsteiner, L. (2004b). An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *The Journal of Applied Ecology*, 41(2), 263–274.

- Estrada, E., & Bodin, O. (2008). Using network centrality measures to manage landscape connectivity. *Ecological Applications: A Publication of the Ecological Society of America*, 18(7), 1810–1825.
- Eugster, M. J. A., & Schlesinger, T. (2010). osmar: OpenStreetMap and R. *R Journal*. Retrieved from <http://osmar.r-forge.r-project.org/RJpreprint.pdf>
- F. Dormann, C., M. McPherson, J., B. Araújo, M., Bivand, R., Bolliger, J., Carl, G., ... Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30(5), 609–628.
- Feeley, K. J., & Silman, M. R. (2011). The data void in modeling current and future distributions of tropical species. *Global Change Biology*, 17(1), 626–630.
- Ferreguetti, Á. C., Tomás, W. M., & Bergallo, H. G. (2017). Density, occupancy, and detectability of lowland tapirs, *Tapirus terrestris*, in Vale Natural Reserve, southeastern Brazil. *Journal of Mammalogy*, 98(1), 114–123.
- Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1), 38–49.
- Fink, D., Hochachka, W. M., Zuckerberg, B., Winkler, D. W., Shaby, B., Munson, M. A., ... Kelling, S. (2010). Spatiotemporal exploratory models for broad-scale survey data. *Ecological Applications: A Publication of the Ecological Society of America*, 20(8), 2131–2147.
- Fletcher, R. J., McCleery, R. A., Greene, D. U., & Tye, C. A. (2016). Integrated models that unite local and regional data reveal larger-scale environmental relationships and improve predictions of species distributions. *Landscape Ecology*, 31(6), 1369–1382.
- Foerster, C. R., & Vaughan, C. (2002). Home Range, Habitat Use, and Activity of Baird's Tapir in Costa Rica. *Biotropica*, 34(4), 423–437.
- Foley, J. A., Defries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., ... Snyder, P. K. (2005). Global consequences of land use. *Science*, 309(5734), 570–574.
- Foltete, J.-C., Clauzel, C., Vuidel, G., & Tournant, P. (2012). Integrating graph-based connectivity metrics into species distribution models. *Landscape Ecology*, 27, 557–569.
- Fort, J. L. (2016). Large carnivore occupancy and human-wildlife conflict in Panamá. Southern Illinois University at Carbondale.
- Foster, R. J., & Harmsen, B. J. (2012). A critique of density estimation from camera-trap data. *The Journal of Wildlife Management*, 76(2), 224–236.
- Frankham, R. (2005). Genetics and extinction. *Biological Conservation*, 126(2), 131–140.

- Franklin, J., Wejnert, K. E., Hathaway, S. A., Rochester, C. J., & Fisher, R. N. (2009). Effect of species rarity on the accuracy of species distribution models for reptiles and amphibians in southern California. *Diversity and Distributions*, 15(1), 167–177.
- Franklin, J. (2010a). *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press.
- Franklin, J. (2010b). Moving beyond static species distribution models in support of conservation biogeography. *Diversity and Distributions*, 16(3), 321–330.
- García, M., Jordan, C. A., O’Farril, G., Poot, C., Meyer, N., Estrada, N., ... Ruiz-Galeano, M. (2016). *Tapirus bairdii*. <https://doi.org/10.2305/IUCN.UK.2016-1.RLTS.T21471A45173340.en>
- Gardner, B., Reppucci, J., Lucherini, M., & Royle, J. A. (2010). Spatially explicit inference for open populations: estimating demographic parameters from camera-trap studies. *Ecology*, 91(11), 3376–3383.
- Gaston, K. J. (1991). How Large Is a Species’ Geographic Range? *Oikos*, 61(3), 434–438.
- Gaston, K. J. (1994). *Rarity*. Chapman & Hall.
- Gelfand, A. E., Silander, J. A., Shanshan, W., Andrew, L., Lewis, P. O., Rebelo, A. G., & Mark, H. (2006). Explaining species distribution patterns through hierarchical modeling. *Bayesian Analysis*, 1(1), 41–92.
- Getis, A., & Ord, J. K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, 24(3), 189–206.
- Girof, P. O. (2002). The Darien Region Between Colombia and Panama: Gap or Seal? *Human Rights and the Environment: Conflicts and Norms in a Globalizing World*, 174. Retrieved from <https://books.google.com/books?hl=en&lr=&id=PrEd8lcoz0EC&oi=fnd&pg=PA172&dq=darien+colombia+conflict+development&ots=L2Maa4Cg3d&sig=fcccvatRhNsP6imYCSIkKMf-jWY>
- Glista, D. J., DeVault, T. L., & DeWoody, J. A. (2009). A review of mitigation measures for reducing wildlife mortality on roadways. *Landscape and Urban Planning*, 91(1), 1–7.
- Golding, N., August, T. A., Lucas, T. C. D., Gavaghan, D. J., van Loon, E. E., & McInerny, G. (2017). The zoon R package for reproducible and shareable species distribution modelling. *Methods in Ecology and Evolution* / British Ecological Society. <https://doi.org/10.1111/2041-210X.12858>
- Golub, G. H., & Van Loan, C. F. (2012). *Matrix Computations*. JHU Press.

- González-Maya, J. F., Schipper, J., Polidoro, B., Hoepker, A., Zárrate-Charry, D., & Belant, J. L. (2012). Baird's tapir density in high elevation forests of the Talamanca region of Costa Rica. *Integrative Zoology*, 7(4), 381–388.
- Google Earth Engine Team. (2015). Google Earth Engine: A planetary-scale geo-spatial analysis platform [Data set]. Retrieved from <https://earthengine.google.com>
- Grajales, J. (2013). State Involvement, Land Grabbing and Counter-Insurgency in Colombia. *Development and Change*, 44(2), 211–232.
- Grinnell, J. (1917). The Niche-Relationships of the California Thrasher. *The Auk*, 34(4), 427–433.
- Gschweng, M., Kalko, E. K. V., Berthold, P., Fiedler, W., & Fahr, J. (2012). Multi-temporal distribution modelling with satellite tracking data: predicting responses of a long-distance migrant to changing environmental conditions. *The Journal of Applied Ecology*, 49(4), 803–813.
- Guillera-Arroita, G., Ridout, M. S., & Morgan, B. J. T. (2010). Design of occupancy studies with imperfect detection. *Methods in Ecology and Evolution / British Ecological Society*, 1(2), 131–139.
- Guisan, A., Antoine, G., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2-3), 147–186.
- Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., ... Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, 16(12), 1424–1435.
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., ... Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160), 850–853.
- Hastie, T., & Fithian, W. (2013). Inference from presence-only data; the ongoing controversy. *Ecography*, 36(8), 864–867.
- Hefley, T. J., & Hooten, M. B. (2016). Hierarchical Species Distribution Models. *Current Landscape Ecology Reports*. <https://doi.org/10.1007/s40823-016-0008-7>
- Hernandez, P. A., Graham, C. H., Master, L. L., & Albert, D. L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, 29, 773–785.
- Higgins, S. I., O'Hara, R. B., & Römermann, C. (2012). A niche for biology in species distribution models. *Journal of Biogeography*, 39(12), 2091–2095.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15), 1965–1978.

- Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2016). *dismo*: Species Distribution Modeling. R Package Version 1.1-1. Retrieved from <https://CRAN.R-project.org/package=dismo>
- Hijmans, R. J., van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J. A., ... Others. (2016). Package “raster.” R Package. <https://cran.r-project.org/web/packages/raster/index.html> (accessed 1 October 2016). Retrieved from <http://healthstat.snu.ac.kr/CRAN/web/packages/raster/raster.pdf>
- Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, 199(2), 142–152.
- Holderegger, R., & Di Giulio, M. (2010). The genetic effects of roads: A review of empirical evidence. *Basic and Applied Ecology*, 11(6), 522–531.
- Holt, R. D., & Gaines, M. S. (1992). Analysis of adaptation in heterogeneous landscapes: Implications for the evolution of fundamental niches. *Evolutionary Ecology*, 6(5), 433–447.
- Hurlbert, A. H., & Jetz, W. (2007). Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *Proceedings of the National Academy of Sciences of the United States of America*, 104(33), 13384–13389.
- Hurlbert, A. H., & White, E. P. (2005). Disparity between range map- and survey-based analyses of species richness: patterns, processes and implications. *Ecology Letters*, 8(3), 319–327.
- Hutchinson, G. E. (1957). Cold spring harbor symposium on quantitative biology. Concluding Remarks.
- Illian, J., Penttinen, P. A., Stoyan, H., & Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. Wiley.
- Isaac, N. J. B., Turvey, S. T., Collen, B., Waterman, C., & Baillie, J. E. M. (2007). Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PloS One*, 2(3), e296.
- IUCN. (2001). *IUCN Red List Categories and Criteria*. International Union for Conservation of Nature.
- IUCN. (2016a). *Guidelines for Using the IUCN Red List Categories and Criteria*. IUCN.
- IUCN. (2016b). *The IUCN Red List of Threatened Species [Data set]*. Retrieved from <http://www.iucnredlist.org>
- IUCN-ORMACC. (2016). *Map of the Indigenous Peoples, Protected Areas and Natural Ecosystem of Central America [Data set]*. Retrieved from ormacc@iucn.org
- Jackson, S. T., & Overpeck, J. T. (2000). Responses of plant populations and communities to environmental changes of the late Quaternary. *Paleobiology*, 26(sp4), 194–220.

- Jetz, W., McPherson, J. M., & Guralnick, R. P. (2012). Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in Ecology & Evolution*, 27(3), 151–159.
- Jiguet, F., Barbet-Massin, M., & Henry, P.-Y. (2010). Predicting potential distributions of two rare allopatric sister species, the globally threatened *Doliornis cotingas* in the Andes. *Journal of Field Ornithology*, 81(4), 325–339.
- Jiménez-Valverde, A., Lobo, J. M., & Hortal, J. (2008). Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions*, 14(6), 885–890.
- Joppa, L. N., Loarie, S. R., & Pimm, S. L. (2008). On the protection of “protected areas.” *Proceedings of the National Academy of Sciences*, 105(18), 6673–6678.
- Jordan, C. A. (2015). The dynamics of wildlife and environmental knowledge in a bioculturally diverse coupled natural and human system in the Caribbean region of Nicaragua (Ph.D.). (G. R. Urquhart, Ed.). Michigan State University.
- Jordan, C. A., Galeano, M. R., & Alonzo, A. S. (2014). La Cacería Histórica de Tapires Centroamericanos (*Tapirus bairdii*) en la RAAS, Nicaragua. *Estud Ambient*, 1(1), 73–87.
- Jordan, C. A., Schank, C. J., Urquhart, G. R., & Dans, A. J. (2016). Terrestrial Mammal Occupancy in the Context of Widespread Forest Loss and a Proposed Interoceanic Canal in Nicaragua’s Decreasingly Remote South Caribbean Region. *PloS One*, 11(3), e0151372.
- Karger, D. N., Conrad, O., Böhrner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., ... Kessler, M. (2017). Climatologies at high resolution for the earth’s land surface areas. *Scientific Data*, 4, 170122.
- Keith, D. A., Akçakaya, H. R., Thuiller, W., Midgley, G. F., Pearson, R. G., Phillips, S. J., ... Rebelo, T. G. (2008). Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. *Biology Letters*, 4(5), 560–563.
- Kéry, M. (2011). Towards the modelling of true species distributions. *Journal of Biogeography*, 38(4), 617–618.
- Kery, M., Gardner, B., & Monnerat, C. (2010). Predicting species distributions from checklist data using site-occupancy models. *Journal of Biogeography*, 37, 1851–1862.
- Kéry, M., Guillera-Aroita, G., & Lahoz-Monfort, J. J. (2013). Analysing and mapping species range dynamics using occupancy models. *Journal of Biogeography*, 40(8), 1463–1474.

- Kery, M., & Royle, J. A. (2015). *Applied Hierarchical Modeling in Ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 1: Prelude and Static Models*. Elsevier Science.
- Koshkina, V., Wang, Y., Gordon, A., Dorazio, R. M., White, M., & Stone, L. (2017). Integrated species distribution models: combining presence-background data and site-occupancy data with imperfect detection. *Methods in Ecology and Evolution / British Ecological Society*, 8(4), 420–430.
- Koster, J. M. (2006). Assessing the sustainability of Baird's tapir hunting in the Bosawas Reserve, Nicaragua. *Tapir Conservation*, 15(2), 23–28.
- Lahoz-Monfort, J. J., Guillera-Arroita, G., & Wintle, B. A. (2014). Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography: A Journal of Macroecology*, 23(4), 504–515.
- Latimer, A. M., Wu, S., Gelfand, A. E., & Silander, J. A., Jr. (2006). Building statistical models to analyze species distributions. *Ecological Applications: A Publication of the Ecological Society of America*, 16(1), 33–50.
- Leathwick, J. R., & Austin, M. P. (2001). Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology*. Retrieved from [http://onlinelibrary.wiley.com/doi/10.1890/0012-9658\(2001\)082%5B2560:CIBTSI%5D2.0.CO;2/full](http://onlinelibrary.wiley.com/doi/10.1890/0012-9658(2001)082%5B2560:CIBTSI%5D2.0.CO;2/full)
- Lesmeister, D. B., Nielsen, C. K., Schauber, E. M., & Hellgren, E. C. (2015). Spatial and temporal structure of a mesocarnivore guild in midwestern north America. *Wildlife Monographs*, 191(1), 1–61.
- Liu, C., Berry, P. M., Dawson, T. P., & Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28(3), 385–393.
- Lobo, J. M., Jiménez-Valverde, A., & Hortal, J. (2010). The uncertain nature of absences and their importance in species distribution modelling. *Ecography*, 33(1), 103–114.
- Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2008). AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography: A Journal of Macroecology*, 17(2), 145–151.
- Loiselle, B. A., Howell, C. A., Graham, C. H., Goerck, J. M., Brooks, T., Smith, K. G., & Williams, P. H. (2003). Avoiding Pitfalls of Using Species Distribution Models in Conservation Planning. *Conservation Biology: The Journal of the Society for Conservation Biology*, 17(6), 1591–1600.
- Lomba, A., Pellissier, L., Randin, C., Vicente, J., Moreira, F., Honrado, J., & Guisan, A. (2010). Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation*, 143(11), 2647–2657.

- Lomolino, M. V. (2004). Conservation biogeography. *Frontiers of Biogeography: New Directions in the Geography of Nature*, 293–296.
- Lomolino, M. V., Riddle, B. R., Brown, J. H., & Whittaker, R. J. (2010). *Biogeography*. Sinauer Associates.
- MacKenzie, D. I., Nichols, J. D., Hines, J. E., Knutson, M. G., & Franklin, A. B. (2003). Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, 84(8), 2200–2207.
- MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., & Langtimm, C. A. (2002). Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One. *Ecology*, 83(8), 2248–2255.
- MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., & Hines, J. E. (2006). *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. Elsevier.
- MacKenzie, D. I., Nichols, J. D., Seamans, M. E., & Gutiérrez, R. J. (2009). Modeling species occurrence dynamics with multiple states and imperfect detection. *Ecology*, 90(3), 823–835.
- Mackenzie, D. I., & Royle, J. A. (2005). Designing occupancy studies: general advice and allocating survey effort. *The Journal of Applied Ecology*, 42(6), 1105–1114.
- Mair, L., & Ruete, A. (2016). Explaining Spatial Variation in the Recording Effort of Citizen Science Data across Multiple Taxa. *PloS One*, 11(1), e0147796.
- McCann, N. P., Wheeler, P. M., Coles, T., & Bruford, M. W. (2012). Rapid ongoing decline of Baird's tapir in Cusuco National Park, Honduras. *Integrative Zoology*, 7, 420–428.
- McClintock, B. T., White, G. C., Antolin, M. F., & Tripp, D. W. (2009). Estimating Abundance Using Mark–Resight When Sampling Is with Replacement or the Number of Marked Individuals Is Unknown. *Biometrics*, 65(1), 237–246.
- McDonald, T. L. (2013). The point process use-availability or presence-only likelihood and comments on analysis. *The Journal of Animal Ecology*, 82(6), 1174–1182.
- McGill, B. J. (2010). Towards a unification of unified theories of biodiversity: Towards a unified unified theory. *Ecology Letters*, 13(5), 627–642.
- McInerny, G. J., & Etienne, R. S. (2012). Ditch the niche--is the niche a useful concept in ecology or species distribution modelling? *Journal of Biogeography*, 39(12), 2096–2102.
- Medici, E. P., Carrillo, L., Montenegro, O. L., Miller, P. S., Carbonell, F., Chassot, O., ... Mendoza, A. (2005). Baird's Tapir (*Tapirus bairdii*) Conservation Workshop Population and Habitat Viability Assessment (PHVA). IUCN/SSC Tapir Specialist Group.

- Medici, E. P., & Desbiez, A. (2012). Population viability analysis: using a modeling tool to assess the viability of tapir populations in fragmented landscapes. *Integrative Zoology*, 7(4), 356–372.
- Meijaard, E., Sheil, D., Nasi, R., Augeri, D., Rosenbaum, B., Iskandar, D., ... O'Brien, T. (2005). *Life After Logging: Reconciling Wildlife Conservation and Production Forestry in Indonesian Borneo*. CIFOR.
- Mejía-Correa, S., Diaz-Martinez, A., & Molina, R. (2014). Densidad y hábitos alimentarios de la danta *Tapirus bairdii* en el Parque Nacional Natural Los Katios, Colombia. *Tapir Conservation*, 23(32), 16–23.
- Mendoza, E., Fuller, T. L., Thomassen, H. A., Buermann, W., Ramírez-Mejía, D., & Smith, T. B. (2013). A preliminary assessment of the effectiveness of the Mesoamerican Biological Corridor for protecting potential Baird's tapir (*Tapirus bairdii*) habitat in southern Mexico. *Integrative Zoology*, 8(1), 35–47.
- Merow, C., Wilson, A. M., & Jetz, W. (2017). Integrating occurrence data and expert maps for improved species range predictions. *Global Ecology and Biogeography: A Journal of Macroecology*, 26(2), 243–258.
- Meyer, C. (2015). *Limitations in Global Information on Species Occurrences* (Ph.D.). (H. Kreft, Ed.). Universität Göttingen.
- Meyer, N. F. V., Esser, H. J., Moreno, R., van Langevelde, F., Liefting, Y., Ros Oller, D., ... Jansen, P. A. (2015/7). An assessment of the terrestrial mammal communities in forests of Central Panama, using camera-trap surveys. *Journal for Nature Conservation*, 26, 28–35.
- Meyer, N., Moreno, R., & Jansen, P. A. (2013). Distribution and conservation status of the Baird's tapir in Panama. *Tapir Conservation*, 22(30), 10–13.
- Meyer, N., Moreno, R., Sanches, E., Ortega, J., Brown, E., & Jansen, P. A. (2016). Do protected areas in Panama support intact assemblages of ungulates? *THERYA*, 7(1), 65–76.
- Michélot, T., Langrock, R., & Patterson, T. A. (2016). moveHMM: an R package for the statistical modelling of animal movement data using hidden Markov models. *Methods in Ecology and Evolution / British Ecological Society*, 7(11), 1308–1315.
- Millennium Ecosystem Assessment. (2005). *Ecosystems and Human Well-being: Synthesis*. World Resources Institute.
- Miller, J. A. (2012). Species distribution models: Spatial autocorrelation and non-stationarity. *Progress in Physical Geography*, 36(5), 681–692.
- Miller, J. A. (2014). Virtual species distribution models: Using simulated data to evaluate aspects of model performance. *Progress in Physical Geography*, 38(1), 117–128.

- Miller, J. A., & Holloway, P. (2015). Incorporating movement in species distribution models. *Progress in Physical Geography*, 39(6), 837–849.
- Miller, J., & Franklin, J. (2002). Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecological Modelling*, 157(2–3), 227–247.
- Miller, J., Franklin, J., & Aspinall, R. (2007). Incorporating spatial dependence in predictive vegetation models. *Ecological Modelling*, 202(3–4), 225–242.
- Miller, K., Chang, E., & Johnson, N. (2001). Defining common ground for the Mesoamerican Biological Corridor. World Resources Institute Washington, DC.
- Mills, L. S., Soulé, M. E., & Doak, D. F. (1993). The keystone-species concept in ecology and conservation. *Bioscience*, 43(4), 219–224.
- McPherson, J., & Jetz, W. (2007). Effects of species' ecology on the accuracy of distribution models. *Ecography*, 30(1), 135–151.
- Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., & Worm, B. (2011). How Many Species Are There on Earth and in the Ocean? *PLoS Biology*, 9(8), e1001127.
- Moreno, S. H. (1993). Impact of Development on the Panama Canal Environment. *Journal of Interamerican Studies and World Affairs*, 35(3), 129–149.
- Naimi, B., & Araújo, M. B. (2016). sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography*. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/ecog.01881/full>
- Naranjo, E. J. (1995). Abundancia y uso de hábitat del tapir (*Tapirus bairdii*) en un bosque tropical húmedo de Costa Rica. *Vida Silvestre Neotropical*, 4, 20–31.
- Naranjo, E. J. (2009). Ecology and Conservation of Baird's tapir in Mexico. *Mongabay.com Open Access Journal - Tropical Conservation Science Vol*, 2(2), 140–158.
- Naranjo, E. J., Alcalá, S. A. A., Briones, F. A. F., & Hurtado, R. A. R. (2015). Distribución, abundancia y amenazas a las poblaciones de tapir (*Tapirus bairdii*) y pecarí de labios blancos (*Tayassu pecari*) en México. *THERYA*, 6(1), 227–249.
- NASA. (2017). MODIS Collection 6 NRT Hotspot / Active Fire Detections MCD14DL [Data set]. <https://doi.org/10.5067/FIRMS/MODIS/MCD14DL.NRT.006>
- Nathan, R., Klein, E., Robledo-Arnuncio, J. J., & Revilla, E. (2012). Dispersal kernels: review. *Dispersal Ecology and Evolution*. Oxford University Press. Oxford. Pp, 187–210.
- Nelson, T. A., & Boots, B. (2008). Detecting spatial hot spots in landscape ecology. *Ecography*, 31(5), 556–566.
- Norton, J. E., & Ashley, M. V. (2004). Genetic variability and population structure among wild Baird's tapirs. *Animal Conservation*, 7(2), 211–220.

- Noss, A. J., Gardner, B., Maffei, L., Cuéllar, E., Montaña, R., Romero-Muñoz, A., ... O'Connell, A. F. (2012). Comparison of density estimation methods for mammal populations with camera traps in the Kaa-Iya del Gran Chaco landscape. *Animal Conservation*, 15(5), 527–535.
- O'Connell, A. F., Nichols, J. D., & Ullas Karanth, K. (2010). *Camera Traps in Animal Ecology: Methods and Analyses*. Springer Science & Business Media.
- O'Farrill, G., Calmé, S., Sengupta, R., & Gonzalez, A. (2012). Effective dispersal of large seeds by Baird's tapir: a large-scale field experiment. *Journal of Tropical Ecology*, 28(1), 119–122.
- O'Farrill, G., Galetti, M., & Campos-Arceiz, A. (2013). Frugivory and seed dispersal by tapirs: an insight on their ecological role. *Integrative Zoology*, 8(1), 4–17.
- Oliveira-Santos, L. G. R., Zucco, C. A., Antunes, P. C., & Crawshaw, P. G., Jr. (2010). Is it possible to individually identify mammals with no natural markings using camera-traps? A controlled case-study with lowland tapirs. *Mammalian Biology - Zeitschrift Für Säugetierkunde*, 75(4), 375–378.
- Oliver, W. L. R. (1993). *Pigs, peccaries and hippos: status survey and conservation action plan*. Island Pr.
- Osuri, A. M., Ratnam, J., Varma, V., Alvarez-Loayza, P., Hurtado Astaiza, J., Bradford, M., ... Sankaran, M. (2016). Contrasting effects of defaunation on aboveground carbon storage across the global tropics. *Nature Communications*, 7, 11351.
- Pardini, R., Bueno, A. de A., Gardner, T. A., Prado, P. I., & Metzger, J. P. (2010). Beyond the fragmentation threshold hypothesis: regime shifts in biodiversity across fragmented landscapes. *PloS One*, 5(10), e13666.
- Pastore, M. (2017). *overlapping: Estimation of Overlapping in Empirical Distributions*. Retrieved from <https://CRAN.R-project.org/package=overlapping>
- Pearce, J. L., & Boyce, M. S. (2006). Modelling distribution and abundance with presence-only data. *The Journal of Applied Ecology*, 43(3), 405–412.
- Pearson, R. G., Thuiller, W., Araújo, M. B., Martinez-Meyer, E., Brotons, L., McClean, C., ... Lees, D. C. (2006). Model-based uncertainty in species range prediction. *Journal of Biogeography*, 33(10), 1704–1711.
- Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633), 418–422.
- Peres, C. A., Barlow, J., & Laurance, W. F. (2006). Detecting anthropogenic disturbance in tropical forests. *Trends in Ecology & Evolution*, 21(5), 227–229.
- Peterson, A. T., Soberon, J., & Sanchez-Cordero, V., V. (1999). Conservatism of ecological niches in evolutionary time. *Science*, 285(5431), 1265–1267.

- Pettorelli, N., Ryan, S., Mueller, T., & Bunnefeld, N. (2011). The Normalized Difference Vegetation Index (NDVI): unforeseen successes in animal ecology. *Climate Research*. Retrieved from <http://www.jstor.org/stable/24872307>
- Phillips, S. J., Dudík, M., & Schapire, R. E. (2004). A maximum entropy approach to species distribution modeling. In *Proceedings of the twenty-first international conference on Machine learning* (p. 83). ACM.
- Phillips, S. J., & Elith, J. (2013). On estimating probability of presence from use–availability or presence–background data. *Ecology*, 94(6), 1409–1419.
- Porter-Bolland, L., Ellis, E. A., Guariguata, M. R., Ruiz-Mallén, I., Negrete-Yankelevich, S., & Reyes-García, V. (2012). Community managed forests and forest protected areas: An assessment of their conservation effectiveness across the tropics. *Forest Ecology and Management*, 268, 6–17.
- Potts, S. G., Biesmeijer, J. C., Kremen, C., Neumann, P., Schweiger, O., & Kunin, W. E. (2010). Global pollinator declines: trends, impacts and drivers. *Trends in Ecology & Evolution*, 25(6), 345–353.
- Rabinowitz, A., & Zeller, K. A. (2010). A range-wide model of landscape connectivity and conservation for the jaguar, *Panthera onca*. *Biological Conservation*, 143(4), 939–945.
- Rabinowitz, D. (1981). Seven forms of rarity. In H. Synge (Ed.), *The Biological Aspects of Rare Plant Conservation* (pp. 205–217). Riley.
- Ramsey, D. S. L., Caley, P. A., & Robley, A. (2015). Estimating population density from presence–absence data using a spatially explicit model. *The Journal of Wildlife Management*, 79(3), 491–499.
- Raven, P. H., & Wilson, E. O. (1992). A fifty-year plan for biodiversity surveys. *Science*, 258(5085), 1099–1100.
- R Development Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, . Retrieved from <http://www.R-project.org>
- Reed, D. H. (2004). Extinction risk in fragmented habitats. In *Animal Conservation forum* (Vol. 7, pp. 181–191). Cambridge University Press.
- Renner, I. W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S. J., ... Warton, D. I. (2015). Point process models for presence-only analysis. *Methods in Ecology and Evolution / British Ecological Society*, 6(4), 366–379.
- Renner, I. W., & Warton, D. I. (2013). Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics*, 69(1), 274–281.
- Reside, A. E., Vanderwal, J. J., Kutt, A. S., & Perkins, G. C. (2010). Weather, not climate, defines distributions of vagile bird species. *PloS One*, 5(10), e13569.

- Reyna-Hurtado, R., Sanvicente-López, M., Pérez-Flores, J., Carrillo-Reyna, N., & Calmé, S. (2016a). Insights into the multiannual home range of a Baird's tapir (*Tapirus bairdii*) in the Maya Forest. *THERYA*, 7(2), 271–276.
- Reyna-Hurtado, R., Sanvicente-López, M., Pérez-Flores, J., Carrillo-Reyna, N., & Calmé, S. (2016b). Insights into the multiannual home range of a Baird's tapir (*Tapirus bairdii*) in the Maya Forest. *THERYA*, 7(2), 271–276.
- Ripple, W. J., Newsome, T. M., Wolf, C., Dirzo, R., Everatt, K. T., Galetti, M., ... Van Valkenburgh, B. (2015). Collapse of the world's largest herbivores. *Science Advances*, 1(4), e1400103.
- Rota, C. T., Fletcher, R. J., Jr, & Dorazio, R. M. (2009). Occupancy estimation and the closure assumption. *Journal of Applied*. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2664.2009.01734.x/full>
- Rota, C. T., Millspaugh, J. J., Kesler, D. C., Lehman, C. P., Rumble, M. A., & Jachowski, C. M. B. (2013). A re-evaluation of a case-control model with contaminated controls for resource selection studies. *The Journal of Animal Ecology*, 82(6), 1165–1173.
- Royle, J. A. (2004). N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, 60(1), 108–115.
- Royle, J. A., Chandler, R. B., Sollmann, R., & Gardner, B. (2013). *Spatial Capture-Recapture*. Elsevier Science.
- Royle, J. A., Chandler, R. B., Yackulic, C., & Nichols, J. D. (2012). Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. *Methods in Ecology and Evolution / British Ecological Society*, 3(3), 545–554.
- Royle, J. A., & Dorazio, R. M. (2008). *Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities*. Elsevier Science.
- Rupprecht, F., Oldeland, J., & Finckh, M. (2011). Modelling potential distribution of the threatened tree species *Juniperus oxycedrus*: how to evaluate the predictions of different modelling approaches? *Journal of Vegetation Science: Official Organ of the International Association for Vegetation Science*, 22(4), 647–659.
- Sagastume-Espinoza, K. O., & Romero, S. J. (2017). First photographic evidence of the iconic big mammals of Honduras, Baird's Tapir (*Tapirus bairdii*) and Jaguar (*Panthera onca*), in La Muralla Wildlife Refuge. *Therya*, 8(1). Retrieved from <http://www.redalyc.org/html/4023/402349653009/>
- Salazar, M. (2018, January 17). Denuncian a Marena por obstaculizar investigación en Indio Maíz. *Confidencial*.

- Sanderson, E. W., Redford, K. H., Chetkiewicz, C.-L. B., Medellin, R. A., Rabinowitz, A. R., Robinson, J. G., & Taber, A. B. (2002). Planning to Save a Species: the Jaguar as a Model. *Conservation Biology: The Journal of the Society for Conservation Biology*, 16(1), 58–72.
- Schank, C. J., Cove, M. V., Kelly, M. J., Mendoza, E., O’Farrill, G., Reyna-Hurtado, R., ... Miller, J. A. (2017). Using a novel model approach to assess the distribution and conservation status of the endangered Baird’s tapir. *Diversity and Distributions*, 1–13.
- Schank, C. J., Cove, M. V., Kelly, M. J., Nielsen, C. K., O’Farrill, G., Meyer, N., ... Miller, J. A. (in prep.). A sensitivity analysis of the application of Integrated Species Distribution Models (ISDMs) to mobile species.
- Schank, C. J., Mendoza, E., Vettorazzi, M. J. G., Cove, M. V., Jordan, C. A., O’Farrill, G., ... Others. (2015). Integrating current range-wide occurrence data with species distribution models to map the potential distribution of Baird’s Tapir. *The Newsletter of the IUCN/SSC Tapir Specialist Group*, 24(33), 15–25.
- Schipper, J., Chanson, J. S., Chiozza, F., Cox, N. A., Hoffmann, M., Katariya, V., ... Young, B. E. (2008). The status of the world’s land and marine mammals: diversity, threat, and knowledge. *Science*, 322(5899), 225–230.
- Schwartzman, S., Moreira, A., & Nepstad, D. (2000). Rethinking Tropical Forest Conservation: Perils in Parks. *Conservation Biology: The Journal of the Society for Conservation Biology*, 14(5), 1351–1357.
- Schwartzman, S., Nepstad, D., & Moreira, A. (2000). Arguing Tropical Forest Conservation: People versus Parks. *Conservation Biology: The Journal of the Society for Conservation Biology*, 14(5), 1370–1374.
- Schwerdtfeger, W. (1976). *World survey of climatology volume 12: climates of Central and South America*. Elsevier Scientific Publishing Company Amsterdam.
- Silver, N. (2012). *The Signal and the Noise: Why So Many Predictions Fail--but Some Don’t*. Penguin.
- Snow, N. P., Jarzyna, M. A., & VerCauteren, K. C. (2017). Interpreting and predicting the spread of invasive wild pigs. *The Journal of Applied Ecology*. <https://doi.org/10.1111/1365-2664.12866>
- Soberón, J. (2007). Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, 10(12), 1115–1123.
- Soberón, J., & Nakamura, M. (2009). Niches and distributional areas: concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences of the United States of America*, 106 Suppl 2, 19644–19650.
- Soberón, J., & Peterson, A. T. (2005). Interpretation of models of fundamental ecological niches and species’ distributional areas. *Biodiversity Informatics*, 2, 1–10.

- Stephens-Davidowitz, S. (2017). *Everybody Lies: Big Data, New Data, and what the Internet Can Tell Us about who We Really are*. Harper Collins.
- Stephens, P. A., Sutherland, W. J., & Freckleton, R. P. (1999). What Is the Allee Effect? *Oikos*, 87(1), 185–190.
- Stocks, A., McMahan, B., & Taber, P. (2007). Indigenous, colonist, and government Impacts on Nicaragua's Bosawas Reserve. *Conservation Biology: The Journal of the Society for Conservation Biology*, 21(6), 1495–1505.
- Stockwell, D. R. B., & Townsend, P. (2002). Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, 148(1), 1–13.
- Stuart, S. N., Chanson, J. S., Cox, N. A., Young, B. E., Rodrigues, A. S. L., Fischman, D. L., & Waller, R. W. (2004). Status and trends of amphibian declines and extinctions worldwide. *Science*, 306(5702), 1783–1786.
- Symonds, M. R. E., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, 65(1), 13–21.
- Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography*, 32(3), 369–373.
- Tobler, M. W., Hibert, F., Debeir, L., & Richard-Hansen, C. (2014). Estimates of density and sustainable harvest of the lowland tapir *Tapirus terrestris* in the Amazon of French Guiana using a Bayesian spatially explicit capture–recapture model. *Oryx: The Journal of the Fauna Preservation Society*, 48(03), 410–419.
- Townsend Peterson, A. (2006). Uses and Requirements of Ecological Niche Models and Related Distributional Models. *Biodiversity Informatics*, 3(0). <https://doi.org/10.17161/bi.v3i0.29>
- Townsend Peterson, A., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., & Araújo, M. B. (2011). *Ecological Niches and Geographic Distributions (MPB-49)*. Princeton University Press.
- Tsoar, A., Allouche, O., Steinitz, O., Rotem, D., & Kadmon, R. (2007). A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions*, 13(4), 397–405.
- UNEP-WCMC. (2014). The World Database on Protected Areas [Data set]. Retrieved from protectedplanet.net
- VanDerWal, J., Shoo, L. P., Graham, C., & Williams, S. E. (2009). Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological Modelling*, 220(4), 589–594.
- van Etten, J. (2017). R Package *gdistance*: Distances and Routes on Geographical Grids. *Journal of Statistical Software*, 76(13), 1–21.

- Vaz, U. L., Cunha, H. F., & Nabout, J. C. (2015). Trends and biases in global scientific literature about ecological niche models. *Brazilian Journal of Biology = Revista Brasileira de Biologia*, 75(4 Suppl 1), S17–S24.
- Warton, D. I., & Shepherd, L. C. (2010). Poisson point process models solve the “pseudo-absence problem” for presence-only data in ecology. *The Annals of Applied Statistics*, 4(3), 1383–1402.
- Watsa, M. E. (2014, August 13). “Natural Reserves” no more: illegal colonists deforest huge portions of Nicaraguan protected areas. Mongabay.
- White, G. C. (1982). Capture-recapture and removal methods for sampling closed populations. Los Alamos National Laboratory.
- Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., Guisan, A., & NCEAS Predicting Species Distributions Working Group. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, 14(5), 763–773.
- Worboys, G., Francis, W. L., & Lockwood, M. (2010). *Connectivity Conservation Management: A Global Guide (with Particular Reference to Mountain Connectivity Conservation)*. Earthscan.
- Wulsch, C., Caragiulo, A., Dias-Freedman, I., Quigley, H., Rabinowitz, S., & Amato, G. (2016). Genetic Diversity and Population Structure of Mesoamerican Jaguars (*Panthera onca*): Implications for Conservation and Management. *PloS One*, 11(10), e0162377.
- Yackulic, C. B., Chandler, R., Zipkin, E. F., Royle, J. A., Nichols, J. D., Campbell Grant, E. H., & Veran, S. (2013a). Presence-only modelling using MAXENT: when can we trust the inferences? *Methods in Ecology and Evolution / British Ecological Society*, 4(3), 236–243.
- Yackulic, C. B., Chandler, R., Zipkin, E. F., Royle, J. A., Nichols, J. D., Campbell Grant, E. H., & Veran, S. (2013b). Presence-only modelling using MAXENT: when can we trust the inferences? *Methods in Ecology and Evolution / British Ecological Society*, 4(3), 236–243.
- Zaniewski, A. E., A.Elizabeth, Z., Anthony, L., & Overton, J. M. (2002). Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*, 157(2-3), 261–280.
- Zeller, K. A., McGarigal, K., & Whiteley, A. R. (2012). Estimating landscape resistance to movement: a review. *Landscape Ecology*, 27(6), 777–797.
- Zeller, K. A., & Rabinowitz, A. (2011). Using geographic information systems for range-wide species conservation planning. *Geographic Information Systems*. Nova Science Publishers, Inc. , Hauppauge, NY. Retrieved from https://www.researchgate.net/profile/Alan_Rabinowitz/publication/265999529_U

sing_Geographic_Information_Systems_for_Range-
Wide_Species_Conservation_Planning/links/555b1dc708ae980ca6121234.pdf

Zurell, D., Berger, U., Cabral, J. S., Jeltsch, F., Meynard, C. N., Münkemüller, T., ...
Grimm, V. (2010). The virtual ecologist approach: simulating data and observers.
Oikos , 119(4), 622–635. ll, W. G. 1990. *Form and Style in Thesis Writing, a
Manual of Style*. Chicago: The University of Chicago Press.