Assessing the Connectivity for the jaguar (*Panthera onca*) in the United States-Mexico Border Ecoregions Using Species Distribution Modeling and Factorial Least Cost Path Analysis

by

Cirenia A. Torres

A Thesis Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
MASTER OF SCIENCE
(GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY)

August 2021

Epigraph

Gentle eyes
that see so much,
paws that have
the quiet touch.

Purrs to signal
"all is well"
and show more love
than words can tell.

Graceful movements
touched with pride,
a calming presence
by our side.

A friendship
that will last and grow -
small wonder
why we love them so.

- Author Unknown

Dedication

To *Panthera onca* (Linnaeus, 1758)

Eric Kilby https://www.flickr.com/photos/ekilby/6703324115/

Acknowledgements

Dr. Lopes-Gonzales recommend an ecology book that was very helpful for estimating mammalian home ranges which further helped me with the jaguar modeling. Special thanks to Drs. Wilson and Fleming for their meticulous feedback and helping me reach the summer graduation deadline.

Without Google Earth Engine it would not have been possible to preprocess my vegetation and water index covariate datasets using Landsat data. This is a valuable and free resource which I had not previously used in my SSI classes. It challenged me to learn JavaScript, other programming languages, and understand how to process Landsat data. My sincere thanks and appreciation to Google Earth Engine and the wonderful team who made this free for anyone to use.

Special thanks to my two cat children (Oniyuri and Yoruichi) and Larry Tran who have been patient with me throughout this process.

Before starting the SSI GIST program, I had very limited knowledge of GIS. I am happy that I learned the theory and applications for spatial thinking. All classes were fun and challenging.

Table of Contents

List of Tables

List of Figures

Abbreviations

AUC	Area Under the (ROC) Curve
CBP	U.S. Customs and Border Protection
CITES	Convention on International Trade in Endangered Species of Wild Fauna and Flora
CTI	Compound Topographic Index
DEM	Digital Elevation Model
DHS	Department of Homeland Security
EM	Ensemble Model
ENFA	Ecological Niche Factor Analysis
EVI	Enhanced Vegetation Index
FLCP	Factorial Least Cost Path
FPR	False Positive Rate
GARP	Genetic Algorithm for Ruleset Prediction
GBIF	Global Biodiversity Information Facility
GBM	Generalized Boosted Model
GEE	Google Earth Engine
GIS	Geographic Information System
GLM	Generalized Linear Model
GPS	Global Positioning System
GPW	Gridded Population of the World
IIC	Integral Index of Connectivity
IPP	Inhomogeneous Poisson Process

IUCN	International Union for Conservation of Nature
JCMA	Jaguar Management and Conservation Area
JCU	Jaguar Conservation Unit
JOD	Jaguar Observations Database
LCP	Least Cost Path
MaxEnt	Maximum Entropy Method
MD	Mahalanobis Distance
NDVI	Normalized Difference Vegetation Index
OLS	Ordinary Least Squares
OOB	Out of Bag
PC	Probability of Connectivity
RF	Random Forest
ROC	Receiver Operating Characteristic
SDM	Species Distribution Model
SSI	Spatial Sciences Institute
SVM	Support Vector Machines
TPR	True Positive Rate
TRI	Terrain Ruggedness Index
UNICOR	UNIversal CORridor Network Simulator
US	United States of America
USGS	U.S. Geological Survey

Abstract

Connectivity is important for biodiversity conservation because it can offset the impacts of habitat loss and fragmentation, allowing migration, dispersal, and adequate gene flow. Barriers that cut across a species range such as the United States-Mexico border wall can block dispersal and negatively impact gene flow between populations. It is therefore important to understand how to establish or re-establish wildlife corridors in order to help species survive. The focal species selected for this thesis project was the jaguar *(Panthera onca).* The study area comprised several ecoregions that covered portions of the United States of America (US) and Mexico. The jaguar's suitable habitat was identified using a Random Forest model to predict potential habitats. The factorial least-cost path analysis was used to identify the jaguar's potential corridors. Results predict there is good habitat for jaguars in the Sonoran-Sinaloan subtropical dry forest, Sinaloan dry forests, Sierra Madre Occidental, California montane chaparral and woodlands, Arizona Mountains forest, Sierra Madre Oriental pine-oak forests, Veracruz moist forests, Sierra de la Laguna pine-oak forests, Sierra de la Laguna dry forests, Tamaulipan matorral, and small portions of the Sonoran desert ecoregion. The jaguar's potential corridor modeling suggests that there were previously two high density corridors between the US and Mexico allowing jaguar connectivity. However, if the partially constructed border barriers are completed those jaguar corridors will be lost. Work on nine co-distributed mammals (orders: Carnivora and Artiodactyla): jaguar *(Panthera onca),* mountain lion *(Puma concolor),* ocelot (*Leopardus pardalis*), bobcat *(Lynx rufus),* black bear *(Ursus americanus),* gray fox *(Urocyon cinereoargenteus),* Mexican gray wolf *(Canis lupus baileyi)* Sonoran pronghorn *(Antilocapra americana sonoriensis),* and Bighorn sheep *(Ovis canadensis)* in the US-Mexico border ecoregions will continue after the completion of this work.

Chapter 1 : Introduction

Establishing connectivity between species habitats is important because it can offset the impacts of habitat loss and fragmentation. As anthropogenic pressures continue to push species into isolated habitat patches it is important to understand how to establish wildlife corridors that can help species migrate, disperse, and maintain adequate gene flow. This chapter introduces the concept of landscape connectivity, highlights some of the species at risk from loss of connectivity in the US-Mexico border ecoregions, describes the research objectives, and thesis organization.

1.1. Landscape Connectivity

1.1.1. Habitat Loss and Fragmentation

Current rates of species extinction are about 1,000 times greater than background rates of extinction and are likely underestimated (Pimm et al. 2014). The major threats that can drive worldwide species extinction are habitat degradation (including habitat loss and fragmentation), overexploitation, invasive species, climate change, pollution, and disease (Groom, Meffe, and Carroll 2006). Human caused habitat loss and degradation of natural habitat have large negative effects on biodiversity (Fahrig 2003) and are the main immediate threat to global biodiversity (Groom, Meffe, and Carroll 2006). Habitat destruction usually leads to habitat fragmentation which can significantly reduce biodiversity and damage ecosystems (Haddad et al. 2015).

Habitat fragmentation which reduces and isolates natural habitat, despite having differing effects on species, is “one of the greatest threats to regional and global biodiversity” (Groom, Meffe, and Carroll 2006, 250). Only species that have small home ranges and whose life history requirements can be met within a fragment may survive in fragmented landscapes (Groom, Meffe, and Carroll 2006). The negative effects of habitat fragmentation include edge effects that create barriers and filters that prevent some species from moving, lack of access to important habitat due to large distance between important habitat patches, and other effects that can negatively impact animal dispersal and seasonal migration. Additionally, habitat fragmentation can negatively impact gene flow which can contribute to a greater extinction risk (Riley et al. 2006; Ernest et al. 2014; Wan, Cushman, and Ganey 2018). If species populations become too genetically isolated in fragmented landscapes, they can suffer from inbreeding and low genetic variation which in turn can lead to local extinction (Rudnick et al. 2012). Habitat loss and fragmentation threaten the ability of organisms to stay connected to one another in a landscape by disrupting the ability of individual organisms and their genes to move across landscapes.

1.1.2. Connectivity

Establishing connectivity between natural areas is one way to help counter the effects of fragmentation (Groom, Meffe, and Carroll 2006; Haddad et al. 2015). Landscape connectivity is the amount of movement made possible by the landscape composition (Rudnick et al. 2012). The movement which is made possible by landscape connectivity can be the movement of genes, propagules, individuals, and populations of species. Landscape connectivity can facilitate two important movement processes of species: migration (i.e. seasonal movements) and dispersal (i.e. to occupy previously unoccupied territory by that species). Connectivity can be further classified into two types: structural and functional. Structural connectivity relates to physical landscape characteristics and functional connectivity “describes how well genes, propagules, individuals, or populations move through the landscape” (Rudnick et al. 2012, 2). We can either identify and protect currently existing landscape connectivity or reestablish connectivity in fragmented landscapes to help species gain resilience to future environmental conditions that may put them at a greater extinction risk. Maintaining or reestablishing landscape connectivity across ecoregions or continents for long periods of time is important for the long-term survival of species because it allows them to shift their ranges in response to climate change and other long-term ecological changes; if change happens too fast a species may not be able to adapt (Rudnick et al. 2012).

There are several modeling techniques that can be used to identify and quantify landscape connectivity. Cushman et al. (2013) describes several methods which can be used to estimate resistance including telemetry, landscape genetics, habitat quality, mark-recapture, and combinations of methods. Potential corridors can be identified using least-cost modeling or factorial least-cost paths (Cushman et al. 2013). For example, Elliot et al. (2014) used Global Positioning System (GPS) tracking data from dispersing African lions and resistance surfaces for calculating the factorial least-cost path network. Circuit theory, centrality analyses, resistant kernels, and network-based models are some of the other types of ways we can analyze connectivity (Cushman et al. 2013). Measuring functional connectivity can be achieved more cost effectively and with greater sample sizes using genetic approaches compared to tracking individual animals but the concern with genetic studies is that "current genetic patterns may not reflect the impact of current landscape features" (Rudnick et al. 2012, 6) and "genetic connectivity may be masked in some instances by local adaptation" (Rudnick et al. 2012, 6). Comparing resistance surfaces created with movement data to those created through landscape genetic analyses can be useful to evaluate the robustness of a connectivity model (Cushman et al. 2013). How well a species can travel across a landscape can depend on many factors and is typically predicted using landscape resistance surfaces created by assigning values to cells in a raster. Simple estimates of landscape resistance might not be able to adequately predict resistance to movement of a specialist species or of other species with "species-specific movements” with specific habitat needs (Rudnick et al. 2012, 6). Connectivity modeling needs to consider both functional and structural connectivity components that can ensure “long-term habitat shifts” (Rudnick et al. 2012, 6).

1.2. Co-distributed Species in the US-Mexico Border Regions

1.2.1. Border Ecoregions and Biodiversity Hotspots

An ecoregion is defined as a geographic area that has a distinct collection of “natural communities that share a large majority of their species, ecological dynamics, and similar environmental conditions and whose ecological interactions are critical for their long-term persistence” (Dinerstein et al. 1995, 4). The US-Mexico border crosses eight ecoregions described by Dinerstein et al. (2017): (1) the California Coastal Sage and Chaparral; (2) the Chihuahuan Desert; (3) the Sierra Madre Occidental Pine-Oak Forests; (4) the Sonoran Desert; (5) the Tamaulipan Mezquital; (6) the Western Gulf Coastal Grasslands; (7) the California Montane Chaparral and Woodlands; and (8) the Sierra Madre Oriental Pine-Oak Forests (Figure 1). Within these ecoregions there are five biodiversity conservation hotspots along the border, the Californias, Sonora Desert, Sky Islands, Big Bend, and Lower Rio Grande Valley (Defenders of Wildlife 2018). These biodiversity hotspots are areas of high biological diversity which are threatened with habitat degradation.

1.2.2. Species at Risk Due to Border Dispersal Barriers

Studies have found that the US-Mexico border wall threatens biodiversity. Lasky, Jetz, and Keitt (2011) performed the first large scale evaluation of the border ecoregions to assess the potential threats presented by the border dispersal barrier to species that are non-volant terrestrial vertebrates. They identified the California, Madrean archipelago, and Gulf coast border regions as having high species richness with high number of species at risk from existing border barriers

Figure 1 – Map of study area used for this thesis project showing US-Mexico ecoregions.

and from construction of potential new border barriers. They identified several amphibians, reptiles, and mammals for further study. Peters et al. (2018) found that the border bisects the geographic ranges of 1,077 native terrestrial and freshwater animals and 429 native plants. Of these, 62 species are listed as Critically Endangered, Endangered, or Vulnerable by the International Union for Conservation of Nature (IUCN). Endangered species including Peninsular bighorn sheep *(Ovis canadensis nelsoni),* Mexican gray wolf *(Canis lupus baileyi)* and Sonoran pronghorn *(Antilocapra americana sonoriensis)*. The study also mentions "if cut off by a border wall, 17% of the 346 species we analyzed, including jaguar (*Panthera onca*) and ocelot (*Leopardus pardalis*), would have residual US populations covering 20,000 square kilometers or less" (Peters et al. 2018, 741). Besides border fences and walls, other anthropogenic threats exist along the border such as artificial night lighting and noise pollution that can further threaten endangered Texas ocelots by discouraging dispersal of Mexican ocelots into Texas (Grigione and Mrykalo 2004). GPS tracked bobcats have died due to forced dispersal during previous border wall construction programs (Gaskill 2011).

1.3. Research Objectives and Thesis Organization

How do old and new border barriers affect connectivity for sensitive borderland mammals? Broadly speaking, this thesis has the objective to contribute to a better understanding of co-distributed border ecoregion species connectivity. The aim of this study is to identify areas on the US-Mexico border that would disrupt cross-border connectivity for one border species that may be sensitive to new border barriers. To accomplish this goal, core habitat areas and corridors for the jaguar in the border ecoregions and surrounding ecoregions were identified (Figure 1). It is important to find out how new border barriers would affect connectivity for mammals since existing and new border barriers can disrupt migration, dispersal, and adequate gene flow; specifically, in this study for the jaguar *(Panthera onca).* The findings can help with management decisions for biodiversity conservation efforts in these border ecoregions. To the best of my knowledge this is the first study to use the Factorial Least Cost Path (FLCP) modeling approach for the jaguar *(Panthera onca)* in this study area. The specific objectives for completing this study are as follows:

1. Identify appropriate border ecoregion focal species that could be impacted by border barriers and that would likely serve as umbrella species.
2. Determine the most important environmental and anthropogenic predictor variables that influence suitable habitat for focal mammals by doing a thorough literature review.
3. Create habitat suitability maps for the jaguar *(Panthera onca)* by building species distribution models.
4. Identify core habitat patches for the jaguar *(Panthera onca)* by using the resistant kernel method.
5. Identify potential corridors through the factorial least cost path method.

The remainder of this thesis is organized as follows. Chapter 2 reviews the related research which others have conducted using the methods outlined above for delineating habitat patches and modeling potential corridors. Chapter 3 describes the data and methods used in this study. Chapter 4 describes the results of the study, and Chapter 5 discusses the significance of these results and concludes by noting the main findings related to the focal species.

Chapter 2 : Related Work

This chapter describes work on multispecies connectivity modeling and the techniques commonly used. The four sections that follow describe multispecies connectivity modeling, species distribution modeling with presence only data, potential corridor modeling methods, and focal species selection.

2.1. Multispecies Connectivity Modeling

Multispecies connectivity modeling approaches use different methods to understand how to best connect habitat patches in fragmented landscapes. Some studies combine methods that predict suitable habitat and identify potential corridors. There are different algorithms for predicting suitable habitat, such as the widely used maximum entropy method (MaxEnt), and potential corridors, such as least cost path methods.

DeMatteo et al. (2017), for example, used MaxEnt to evaluate habitat use, habitat suitability, and potential species richness for jaguars (*Panthera onca*), pumas (*Puma concolor*), ocelots (*Leopardus pardalis*), oncillas (*Leopardus tigrinus*), and bush dogs (*Speothos venaticus*) across northern-central Misiones, Argentina. They next determined the optimal location for primary/secondary corridors that would link the northern-central zones of the Green Corridor in Misiones and identified areas within these corridors needing priority management. This study performed a secondary analysis that compared the multispecies corridor results with the jaguar's unique requirements and effectively demonstrated that their "multispecies approach balanced the preferences of all five species and effectively captured areas required by this highly restricted and endangered carnivore" (DeMatteo et al. 2017, 1). They prefer multispecies approaches over a single species approach when creating corridors. They used data collected with dogs that could detect scat and a DNA extraction protocol to identify species from the scat. The optimal locations for primary/secondary corridors were identified by overlaying the five species-specific least cost path and least cost corridor models. They identified a primary corridor with a width of 7 km and a secondary corridor with a width of 14 km. They choose the five aforementioned species because they were identified from scat analysis to be co-distributed in the study area.

Ersoy, Jorgensen, and Warren (2019) demonstrate an approach that identifies multispecies networks and compared it to a previous study in Sheffield, UK. They modeled least cost corridors for four bird, three mammal, and three reptile species. They identified a mix of landcovers important for supporting a maximum number of species and proposed ways that current Sheffield green networks can be improved.

Khosravi, Hemami, and Cushman (2018) identified core habitat areas and connectivity corridors for six species found in the same geographic area in the central Iranian plateau. They used an ensemble model (EM) of habitat suitability using MaxEnt, GLM, GBM, and the biomod2 package in R to predict potential habitats of Persian leopard (*Panthera pardus saxicolor*), Asiatic cheetah (*Acinonyx jubatus venaticus*), caracal (*Caracal caracal*), wild cat (*Felis silvestris*), sand cat (*Felis margarita*), and grey wolf (*Canis lupus*). They then use resistant kernels and factorial least-cost path modelling to predict important core habitats and potential corridors between habitat patches using the UNIversal CORridor network simulator (UNICOR) (Landguth et al. 2012). This program uses Dijkstra’s algorithm (Dijkstra, 1959) to solve the single shortest path problem from each species occurrence point to every other occurrence point (Landguth et al. 2012). The contribution of each core habitat to the connectivity network was assessed using graph network algorithms in Conefor 2.2 (Saura and Tornė 2009). The next section describes species distribution modeling with presence-only data. These three aforementioned examples covered in this section show how species corridor modeling can be combined with species distribution modeling.

2.2. Species Distribution Modeling with Presence-only Data

Species distribution modeling (SDM) has many names including climatic envelope-modeling, habitat modeling, environmental niche modeling, and ecological niche modeling (Hijamans and Elith 2017). SDMs can use species occurrences in the form of presence and/or absence georeferenced locations along with predictor variables (environmental, topographic, climatic, anthropogenic, etc.) with the goal of predicting the species distribution, abundance, presence, or occurrence. The outcome one ends up predicting depends on the modeling method. There are numerous statistical methods that can be used to produce predictions and primarily depend on the type of data that is available. If you have a well-defined sampling method with presence and absence data, you should use statistical methods that were developed to handle those data. If you have information only about a species presence, you should use statistical methods developed to handle presence-only data; however, it is possible to use a modeling method that requires absence data so long as you can substitute absences with background data (Hijamans and Elith 2017). The following review of SDM methods is restricted to presence-only data because these were the data available for this study.

2.2.1. Maximum Entropy Method (MaxEnt) for modeling species geographic distributions

MaxEnt is a maximum-entropy technique that uses a machine learning algorithm called sequential-update which is analogous to the AdaBoost algorithm to predict a species' potential distribution (π) (Phillips, Dudik, and Schapire 2004). The potential distribution π can exhibit sampling bias. This can happen for several reasons such as when some locations are sampled more because of easier access to them. Sampling bias will cause the distribution to be weighted more towards the areas and environmental conditions where species localities were sampled more. For this reason, π and $\hat{\pi}$ are best interpreted as a relative index of environmental suitability (Phillips, Anderson, and Schapire 2006). The unknown distribution π is estimated by $\hat{\pi}$ of maximum entropy which depends on the condition that $\hat{\pi}$ is equal to the empirical distribution $\overline{\pi}$ for all of the features f_j (i.e. $\hat{\pi}[f_j] = \overline{\pi}[f_j]$) (Phillips, Dudik, and Schapire 2004). In reality, $\hat{\pi}$ is close but not exactly equal to $\overline{\pi}$ and the regularization equation estimates how close it is to the empirical value (Phillips, Dudik, and Schapire 2004). The maximum entropy (MaxEnt) distribution equation in which entropy is maximized is equivalent to the maximum likelihood Gibbs distribution equation, which is the distribution that minimizes relative entropy (Kullback-Leiber divergence) (Phillips, Dudik, and Schapire 2004; Phillips, Anderson, and Schapire 2006).

The MaxEnt program, used for maximum entropy modelling of species geographic distributions (Phillips, Dudík, Schapire, nd, Ver. 3.4.1), uses the sequential-update algorithm "that modifies one weight λ_j at a time" (Phillips, Dudik, and Schapire 2004, 667) and converges to the optimal maximum entropy distribution. MaxEnt predicts environmental suitability (or habitat suitability) as a function of the environmental variables by finding the probability distribution of maximum entropy, which is the distribution that is as close to uniform as possible. It uses presence-only data (positive examples), such as from museum records and herbarium collections and does not require information about where species are absent (negative examples) (Phillips, Dudik, and Schapire 2004; Phillips, Anderson, and Schapire 2006). The MaxEnt models show probability distributions over pixels but pixels without species records cannot be considered as places where a species is absent because it is only working with data about presence localities (Phillips, Anderson, and Schapire 2006).

MaxEnt has always been freely available but its source code was not publicly available because it was owned by AT&T. In December 2016, MaxEnt became open-source, and a new version (ver. 3.4.0) was released with an update to include a default log-log (cloglog) transform which can be interpreted as an estimate of the probability of occurrence (Phillips et al. 2017). This new cloglog transform uses a maximum likelihood exponential model obtained from an inhomogeneous Poisson process (IPP). Although it is an unlikely assumption, if sampling effort is unbiased then the new raw MaxEnt output can be interpreted as the relative abundance for the species and the cloglog transformation converts it to the probability of species presence (Phillips et al. 2017) with the following equation:

$$\text{Probability of presence} = 1-\exp(-\exp(H)p_{\lambda}(z)) \tag{1}$$

The previous versions of MaxEnt used the logistic transform as the default output with the following equation to give an estimate of the probability of presence (Phillips and Dudik 2008):

$$Q\,(y = 1|z) = e^{H}q_{\lambda}{}^{(x(z))} / 1 + e^{H}q_{\lambda}{}^{(x(z))} \tag{2}$$

The logistic transform is still available in the new version. Figure 2 shows a graph comparing these two equations.

The species datasets used in this study are from presence-only mammal species occurrence locations recorded from human observations, camera traps, telemetry, and scientific literature without knowledge of species absence locations. This method was proven to perform better than other methods that do not require absence data, such as the Genetic Algorithm for Ruleset Prediction (GARP), which was tested with presence-only species occurrence records for birds and mammals (Phillips, Dudik, and Schapire 2004; Phillips, Anderson, and Schapire 2006).

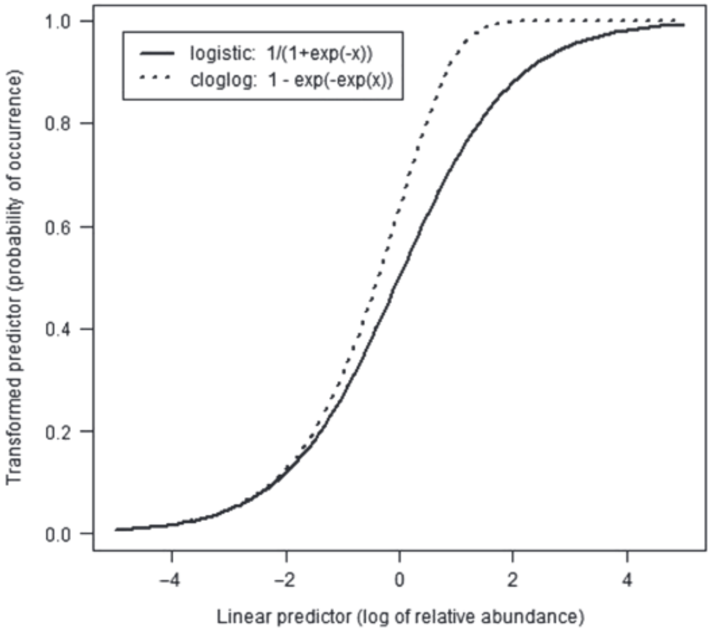

Figure 2 – Graph showing the distribution produced using the cloglog and logistic transforms. Source: Phillips et al. (2017).

2.2.2. Generalized Linear Models

In statistics, a generalized linear model (GLM) describes any model with an expected value (μ) of the response (dependent) variable (Y) with linear explanatory (independent) variables x_1, x_2, … x_p (Upton and Cook 2014).

The following model described in Upton and Cook (2014), is a link function where the parameters $\beta_1, \beta_2, \ldots, \beta_p$ are not known:

$$g(\mu) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p \qquad (3)$$

Nelder and Wedderburn (1972) developed this class of GLMs which includes the mathematical likelihood procedure to fit models based on normal, binomial, Poisson, or gamma distributions.

A GLM is a generalization of ordinary least squares regression (OLS) (Hijamans and Elith 2017). Species presence and absence (or background) data can be analyzed using logistic regression methods commonly used with GLMs (Hijamans and Elith 2017).

Species distribution modeling involves model fitting of species distributions to environmental variables. The environmental variables are the explanatory (predictor) variables, and the response variables are the species distributions. In R (R Core Team 2018), GLMs are fitted using a maximum likelihood procedure and related to response variables through link functions that allow the variance of each measurement to be a function of the predicted value (Hijamans and Elith 2017). These models can be fitted using the glm() function in R (Kabacoff 2017). Table 1 shows the different family types that can be used as part of the glm() function in R to fit GLMs.

Table 1 – List of the eight family types and their associated default link functions to be used with the glm() function in R. The glm function has the form: **glm(***formula***, family=***familytype***(link=***linkfunction***), data=).** Source: Kabacoff (2017).

Family	Default Link Function
binomial	(link = “logit”)
gaussian	(link = “identity”)
Gamma	(link = “inverse”)
inverse.gaussian	(link = “1/mu^2”)
poisson	(link = “log”)
quasi	(link = “identity”, variance = “constant”)
quasibinomial	(link = “logit”)
quasipoisson	(link = “log”)

2.2.3. Generalized Boosted Models

The Generalized Boosted Models (GBM), gbm package in R is based on Friedman’s (2001, 2002) gradient boosting functions such as the gradient boosting machine and Freund and Shapire’s (1997) AdaBoost algorithm (Ridgeway 2019). The gbm package uses the AdaBoost exponential loss function and Friedman’s gradient descent algorithm (Ridgeway 2019). The gbm package supports the Gaussin, AdaBoost, Bernoulli, Laplace, Quantile regression, Cox Proportional Hazard, Poisson, and Pairwise distributions. Friedman (2001, 2002) describes an estimation function which estimates f(x) and includes a loss function $\Psi(y, F(x))$ written in the regression function below as $L(y, F(x))$:

$$F^* = \underset{F}{\arg\min}\, E_{y,x} L(y, F(x)) = \underset{F}{\arg\min}\, E_x[E_y(L(y, F(x))) \mid x] \qquad (4)$$

Sometimes this is written as the regression function that estimates f(x) (Ridgeway 2019):

$$\hat{f}(x) = \underset{f(x)}{\arg\min}\, E_{y|x} [\Psi(y, f(x))|x] \qquad (5)$$

Freedman’s gradient boosting machine contains Friedman’s gradient boost algorithm.

The AdaBoost algorithm is a machine learning algorithm, specifically described by Freund and Shapire (1997) as an adaptive boosting algorithm with the goal of identifying a hypothesis that has a low error relative to the distribution over the training examples. The first hypothesis that is identified is used for obtaining the next weight vector and the process repeats to generate more weight vectors in an iterative process that adjusts adaptively to errors in weak hypotheses (Freund and Shapire 1997).

2.2.4. Random Forests Models

Random forests (RF) is an ensemble machine learning algorithm developed by Breiman (2001). The general process by which RF works is through growing an ensemble of trees, allowing each tree to vote for the most popular class, and finally the forest chooses the classification with the most votes from all the trees in the forest. To grow the ensembles, random vectors are produced independent of past random vectors with the same distribution. The procedures implemented by Breiman's (2001) RF algorithm are called random forests. Breiman (2001) provides a definition for an RF classifier as follows:

> "A random forest is a classifier consisting of a collection of tree-structured classifiers {h(x, k), k = 1,...} where the {k } are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x" (Breiman 2001, 6).

Random Forests has a well-established history in many disciplines but is less commonly implemented in ecological studies. There are numerous benefits to implementing an RF method in ecology including a high classification accuracy, the ability to determine variable importance, the ability to model complex interactions among predictor variables, it can handle thousands of input variables, it is robust against overfitting, it can perform regression and classification even if there are missing data, and it performs well compared to other classifiers. Cutler et al. (2007) used data from three different species to compare the accuracies of an RF model to those of classification trees, a logistic regression, and linear discriminant analysis, and found RF performed the best. Mi et al. (2017) found RF performed better than MaxEnt for predicting rare species distributions with a limited number of samples over a large area and missing data for several Asian crane species. Torres et al. (2012) evaluated 11 SDMs and although all generally had high AUC values (≥ 0.88), the RF model had the highest AUC (0.96) when testing with 30% of the occurrence locations. Although RF is not as commonly used in ecology compared to other fields it provides many advantages and studies have proven it performs better than many other commonly used algorithms.

The next section describes resistant kernel modeling, factorial least cost path modeling, and graph theory. These techniques can be combined with species distribution modeling with presence-only data as well.

2.3. Species Connectivity Modeling

The resistant kernel, factorial least cost path, and graph theory modeling techniques discussed in this section can be used to model species connectivity for either single or multiple species.

2.3.1. Resistant kernel modeling

The resistant kernel modeling (or resistant kernel estimator) approach calculates the expected density of dispersing individual animals in each pixel on a landscape (Cushman, Lewis, and Landguth 2014). The resistant-kernel estimator method is a hybrid between the kernel estimator method and least-cost paths which use resistance surfaces (Compton et al. 2007). To calculate the resistant kernel estimator a least cost kernel for each cell is first calculated. Each cell represents a source for dispersers, for example a vernal pool, and then all kernels in each cell are added together (Compton et al. 2007). Worton (1989) described the fixed kernel and adaptive kernel methods for estimating the utilization distribution in species home range analyses. Patch isolation can be influenced by distance and the type of land cover matrix (Ricketts 2001). Ricketts (2001) described a maximum likelihood method to estimate relative resistances of two types of landcover for butterfly movement. Resistant surfaces are typically created by assigning resistance values to land cover types, while the least-cost path (LCP) method finds the shortest (less) costly distance between two origins (Compton et al. 2007). A variation to the LCP method involves using a multidirectional approach called a least-cost kernel surface because it measures the functional distance from one cell (source of dispersers) to all other cells in a landscape, giving a probability of dispersal (Compton et al. 2007). A cost is assigned to each cover type which represents a cost an animal incurs when moving across that surface. The lower the cost the easier it will be for an animal to move across that landscape surface.

The UNIversal CORridor network simulator (UNICOR; Landguth et al. 2012) program is used for identifying species connectivity and corridors. The UNICOR program uses a modified version of Dijkstra's algorithm which computes the single shortest path from all locations on a landscape that are specified (Landguth et al. 2012). Dijkstra's algorithm solves two problems, it constructs a tree, which is a graph, that has only one path between every two nodes and finds a path of shortest length between two nodes (Dijkstra 1959). UNICOR version 2.0 includes the resistant kernel technique used by Compton et al. (2007) which predicts habitat connectivity and corridor paths using a resistance surface as an input. There are three advantages to using resistant kernels: 1) They predict and map expected movement rates for every pixel in the study area; 2) the scale dependency of a species dispersal ability can be used to understand the effect of landscape change and fragmentation; and 3) they simulate and map different geographic extents using a combination of species (Landguth et al. 2016). The resistant kernel method used in UNICOR uses the modified version of Dijkstra's algorithm to compute the least-cost dispersal around every specified source cell to create expected density surfaces for dispersing individuals at any location on a landscape (Landguth et al. 2016). This is accomplished by creating surfaces of cost to movement for every specified source which is then transformed to indicate a scale from zero to one (Landguth et al. 2016).

2.3.2. Factorial least cost path modeling

A major limitation to traditional least-cost paths and corridor analysis is that there are only two locations considered (i.e. the source and the destination). Factorial least cost path analysis is a type of least cost path method that relies on a spatially synoptic view to understand connectivity (Cushman, Lewis, and Landguth 2014). By adopting a synoptic view one can understand connectivity from multiple locations to all other locations simultaneously. In this way, factorial least cost path analysis helps us to better understand connectivity by calculating least cost paths for "thousands or millions of combinations of locations" across the study area (Cushman, Lewis, and Landguth 2014, 845). Figure 3 shows a factorial least cost path analysis where densities of all paths are shown from blue to red with red representing the lowest cost paths (Rudnick et al. 2012).

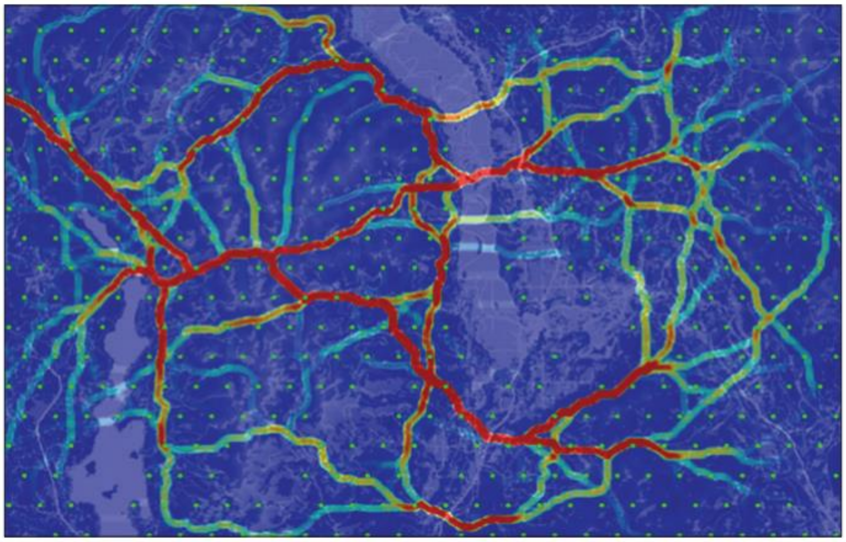

Figure 3 – Example factorial least-cost path analysis. Source: Rudnick et al. (2012)

Both the resistant kernel and factorial least cost path modeling methods can be modeled with the UNICOR software described above. This program can be used for factorial least cost path modeling to calculate the least cost paths for all source pair locations and to create several least cost paths (Cushman, Lewis, and Landguth 2014). The corridor strength is indicated by the higher grid cell values in the raster that is created. Each cell value is calculated by summing the number of cost paths that cross the cell (Cushman, Lewis, and Landguth 2014).

2.3.3. Graph theory

Graph theory is a branch of geometry that originated in 1735 when the Swiss mathematician Leonard Euler solved the Königsberg bridge problem that involved finding a path over seven bridges that traversed a river without crossing any bridge twice (Hosch 2010). Graph theory is now considered a branch of mathematics that is concerned primarily with the statistical description of static networks (Proulx, Promislow, and Phillips 2005). The modern version of the theorem proved by Euler can be stated as follows: "If there is a path along edges of a multigraph that traverses each edge once and only once, then there exist at most two vertices of odd degree; furthermore, if the path begins and ends at the same vertex, then no vertices will have odd degree" (Hosch 2010, 108). In graph theory the term *graph* refers to a set of vertices which are called points or nodes and edges which are the lines connecting vertices (Hosch 2010). A multigraph is one that has any two nodes connected by more than one line and the graph is complete when each one of its nodes is connected to every other node by a line (Hosch 2010). There are also paths in graph theory which can take on different routes in a graph. There are different types of paths (e.g. Eulerian circuits) that can be defined in graphs. Graph theory has applications in many fields ranging from sociology to evolution. It has been widely used in biological networks (Proulx, Promislow, and Phillips 2005). It is useful for problems related to finding optimal paths in a graph, given different criteria, and efficient algorithms (Hosch 2010).

Graph theory has been applied to connectivity analyses in conservation biology. In this case a graph represents a landscape made up of nodes, which could be habitat patches, and lines connecting pairs of nodes, which could represent dispersal (Urban and Keitt 2001).

Pascual-Hortal and Saura (2008) used a graph-based approach to determine functionally connected forest areas within a species distribution and identify the forest habitat areas that are more important for maintaining connectivity for capercaillie. With their results they provided recommendations that could help with the conservation of capercaillie.

Urban and Keitt (2001) used the minimum spanning tree, a graph construct, to understand the relative importance of habitat patches for the Mexican Spotted Owl. They used data from Keitt et al. 1995, USDI Fish and Wildlife Service 1995, to demonstrate the application of minimum spanning trees and found that using this method, large core owl populations and their dispersal routes between core habitats were well maintained.

Another study used a graphed-based approach to understand landscape connectivity indices for prioritizing habitat patches and corridors (Pascual-Hortal and Saura 2006). This study compared 10 graph-based connectivity indices which included a new index, the integral index of connectivity (IIC), with seven different habitat changes, such as habitat patch loss and corridor loss, to see how well each one could identify important landscape elements. They found limitations to existing indices and that the new index was more appropriate because it performed consistently with different habitat changes (Pascual-Hortal and Saura 2006). In a later study, Saura and Pascual-Hortal (2007) looked at the response from nine connectivity indices again including the IIC. This time the IIC came in second place and a new index, the probability of connectivity (PC), placed first, performing consistently across the 13 landscape properties measured.

Conefor Sensinode 2.2 (CS22) (Saura and Torne 2009) is a free software package that is used for quantifying the importance of habitat patches. This program is based on graph theory and works well with geographic information systems (GIS). It can process thousands of nodes and can be used with any standard computer. The time required to complete the analysis will depend on the computer and number of nodes that need to be processed (Saura and Tornė 2009).

The next section describes the species which was selected from the numerous species found to be at risk of border barriers from prior studies.

2.4. Focal Species Selection

Focal species were selected based on available data from various sources and databases including species occurrence data from literature, museums, public organizations, human observations, camera traps, and genetic samples, that indicated their presence occurred across the US-Mexico border and from coast to coast. Menke (2008) created a mountain lion least cost path model (LCP) for locating potential corridors in New Mexico for this felid. The final model found overlap between the mountain lion LCP model of potential corridors and habitat for gray wolf, jaguar, swift fox, and kit fox. This thesis project focuses on the jaguar (*Panthera onca*) and covers a much larger geographic extent than Menke (2008) did. The background for this highly endangered species and prior modeling are described in the two subsections below that conclude this chapter.

2.4.1. Jaguar (Panthera onca) Background

The jaguar (*Panthera onca)* is a keystone species whose historic range once stretched from southwestern United States to southern Argentina, but their range is much smaller today (Seymour 1989) (Figure 4). They are the largest felid predator and the only remaining representative of the genus *Panthera* in the American hemisphere. Presently the jaguar subspecies classification is unclear. Eight subspecies have been recognized by Pocock (1939) and Seymour (1989) but morphological and genetic analysis do not indicate that there are separate subspecies (Larson 1997, Eizirik et al. 2001, Ruiz-Garcia et al. 2006) and Larson (1997) recommends captive jaguars should be managed as a single species.

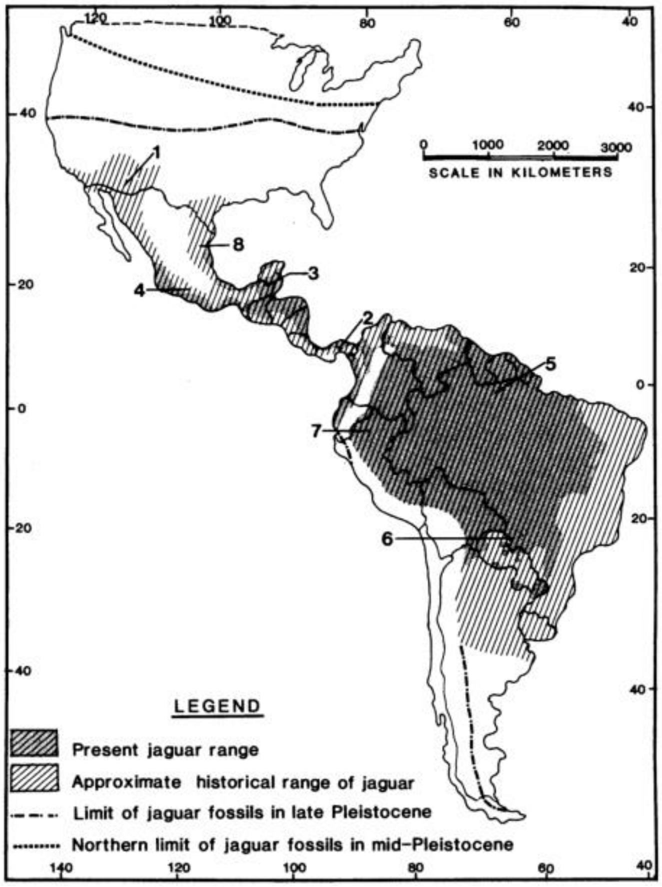

Figure 4 – Map showing approximate current and historic geographic range of the jaguar and the range limits based on found fossils from late Pleistocene and mid Pleistocene. Source: Seymour (1989).

The IUCN Red List classifies jaguars as Near Threatened and they have a decreasing global population trend (Quigley et al. 2017). Major threats to jaguars vary by geographic region and include habitat loss and fragmentation, retaliatory killings due to livestock depredation, illegal body part trade, trophy hunting, and human competition for wild meat (Quigley et al. 2017). Threats to jaguar survival have resulted in severely fragmented populations, a loss of habitat connectivity at local and regional scales, a 49% loss of their historic geographic range, and a 21% loss of their important prey species (white-lipped peccary) (Quigley et al. 2017). Morcatty et al. (2020) found the illegal body part trade for several cat species including the jaguar are connected to Chinese-led development in Central and South America. They found jaguar seizers of body parts to have increased between 2012 – 2018 and found jaguar canines were the most common body part. The main threats to jaguars in northern Mexico include "illegal predator control, illegal hunting, depletion of prey species, and habitat degradation and fragmentation" (Rosas-Rosas and Valdez 2010, 366). In the southwestern United States, the major threats appear to have been livestock settlers. According to Brown (1983), kill data from jaguars indicates that they were eliminated by livestock operators and predator control agents which is concurrent with human settlement and development of the livestock industry. Reports and photographs of jaguars killed in Arizona and New Mexico have been documented as early as 1986 (Brown and López-González, 2000). There are records of jaguars occurring in the southwest US including California, Texas, Arizona, and New Mexico (Brown 1983).

Jaguars have been reported to prey on more than 85 different species with their favorite prey reported as peccaries, capybaras, pacas, agoutis, armadillos, caimans, and turtles (Seymour 1989). In two specific case studies in northeastern Sonora, Mexico, jaguar prey species have been identified but it is not possible to assume that such findings would apply to the entire state of Sonora. In a study area (about 400 km^2) in northeastern Sonora, Mexico, where cattle ranching dominates, jaguars were found to prey mainly on cattle but they also preyed on white-tailed deer, lagomorphs, collard peccary, and coati (Rosas-Rosas, Bender, and Valdez 2008). This study lasted six years and comprised just 27 scat samples that were analyzed by microscope (Rosas-Rosas, Bender, and Valdez 2008). This same study attributed confirmed killings of calfs from ranch records and field surveys to three individual jaguars over a period of six years, and yet there was still a very high calf survival rate averaging 93.3%, since most calves were sold (5728/6136) (Rosas-Rosas, Bender, and Valdez 2008). Cassaigne et al. (2016) investigated the jaguar and puma diet in a study area of about 700 km^2 which partially overlapped the study area of Rosas-Rosas, Bender, and Valdez (2008) using molecular analysis of opportunistically collected scat samples (2012-2013) and kill sites (2011-2013) from two collard jaguars and seven collard pumas. The study found that a variety of small prey weighing less than 15 kg made up the majority of jaguar kill sites (52%) (Cassaigne et al. 2016). Despite a low jaguar density in the study area and just five jaguar scats and kill data for two collard jaguars, they identified a variety of prey species which included birds, deer, calves, coati, and skunks (Cassaigne et al. 2016).

Important native prey for jaguars along the border have been suggested to be Cous white-tailed deer, javelina (a.k.a. collared peccary), coati, opossum, and other medium sized mammals (Brown and López-González 2001). Other examples of borderland jaguar prey have been documented as horse, cattle, elk carrion, white-tailed deer, white-nosed coati, javelina, desert tortoise, frogs, and skunk (Brown and López-González 2001). Potential native prey species for jaguars in the southwestern United States (Arizona and New Mexico) include white-tailed deer, collared peccary, mule deer, coatis, skunks, raccoons, and jackrabbits (Hatten, Averill-Murray, and Van Pelt 2005). Potential domestic prey includes animals such as livestock and horse (Hatten, Averill-Murray, and van Pelt 2005).

2.4.2. Jaguar Modeling Applications

Tôrres et al. (2012) evaluated 11 modeling methods including MaxEnt to predict species distributions and test whether species distribution modeling in general could provide estimates of jaguar population densities in the Neotropics. They used jaguar occurrences from scientific books, research papers, online databases, and field records and complied 1,409 spatially unique jaguar records. The cell size was 0.0417 degrees which is about 4 km^2 because SDM accuracy is limited by "the quality of distribution data and the available climatic and topographic data sets" (Tôrres et al. 2012, 617). The predictor variables used to evaluate all models were precipitation of the coldest quarter, precipitation of the warmest quarter, precipitation seasonality (coefficient of variation), annual precipitation, mean temperature of the wettest quarter, mean temperature of the driest quarter, maximum temperature of the warmest period, minimum temperature of the coldest period, temperature seasonality (coefficient of variation), annual mean temperature, altitude and slope. The number of iterations used was 1,000 and they used pseudo-absences. All the models they evaluated had high AUC values (≥ 0.88), but the Random Forest (RF) model had the highest value (0.96) when testing a subset of 30% of the occurrence locations (Tôrres et al. 2012).

Rosas-Rosas, Bender, and Valdez (2010) found that jaguar cattle kill sites in northeastern Sonora (study site ~ 400 km^2) were positively associated with oak, semitropical thornscrub, and xeric thronscrub and negatively associated with upland mesquite. They also found a positive association of cattle kills sites with proximity to water and roads (Table 2). Other potentially important variables found from the type of vegetation recorded at associated jaguar kill sites in northeastern Sonora, Mexico (400 km^2) include semitropical thornscrub, oak patches, and tropical deciduous forest (Rosas-Rosas and Valdez 2010). Natural prey of jaguar also prefers these types of habitats including white-tailed deer, coatimundi, and collard peccaries (Rosas-Rosas and Valdez 2010).

Table 2 lists important predictor variables found, study information, habitat modeling algorithms used, the input predictor variables, and source for the jaguar modeling studies. Many of these studies use a variety of different SDM algorithms individually or in combination which include MaxEnt, GARP, ENFA, MD, SVM, and EM's.

Important vegetation for jaguars in the borderlands is described in other non-modeling literature from direct observation of the biotic community from jaguars reportedly killed or photographed and include Sinaloan thornscrub (mostly found in Sinaloa), Madrean evergreen woodland (including woodlands of oak and pine), chaparral, and shrub-invaded semidesert grasslands (Brown and López-González 2001). In Arizona and New Mexico, it seems that montane conifer forest and pinon-juniper woodland might also be important for jaguars (Brown and López-González 2001). Other habitat that has been associated with individually studied jaguars with camera traps in the southwestern US includes Sonoran lowland desert, Sonoran desert scrub, mesquite grassland, Madrean oak woodland, and pine-oak woodland (McCain and Childs 2008).

There have been potential corridor models created for jaguars of varying scales, including continental, countrywide, and finer scales. Rabinowitz and Zeller (2010) created a model of potential corridors for the jaguar at a continental scale (or range-wide scale) using GIS and expert opinion to create a cost surface and identify least cost corridors connecting the 90 known jaguar populations from northern Mexico to northern Argentina. They identified 182 potential corridors ranging between 3 to 1,607 km in length. They used six datasets for creating a

Table 2 – Jaguar habitat modeling studies

Important variables found	Study information	Programs used	Input predictor variables	Source
Oak, semitropical thornscrub, and xeric thronscrub, proximity to water, proximity to roads (positive) and upland mesquite (negative)	**Location:** northeastern Sonora, Mexico **Extent:** 400 km^2 **Years:** 1999-2004 **Occurrences:** 45 confirmed jaguar cattle kill-sites **Accuracy:** unknown	MaxEnt Best model found: vegetation type, distance to permanent water type, and distance to roads.	Elevation, vegetation cover type (primary oak forest, semitropical thronscrub, xeric thornscrub, disturbed semitropical thornscrub, tropical deciduous forest, mesquite, and disturbed oak forest), distance to roads, distance to permanent water sources (perennial streams and rivers, springs, ponds, and permanent water developments), elevation, slope, and aspect. Resolution: unknown	Rosas-Rosas, Bender, and Valdez (2010).
Elevation between 1,220 and 1,829 m with scrub grasslands, elevation and biomes, distance to water within 10 km or perennial or intermittent water sources (caution: Euclidean distances measured in mountainous terrain), and terrain ruggedness (intermediate to extreme ruggedness)	**Location:** Arizona **Extent:** 295,234 km^2 **Years:** 1901-2001 **Occurrences:** 57 jaguar sightings alive or dead **Accuracy:** < 8 km (<1.7 km to 8 km)	ArcGIS – habitat suitability map	Coarse-scaled: vegetation biomes (ecosystems) and series (defined by dominant or characteristic species), elevation and terrain ruggedness, proximity to perennial or intermittent water sources (streams, rivers, lakes, or springs), and human density. Resolution: 30 m DEM resampled to 1 km^2 cells.	Hatten, Averill-Murray, and van Pelt (2005).
Important variables not possible to obtain from discriminant analysis. Important variables obtained from histograms: precipitation, elevation, slope, and temperature, shrubland, grassland, and forest (females: shrubland, deciduous broadleaf forest, and grassland; but also, needleleaf forest and mixed forest).	**Location:** Arizona, New Mexico, Texas panhandle, Sonora, and Chihuahua. **Extent:** approximately 1.1 million km^2 **Years:** ~1900 -2003 **Occurrences:** 142 (100 male, 42 female) jaguar occurrence records from museums, photos, verified kills, universities, conservation organizations, interviews with residents in Mexico. **Accuracy:** 25km^2	Genetic Algorithm for Rule Set Production (GARP) 3 models: Males + Females Males Females	20 environmental layers for climate and landscape, including temperature, wetness, vapor pressure, frost days, snow accumulation, radiation, soil type, elevation, aspect, slope, compound topographic index, water flow, and runoff Resolution: resampled to 25km^2 pixels	Boydston and Lopez-Gonzalez (2005).
From ENFA and MaxEnt: Prefers: Tropical rain forest, prey, and regularly flooded vegetation, Avoids: Higher Elevations, arid vegetation, and grassland	**Location:** Mexico **Extent:** 1,972,550 km^2 **Years: Calibration dataset:** 1990 – 2008 **Evaluation dataset:** 2000 – 2008 **Occurrences: Two data sets: 1) Calibration: 197** (1 data point per locality, 1 record of presence for each 5 km^2) occurrences from literature, CONABIO,	Ensemble Model (EM) Ecological Niche Factor Analysis (ENFA) Mahalanobis distance (MD) MaxEnt	6 environmental and anthropogenic factors assumed to be important Dry forest, tropical rain forest, other forest, arid vegetation, grassland, regularly flooded vegetation, agriculture, anthropogenic perturbation: roads (distance) and human population density,	Rodríguez-Soto et al. (2011).

	GBIF, MaNIS, and Jaguar GIS. **2) Evaluation: 104** from VHF-locations for 5 collard adult female jaguars in 5 different regions, camera-traps from 8 additional regions. **Accuracy:** unknown		elevation, slope, prey-species richness. Resolution: All layers resampled to 1-km^2 cell size.	
From MaxEnt: positive relation to risk of attack by jaguar: tree cover percentage, percentage of animals in free grazing areas, and altitude Negative correlation to risk of attack by jaguar: arid vegetation	**Location:** Mexico **Extent:** 1,953,162 km^2 **Years:** 1990 – 2010 **Occurrences:** 222 from felid attacks on livestock (jaguar – 152, puma – 70) **Accuracy:** unknown	Jaguar: MaxEnt, GARP-with best subsets, and Support Vector Machines (SVM) Puma: MaxEnt, Environmental distance, ENFA, GARP- with best subsets, and GARP – single run	Landscape, livestock management, and anthropogenic: Topographic: altitude and slope Vegetative associations of the National Forest Inventory: forest (conifers, oaks and riverside vegetation), dry forest, rainforests, arid vegetation, underwater vegetation, and agriculture Percent tree cover Livestock density, percentage of free grazing Human population density, and distance to paved roads. 1 datum per pixel Resolution: 1 -km^2	Zarco-González et al. (2013).

permeability matrix which was used to create a cost surface to be used in a least cost path analysis (Table 3).

Other studies have focused on countrywide scale corridor studies such as in Mexico (Rodríguez-Soto, Monroy-Vilchis, and Zarco-González 2013), or a biome-scale in Brazil (Morato et al. 2014), and for smaller study areas within countries such as Nicaragua (Zeller et al. 2011) and Argentina (DeMatteo et al. 2017). Rodríguez-Soto, Monroy-Vilchis, and Zarco-González (2013) used a previously created ensemble model (Rodríguez-Soto et al. 2011, Figure 2) of the potential distribution of the jaguar in Mexico and identified jaguar management and conservation areas (JCMA) which they used for identifying potential jaguar corridors between them in their 2013 study (Table 3). The variables used for the ensemble model in Rodríguez-Soto et al. (2011) are listed in Table 2. This ensemble model of potential jaguar distribution in Mexico was used as a cost raster or permeability map to create the potential jaguar corridors in Mexico (Rodríguez-Soto, Monroy-Vilchis, and Zarco-González 2013).

Rodríguez-Soto, Monroy-Vilchis, and Zarco-González (2013) used the inverse of the habitat suitability map previously created from an ensemble model (Rodríguez-Soto et al. 2011) as the permeability (cost raster) map. They identified JCMAs as habitat patches using the Corridor Designer in ArcGIS. The permeability map and habitat patches were used as inputs to create the potential jaguar corridors for Mexico by calculating the cost-distance of each pixel using Corridor Designer in ArcGIS. In contrast, Rabinowitz and Zeller (2010) assigned cost values to landscape layer pixels obtained from 15 jaguar experts. The cost values ranged from 0, which indicates no cost to jaguar movement to a value of 10, which indicates a high cost for jaguar movement as shown in Table 4. A final cost surface or permeability matrix was created by using the Raster Calculator in ArcGIS and reclassifying the output raster. They created movement cost grids from each of the 90 Jaguar Conservation Units (JCU) and delineated least cost corridors using the Corridor function from the Spatial Analysist toolbox in ArcGIS (Rabinowitz and Zeller 2010).

Table 3 – Data layers used by different studies for creating jaguar permeability matrices

Study / Notes	Input layers	Dataset name and scale	Year of data	Data Source
Rabinowitz and Zeller (2010) datasets for creating the jaguar permeability matrix / **Extent:** Continental (northern Mexico to northern Argentina)	Elevation	Global 30 arc-second elevation data set 1 km resolution	1996	Center for earth resources observation and science (EROS)
	Landcover type	Global land cover 2000 1 km resolution	1999–2000	Global land cover 2000
	Percent tree and shrub cover	Continuous vegetation fields 500 m resolution	2000	Global land cover facility
	Population settlements	Vector map level 0 population settlements 1:1,000,000 scale	1960s–1990s	National imagery and mapping agency (NIMA)
	Human population density	Gridded population of the world v3 2.5 min resolution	2000	Center for international earth science information network (CIESIN)
	Roads	Vector map level 0 roads 1:1,000,000 scale	1960s–1990s	National imagery and mapping agency (NIMA)
Rodríguez-Soto, Monroy-Vilchis, and Zarco-González (2013) variables **Extent:** country of Mexico * variables for layers and datasets are from Table 1 in Rodríguez-Soto, Monroy-Vilchis, and Zarco-González (2013), but see variables listed in Table 2 for Rodríguez-Soto et al. (2011).	*Vegetation cover	National Forest Inventory 1:250 000	2001	SEMARNAT et al.
	*****Human disturbance: Agriculture, Road network, and Human population density	Agriculture, Road network, and Human population density	2001, 2008, and 2005	SEMARNAT et al., CONABIO, and FAO
	*Protected Natural Areas (PNA)	Protected Natural Areas	2007	Consejo Nacional de Areas Protegidas (CONAP)
	*Elevation	Digital elevation model	2007	U.S. Geological Survey (USGS)

Table 4 – Cost values (cost to jaguar movement) obtained from 15 jaguar experts assigned to various landscape layers, from Rabinowitz and Zeller (2010).

Landscape Layer	**Cost Value**	**Landscape Layer**	**Cost Value**
Land cover type		*Tree and Shrub Cover (%)*	
Tree Cover, broadleaved, evergreen	0	0 – 10	9
Tree Cover, broadleaved, deciduous	0	10- 20	7
Tree Cover, needle- leaved, evergreen	1	20 - 40	5
Tree Cover, mixed leaf Type	0	40 - 60	2
Tree Cover, regularly flooded, fresh water	2	60 - 80	0
Tree Cover, regularly flooded, saline water	2	80 - 100	0
Mosaic: Tree cover/other natural vegetation	1	*Human Population Density (people/Km2)*	
Shrub Cover, evergreen	2	0-20	1
Shrub Cover, deciduous	3	20-40	5
Herbaceous Cover	5	40-80	7
Sparse herbaceous or sparse shrub cover	6	80-160	9
Regularly flooded shrub and/or herbaceous cover	5	160-320	10
Cultivated and managed areas	8	>320	N/A
Mosaic: Cropland/Tree Cover/ Other natural Vegetation	5	*Elevation (m)*	
Mosaic: Cropland/Shrub or grass cover	7	0 – 1000	0
Bare areas	8	1000–2000	2
Water Bodies	6	2000–3000	7
Snow and Ice	N/A	3000–5000	10
Artificial surfaces and associated areas	10	>5000	N/A
Distance from Roads (Km)		*Distance from Settlements (Km)*	
0 to 2	7	0–2	8
2 to 4	4	2–4	5
4 to 8	2	4–8	4
80 to 160	1	8-16	1
> 16	0	> 16	0

Chapter 3 : Methods

3.1. Research Design

The overarching research question that this study hopes to answer is: How will new border barriers affect connectivity for jaguars? By answering this question, it will be possible to identify dispersal corridors which may be disrupted by new border barriers. Additionally, these results can be compared to the connectivity for the eight co-distributed charismatic mammals (orders: Carnivora and Artiodactyla): mountain lion *(Puma concolor),* ocelot (*Leopardus pardalis*), bobcat *(Lynx rufus),* black bear *(Ursus americanus),* gray fox *(Urocyon cinereoargenteus),* Mexican gray wolf *(Canis lupus baileyi)* Sonoran pronghorn *(Antilocapra americana sonoriensis),* and Bighorn sheep *(Ovis canadensis.)* All data were processed using ArcGIS Pro (Version 2.3.2, Esri Redlands, CA) and Google Earth Engine 2019 (GEE, Google, Mountain View, CA). All modeling and analysis tasks were completed using two software programs: R (Version 3.6.0 (2019-04-26) -- "Planting of a Tree"), and UNICOR (Version 2.0, 2016).

The general methodology which was employed was inspired by and constitutes a variation to that used by Khosravi, Hemami, and Cushman (2018). This earlier study focused on a small geographic region unlike this thesis project which focused on a greater geographic extent and therefore required a revised methodology. Different species data points cover different areas along the US-Mexico border ecoregions which can help meet the intended goal of understanding how new border barriers would affect connectivity for the jaguar and several border mammals that may be impacted by new border barriers. The realization that the occurrence points for several individual species span the US-Mexico border helped define a study area. For example, jaguar records are found in the above mentioned US and Mexican states, but they are also found in Texas and northeastern Mexico. The desire to understand how the connectivity of border mammals would be affected by border barriers motivated this study which followed the US-Mexico border from coast to coast.

Important predictor variables were identified from scientific literature for the jaguar and several focal mammals from relevant study sites, in, or as close as possible to the borderlands of US and Mexico. This approach provided a greater probability of selecting appropriate predictor variables for use in the chosen study area. However, literature was also reviewed for focal species from other geographic areas, such as Argentina, Brazil, and other central and south American countries. It is important to select appropriate covariates to build habitat suitability maps for the chosen species. Building habitat suitability maps helps to identify core habitat patches for specific species by using the resistant kernel method. Potential corridors can be identified using the factorial least cost path method, and the importance of patches and corridors can be assessed using graph network algorithms.

3.2. Focal Species Datasets

Several species of concern whose dispersal, migration, and gene flow might be impacted, have been identified and are listed in Chapter 1 and described in Chapter 2. The focal species are jaguar *(Panthera onca),* mountain lion *(Puma concolor),* ocelot (*Leopardus pardalis*), bobcat *(Lynx rufus)*, black bear *(Ursus americanus),* gray fox *(Urocyon cinereoargenteus)*, Mexican gray wolf *(Canis lupus baileyi)* Sonoran pronghorn *(Antilocapra americana sonoriensis),* and Bighorn sheep *(Ovis canadensis).* These species represent the mammalian orders, Carnivora and Artiodactyla.

The occurrence localities that were used for the jaguar and other focal species were obtained from scientific literature, museums, universities, institutions, online databases, and a conservation organization. This includes occurrence data with GPS coordinates from photographs, camera trap image/video data, scat, hair, tracks, reports, preserved specimens, species mortalities, genetic sampling location data, human observations, hunter harvested individuals, road kills, and GPS telemetry. Three main online databases were used to obtain most of the occurrence data: (1) the Global Biodiversity Information Facility (GBIF); (2) VertNet; and (3) the Jaguar Observations Database (JOD). Occurrence records were cleaned using ArcGIS Pro by removing occurrence points found in the ocean, records with high uncertainties, fossils, zoo records, duplicate records, and by examining metadata files. Final cleaned species records were saved in separate AcrGIS Pro geodatabases including for the jaguar which was used in this thesis project.

3.3. Focal Species Covariate Datasets

There are many studies that advocate for multi-scale optimization when modeling species habitats because species may respond differently to predictor variables at different spatial scales. Multi-scale habitat modeling refers to the scale(s) that is(are) important to an individual organism because of how the individual interacts with the environment. Wiens (1989) mentions that when studies asking the same questions are conducted at different scales their findings are not always consistent. Different species may respond differently at different scales which can lead to issues when designing nature reserves. As Wiens (1989, 385) mentions, “the very foundation of geography is scaling”. The term “scale” is context specific but regarding spatial analysis and modeling, O’Sullivan and Perry (2013) define it as a term that describes spatial grain, spatial extent, temporal grain, and temporal extent. Spatial grain refers to the resolution used to collect the data which the data (e.g. pixel, or cell size), temporal grain refers to the frequency used to collect the data, spatial extent refers to the total area that the dataset covers, and temporal extent is the date range over which the data were collected (O’Sullivan and Perry

2013). When considering spatial scale in ecological studies “expanding the extent of a study usually also entails enlarging the grain. The enhanced ability to detect broad-scale patterns carries the cost of a loss of resolution of fine-scale details” (Wiens 1989, 387). When spatial scale is increased temporal scale should also be increased since processes will operate at slower rates. The acquisition of occurrence data for a broad temporal range supports the large spatial extent chosen for this study. Finally, Wiens (1989) recommends that ecologists should adopt a multiscale approach to species studies.

Several studies have found that multi-scale species models outperform single-scale species models in terms of predictive power (e.g., McGarigal et al. 2016; McGarigal, Zeller, and Cushman 2016; Timm et al. 2016; Wan et al. 2017) and provide greater predictive capacity (Timm et al. 2016). For models to be robust they should include a full set of “covariates relevant to habitat selection by the species as is possible” and at spatial scales important to the focal species (Timm et al. 2016, 1210). This can be accomplished by varying the bandwidth for each covariate and each location in ArcGIS Pro. Unfortunately scale optimization procedures are very rarely used in habitat modeling studies (McGarigal et al. 2016). Timm et al. (2016, 1210) recommends using multi-scale models to reduce “investigator-driven bias”. This author determined the covariates from literature reviews and discussions with species experts. This thesis project also used literature reviews to determine important variables and understand the jaguar’s ecology. Feedback for variable selection was received from Dr. Cushman. This approach provided a full list of the best possible variables to use and the inclusion of species-important scale optimization provided a more robust modeling framework.

All covariate datasets required varying levels of preprocessing with ArcGIS Pro and GEE prior to implementing a multi-scale optimization method. Figure 5 shows the average human population density covariate layer for the period 2000-2020. It was created using ArcGIS Pro Cell Statistics and 5 input rasters for the years 2000, 2005, 2010, 2015, and 2020. A total of 52 covariate layers were ultimately processed.

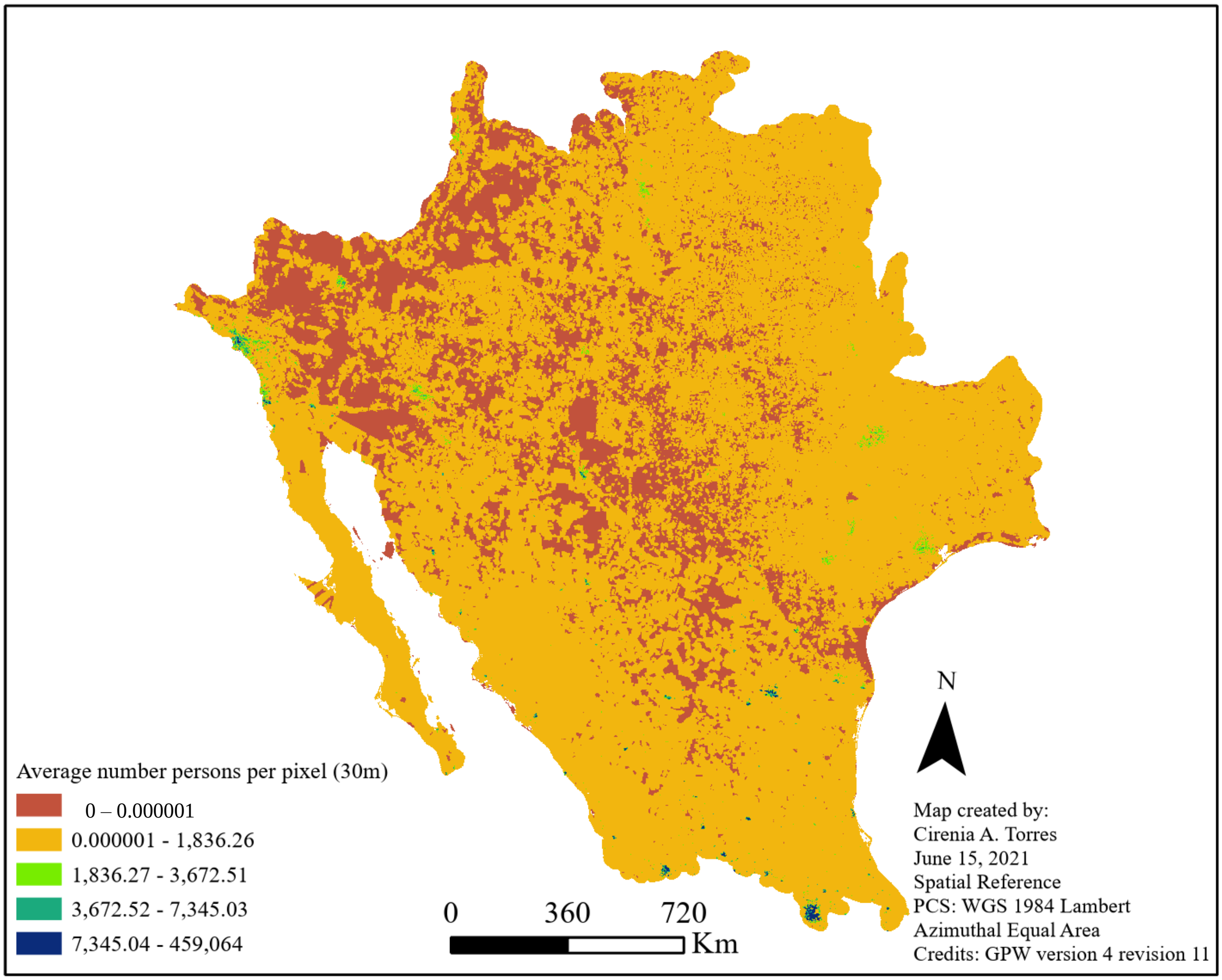

Figure 5 – Average human population density for the period 2000 - 2020. Average number of persons per 30m pixel, 2000 – 2020. Source: GPW ver. 4 rev. 11.

Eight anthropogenic, two climate, 15 ecoregion, 10 land cover, six topographic, seven vegetation, and four water covariate layers were further processed at five different scales, to generate a total of 260 covariates. Dr. Wan's python script was used for multi-scale optimization.

Results for some of the 260 covariates processed are shown in Figures 6 to 9. The univariate scaling of the compound topographic index (CTI) and mean annual temperature layers at 16 km are shown in Figures 6 and 7. Results for the univariate scaling of slope position and roughness layers at 8 km are shown in Figures 8 and 9. Roughness was rescaled from 1 to 10 using a log transform function for proper visualization of landscape features with low (1) to high (10) surface roughness (Figure 9).

The slope position, roughness, and compound topographic index were processed using ArcGIS Pro and the ArcGIS Geomorphometry and Gradient Metrics toolbox (Evan et al. 2014). Slope position calculates the scalable slope position by subtracting a focal mean raster from the original elevation raster based on Berry's (2002) methodology for calculating a surface area ratio. Slope position values range from low (negative) to high (positive). Evan et al. (2014) describe roughness as a representation of a continuous raster within a specified window and this metric is based on the research conducted by Riley, DeGloria and Elliot (1999) and Blaszczynski (1997). Roughness (or ruggedness) is an estimate of terrain heterogeneity which is an important variable for predicting species potential habitats and densities (Riley, DeGloria, and Elliot 1999). It is related to the Terrain Ruggedness Index (TRI) which describes level ground to extremely rugged terrain. The CTI is a steady state wetness index and a function of the slope and upstream contributing area per unit width orthogonal to the flow direction. CTI is described by the following equation:

$$\mathrm{CTI} = \ln\left(As / (\tan(\beta)\right) \qquad (6)$$

where As is equal to the area value calculated as flow accumulation plus 1 multiplied by the pixel area in square meters. β is the slope expressed in radians (Evan et al. 2014).

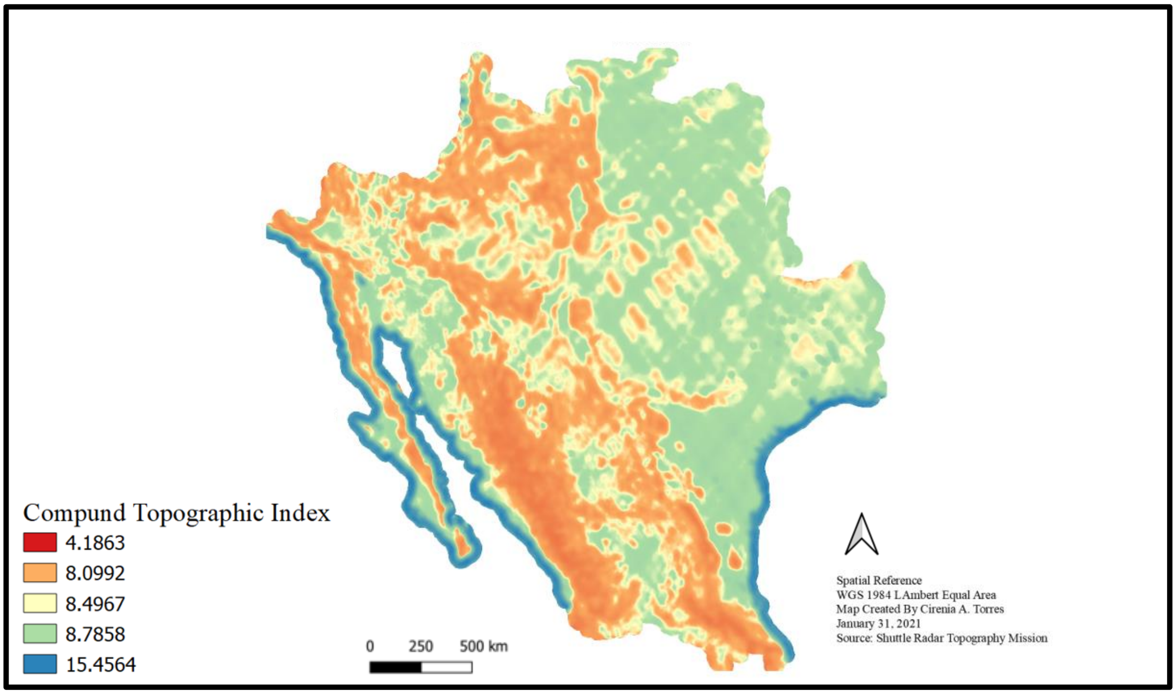

Figure 6 –16 km univariate scaling for the compound topographic index.

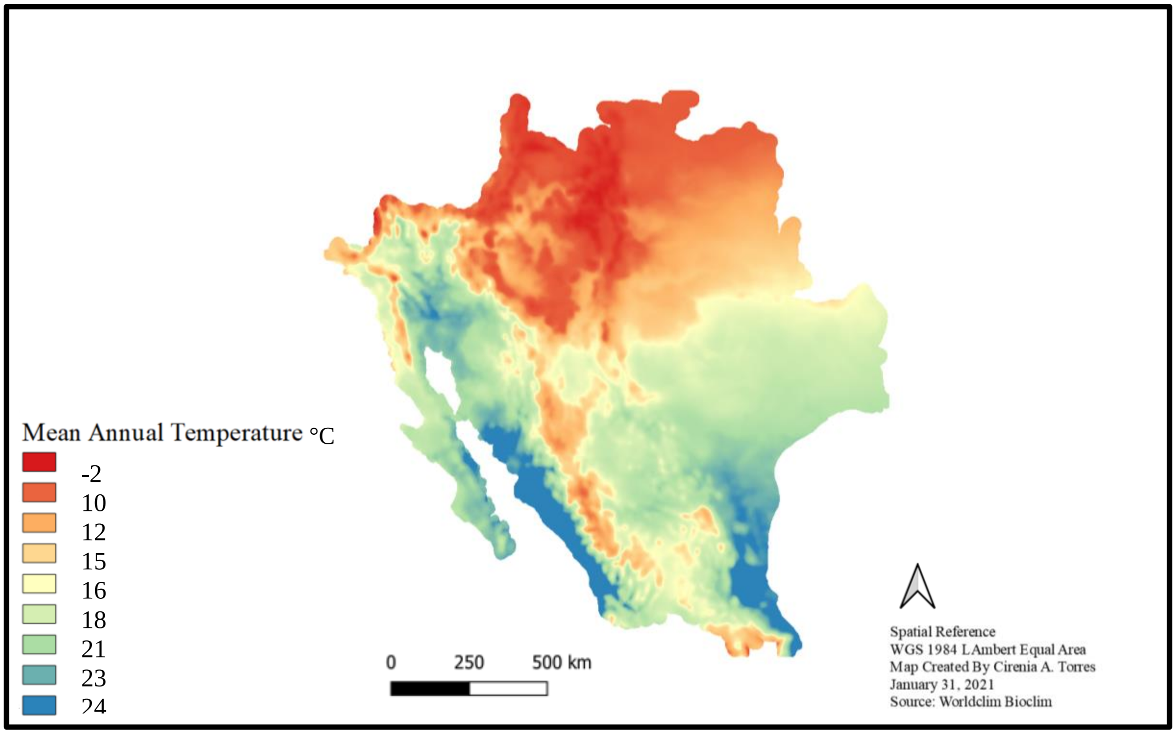

Figure 7 – 16 km univariate scaling for annual mean temperature.

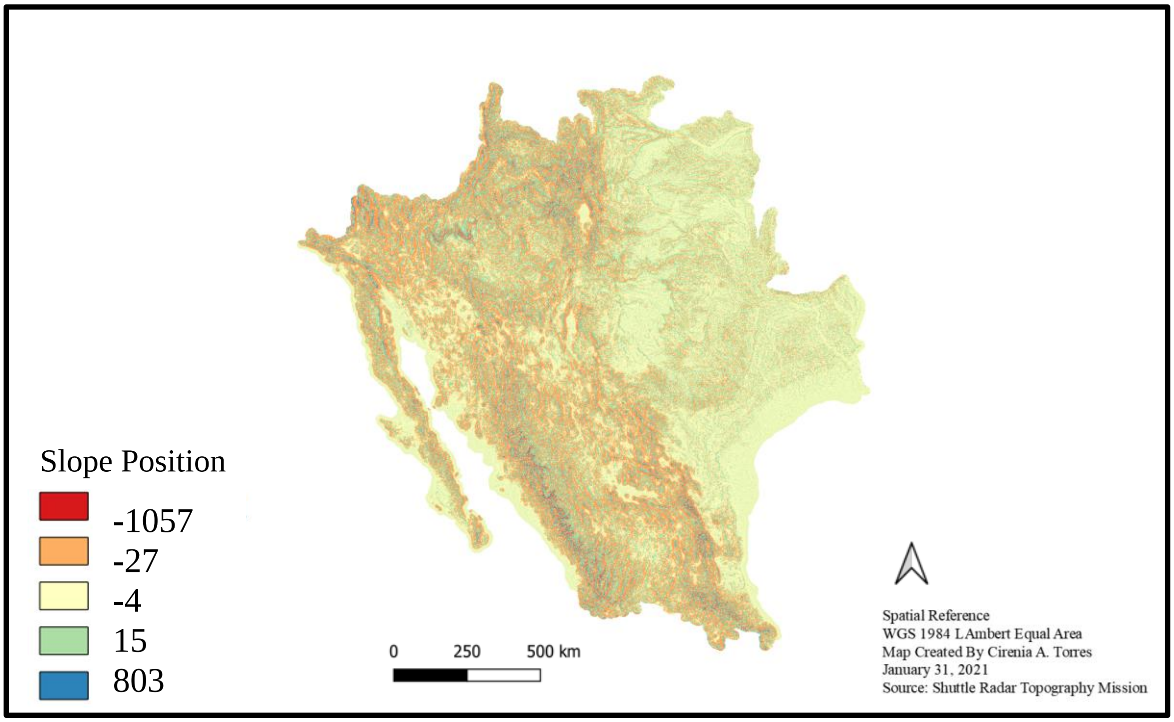

Figure 8 – 8 km univariate scaling for slope position.

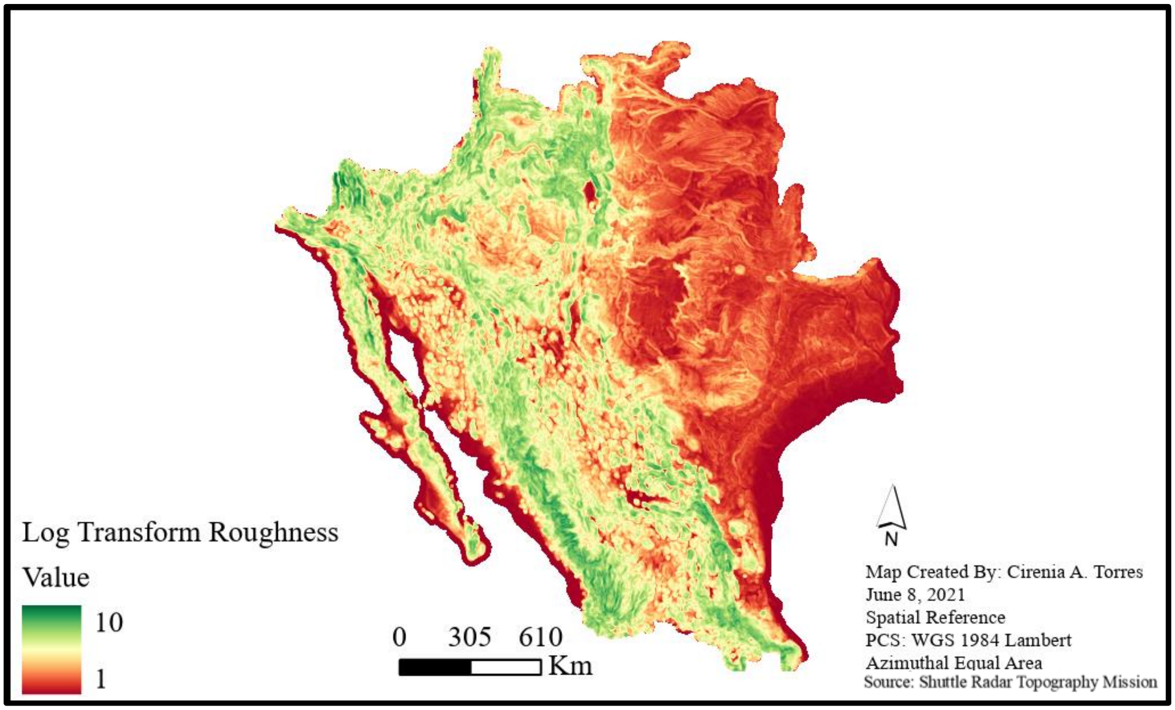

Figure 9 – 8 km univariate scaling for roughness.

3.4. Focal Species Habitat and Corridor Modeling

The focal species suitable habitat models and corridor modeling followed the methodology of Khosravi, Hemami, and Cushman (2018) with some modifications. A Random Forest model was built and a resistance surface was created for input into the UNICOR program. Scale optimization was performed for each covariate layer determined to be important to the jaguar once the final set was determined.

3.4.1. Random Forests

After reviewing several commonly used SDM algorithms as described in Chapter 2 the Random Forests (RF) algorithm was chosen. Random forests is a machine learning algorithm that generates multiple tree predictors where each tree depends on the values of a random vector which is sampled independently and with the same distribution for all trees of a forest (Breiman 2001). It can perform both classification and regression accurately even if there are missing data. The improvements to classification documented in numerous applications can be attributed to the growth of an ensemble of trees and "letting them vote for the most popular class" (Breiman 2001, 5). Random Forests models will not be overfit, and they will work with large datasets of higher dimensionality. The growth of the trees occurs by allowing random vectors to be generated leading the way for the ensemble tree growth. The random vector is generated independently but with the same distribution as the past random vectors. The error rates accompanying the random selection of features to split up each node are comparable to the error rates of the Adaboost algorithm, but these methods are more robust when dealing with noise (Breiman 2001).

The Random Forest R package "randomForest" is an ensemble machine learning algorithm for classification and regression which implements the Breiman (2001) random forests algorithm and it is based on Breiman’s and Cutler’s Fortran code (Breiman et al. 2018). Random Forests was used because it is a strong classifier that uses a robust algorithm and prior studies have used this approach. For example, Mi et al. (2017) found that it performed better than MaxEnt for predicting rare species distributions with a limited number of samples over a large area and missing data for several Asian crane species. Torres et al. (2012) evaluated 11 SDMs and although all generally had high AUC values (≥ 0.88), the RF model had the highest AUC (0.96) when testing with 30% of the occurrence locations. Random Forest has an established history and good predictive performance ability. One great feature of MaxEnt is the capability to perform a jackknife analysis to identify the most important predictor variables; however, Random forest can use the extractor function to measure variable importance. Dr. Wan’s R script with a slight modification by the author was used to create the Random Forest potential habitat model. The author added some lines of code to make use of multiple computer cores to speed up the prediction map output.

3.4.2. Calculating resistant kernels and factorial least-cost paths

The resistant kernels and factorial least-cost paths were generated using the UNICOR program described in Chapter 2. Once the SDM was created for the jaguar, the habitat suitability model was converted to a resistance map following the methodology of Khosravi, Hemami, and Cushman (2018). The predicted habitat patches were then determined and a sensitivity analysis was performed to determine the robustness of predicted patches that reflect the dispersal abilities of the jaguar.

3.4.3. Using the graph network algorithm

The graph network algorithm described in Chapter 2 was used next to determine the contribution of core areas to the network following the methodology of Khosravi, Hemami, and Cushman (2018).

3.5. Model Selection, Validation, and Evaluation

Both the receiver operating characteristic (ROC) and area under the ROC curve (AUC) have been widely used in SDMs to evaluate model accuracy. In general, the ROC is a curve that plots the true positive rate (TPR) on the y-axis against the false positive rate (FPR) on the x-axis. The ROC is used for visualizing the performance of a binary classifier widely used in machine learning. The TPR is also known as the sensitivity (or the probability of detection) and the FPR is also known as "1-specificity" (or probability of false detection). The AUC summarizes the performance of the classifier. AUC values closer to 1 indicate better model performance (Phillips, Dudík, and Schapire, nd). Generally, a model is considered good if it has an AUC value greater than 0.75 (Elith et al. 2006). The ROC and AUC was used to determine the success of the models to predict occurrence patterns.

Cohen's (1960) Kappa is a coefficient of interjudge agreement for nominal scales and has the following equation:

$$\mathrm{k} = P_o - P_c \,/\, 1 - P_c \qquad (7)$$

where P_o is equal to the proportion of units in which the judges are in agreement, P_c is equal to the proportion of units in which the agreement is expected by chance, therefore k is the proportion of agreement after chance agreement has been removed. It can measure the agreement between predicted presences and absences (or pseudo-absences) with the actual presences and absences (or pseudo-absences) corrected for agreement that might occur only by chance. It has a range between -1 to +1 and if the value is less than 0 this indicates that the agreement is less than expected by chance (or performance is no better than by random chance). Since SDMs were created using RF with presence-only data, a methodology like that used by Evans and Cushman (2009) and Mi et al. (2017) was used to validate SDMs in this thesis project.

Chapter 4 : Results

This chapter describes the results for the RF habitat model, resistance surface, and the UNICOR corridor model for the jaguar.

4.1. Jaguar Habitat Map

The jaguar habitat map was modeled using 51 covariates at the optimal scale (i.e the lowest Out of Bag (OOB) error rates from univariate scaling). The multicollinearity test found three of the 51 covariates to be correlated and these were removed from further analysis. Those variables were the Enhanced Vegetation Index (EVI) at 16 km for all years averaged, the Normalized Difference Vegetation Index (NDVI) at 16 km for all years averaged, and major roads and links at 16 km. The data with removed variables was found to be balanced and the final most parsimonious model determined with the model improvement ratio which was not fitting noise selected a model with 36 variables. The selected model had an Area under the ROC curve value of 0.852, a cross-validation Kappa value of 0.7073, a cross-validation OOB Error of 0.147561, and a cross-validation error variance of 1.642579e-05. Figure 10 shows the final jaguar potential habitat map. The color gradient displays areas of high probability of occurrence (red) to low (blue) for the jaguar.

Results predict there is good habitat for jaguars in the Sonoran-Sinaloan subtropical dry forest, Sinaloan dry forests, Sierra Madre Occidental, California montane chaparral and woodlands, Arizona Mountains forests, Sierra Madre Oriental pine-oak forests, Veracruz moist forests, Sierra de la Laguna pine-oak forests, Sierra de la Laguna dry forests, Tamaulipan matorral, and small portions of the Sonoran desert ecoregions.

Figure 11 shows the results for the scaled variable importance graph (a), bootstrap error convergence (b), and the ROC curve (c) for the 36 selected covariates. The scaled variable

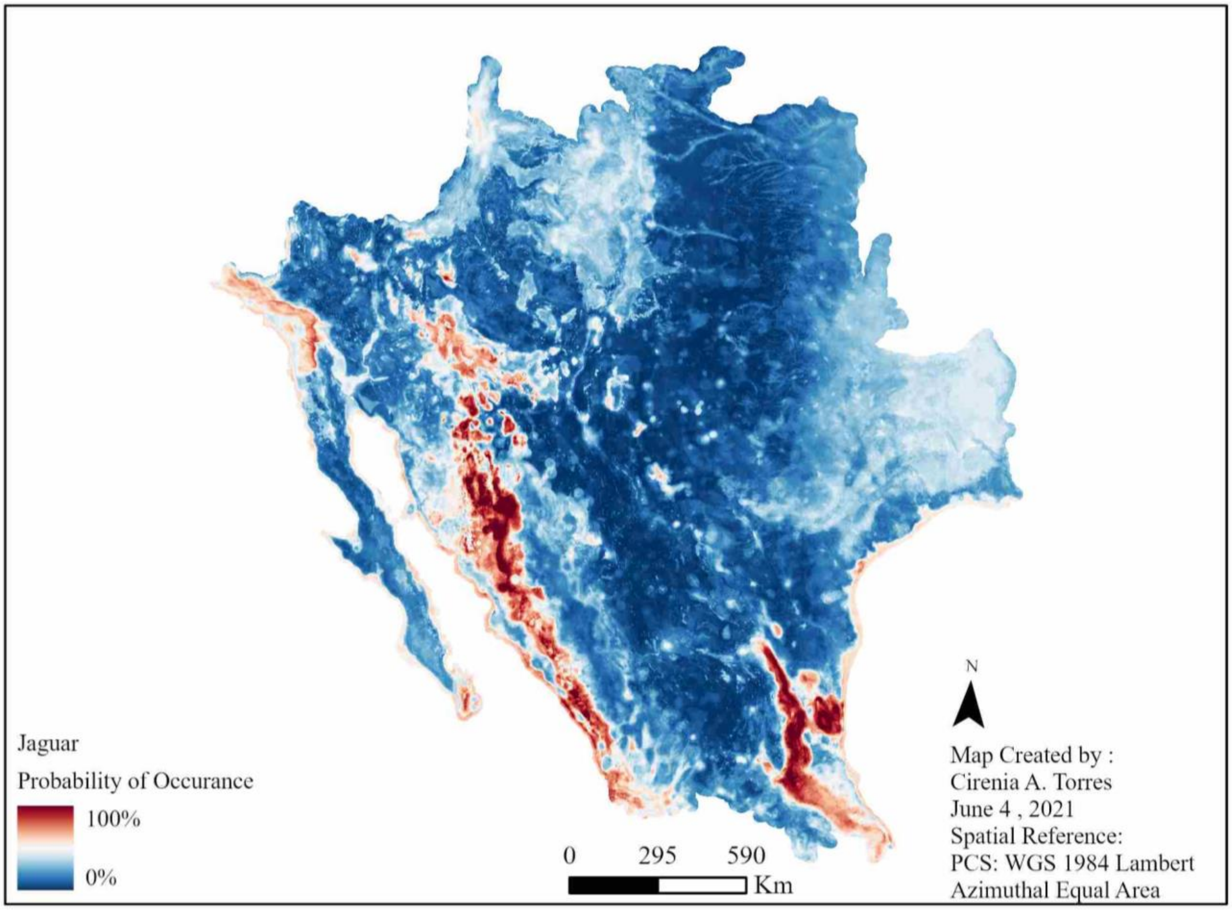

Figure 10 – Jaguar potential habitat map in the US/Mexico border ecoregions from the RF model. The color gradient shows the probability of species occurrence.

importance graph (b) lists the variables scaled from less (0) to more (0.05) important for the RF model. The bootstrap error convergence graph (b) shows the convergence of bootstrap error estimates with the error on the y-axis and the number of trees on the x-axis. Lastly, the ROC curve (c) graphically displays the hit rate (y-axis) and the false alarm rate (x-axis). These results provide confidence in the RF model.

The partial dependency plots for Pav_16km, u17_16km, and MaRd_16km and a presence / absence proximity matrix are shown in Figure 12. The partial dependency plots (a) – (c) indicate the probability for either Pav_16 km, u17_16 km, or MaRd_16 km. The presence /

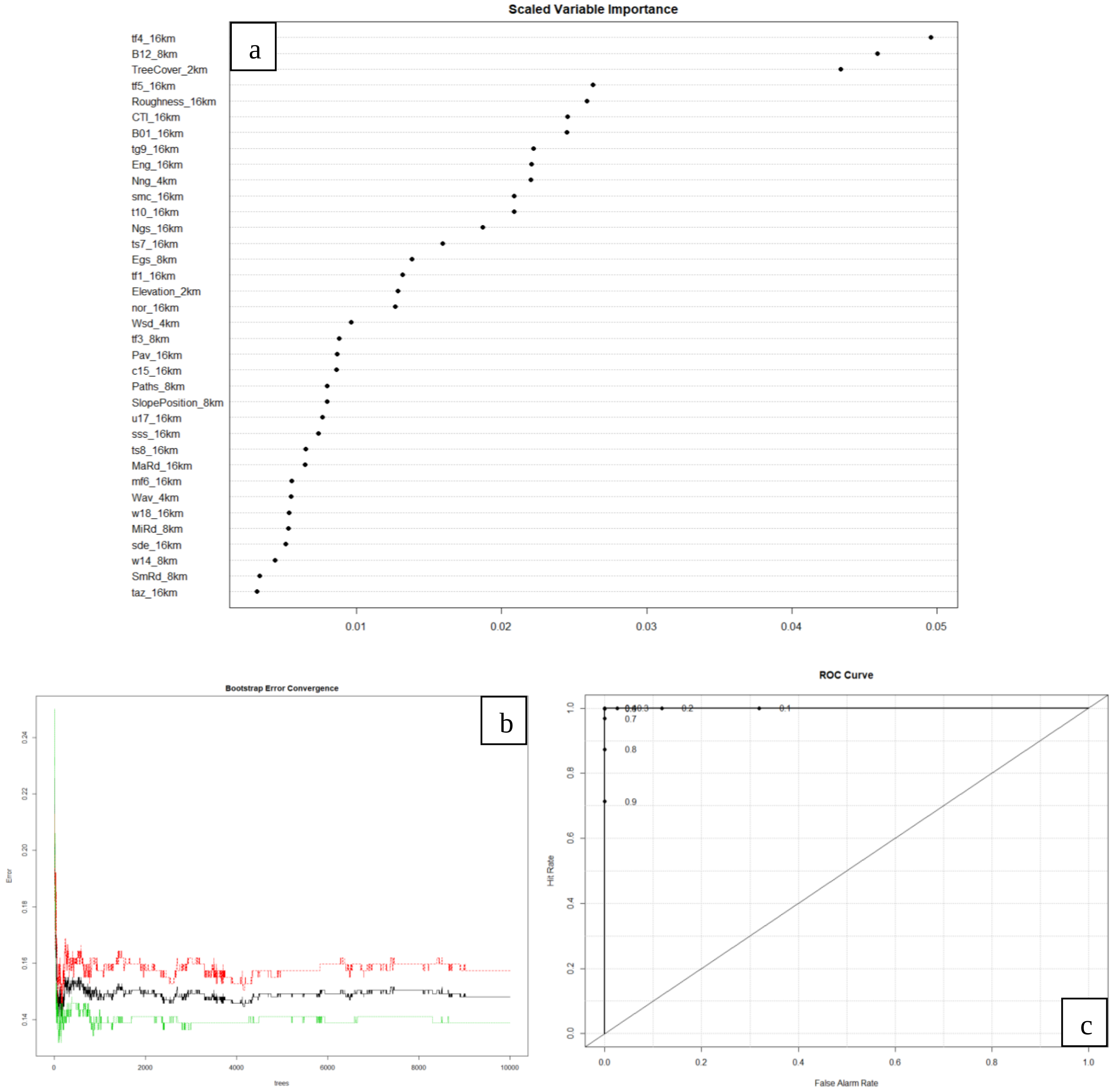

Figure 11 – The scaled variable importance (a), bootstrap error convergence (b), and the ROC curve (c) graphs for the random forest jaguar potential habitat map.

absence proximity matrix (d) is a two-dimensional graph which plots the absences versus the presences used in the RF model.

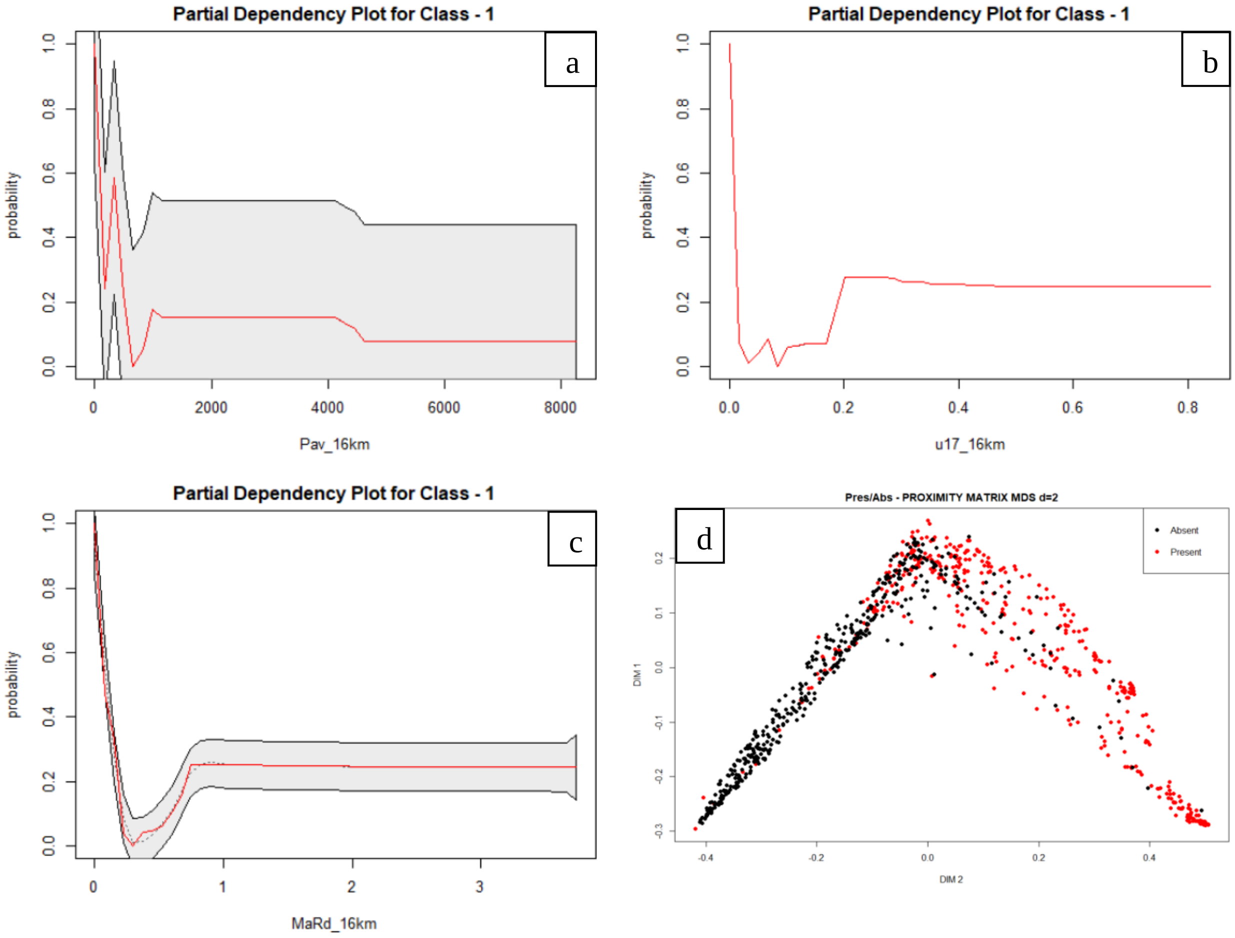

Figure 12 – Partial dependency plots (a) – (c), and (d) Presence/Absence proximity matrix.

4.2. Jaguar Resistance Surface

Figure 13 shows the jaguar resistance surface created using a negative exponential function rescaled from 1 to 100. The color gradient represents areas of high resistance (blue) to low areas (red) for jaguars to traverse the landscape. Areas of low resistance are assumed to comprise landscape features which would be easier or less costly for jaguars to traverse the landscape. Areas of high resistance represent a greater cost to traversal. The less costly areas in this jaguar resistance surface are assumed to be more favorable for use by jaguars.

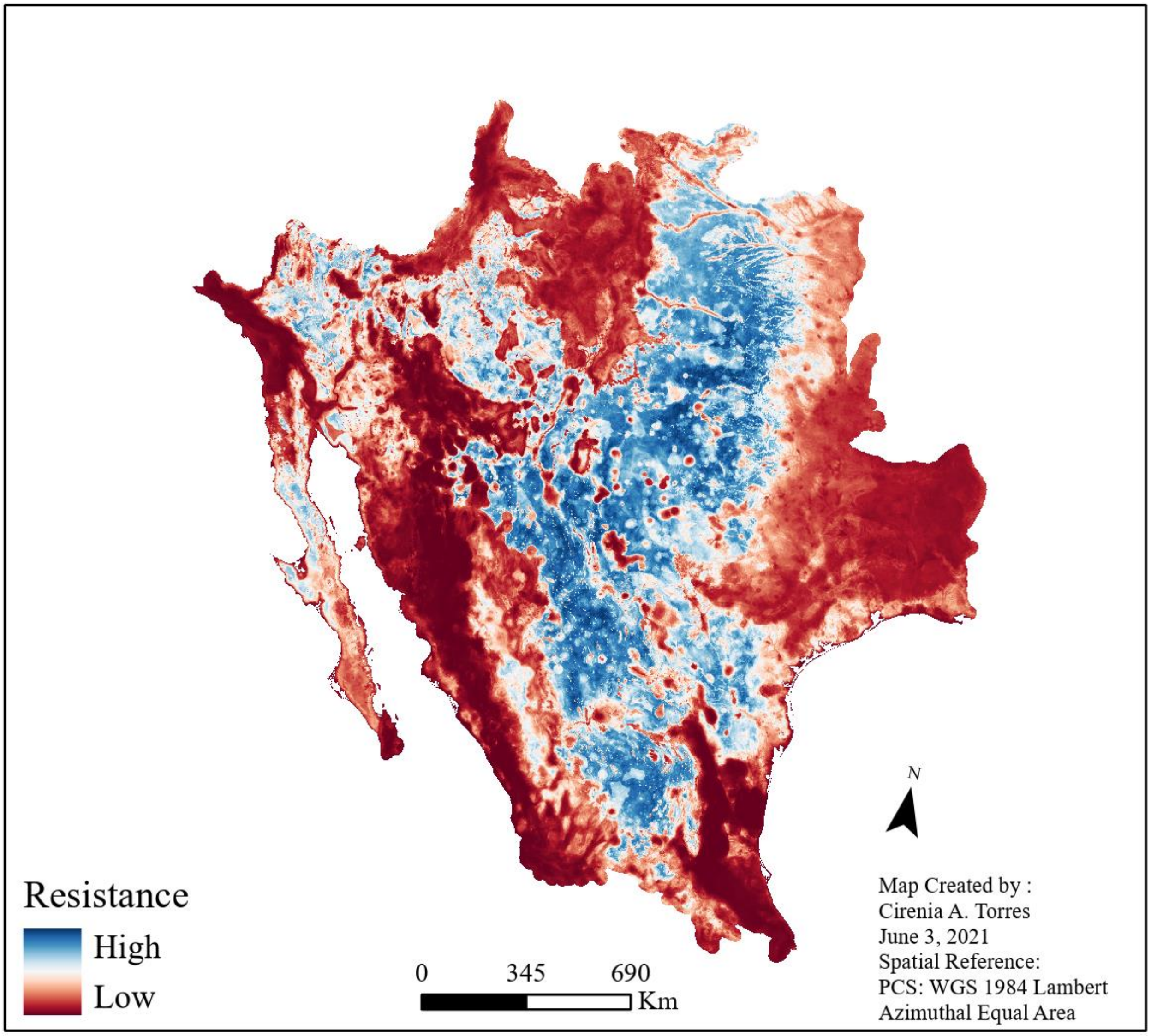

Figure 13 - Jaguar resistance surface created using a negative exponential function and rescaled from 1 to 100.

Figure 14 shows the resistance surface with the major interstate freeways, national and divided roads, and the border wall for reference. From this map we can see that even though there is habitat which is assumed to have a lower resistance for jaguars there may still be complete barriers or partially complete barriers to jaguar movement. For example, a complete barrier could be assumed to be an extremely tall border wall since this would completely block

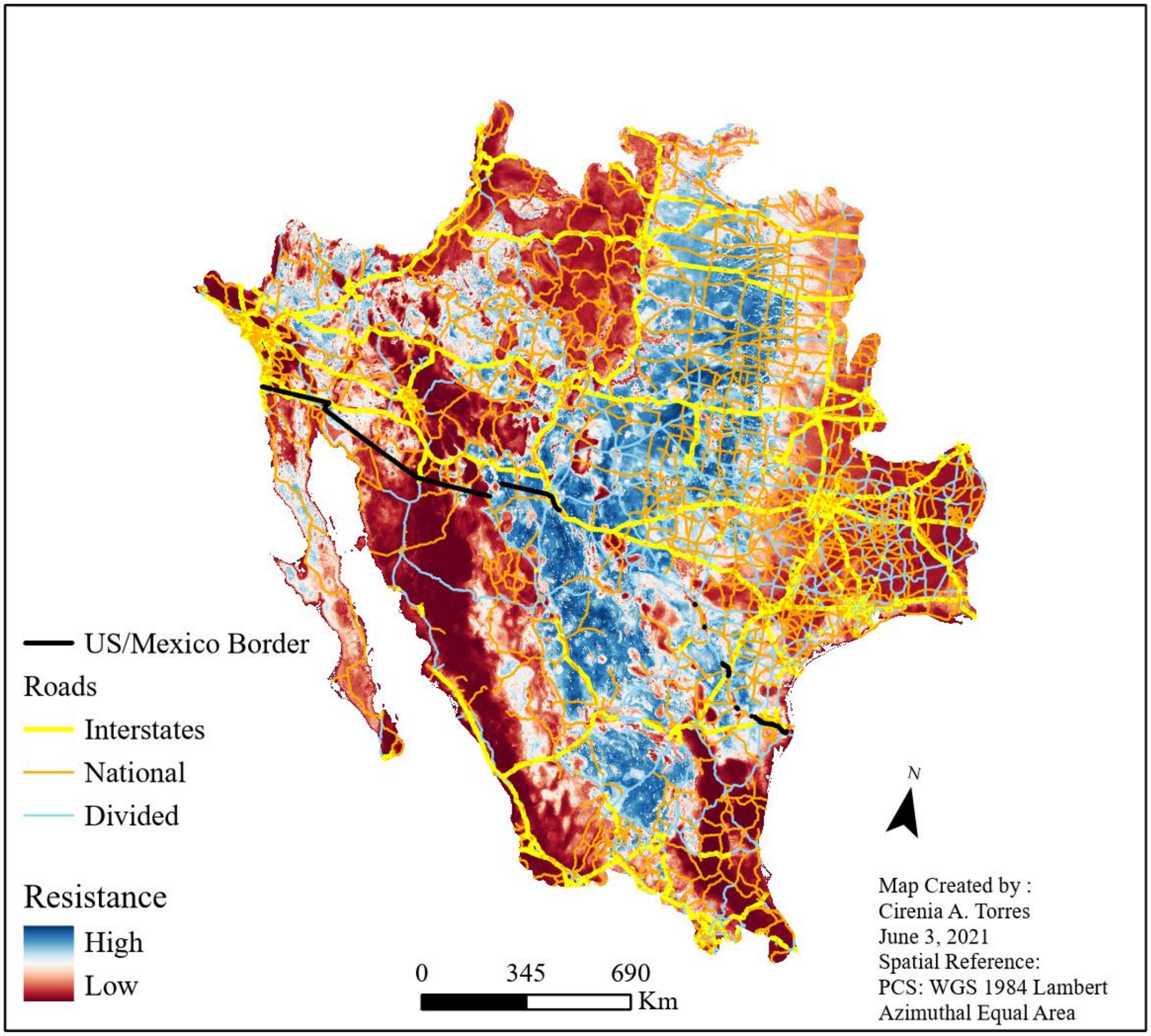

Figure 14 – Map of jaguar resistance surface with major roads and the US-Mexico border wall overlaid.

jaguar movement. Complete to partial barriers can be assumed for different types of roads such as interstate highways, primary national roads, or other important typically divided roads.

4.3. Jaguar Corridor Modeling

The jaguar's potential corridor modeling suggests that there were previously two high density corridors between the US and Mexico allowing jaguar connectivity. However, if the partially constructed border barriers are completed those jaguar corridors will be lost. Additionally, only one jaguar corridor was completely unobstructed, one partially unobstructed, and two already blocked by previous border wall construction. The predicted corridors for the jaguar are shown in Figure 15. Figures 16 and 17 show where the portions of the US-Mexico border wall intersect the jaguar predicted corridors. Figure 16 shows the intersections and unobstructed areas of the previous border wall and the predicted jaguar corridors. Figure 17 shows the new border wall with a status of “partially constructed”, previous border wall, and the intersections with the predicted jaguar corridors. These three maps show the color gradients from high (red) to low (blue) corridor density. These results suggest that if the partially constructed border wall goes to completion, it will block these remaining unobstructed corridors found between the US and Mexico.

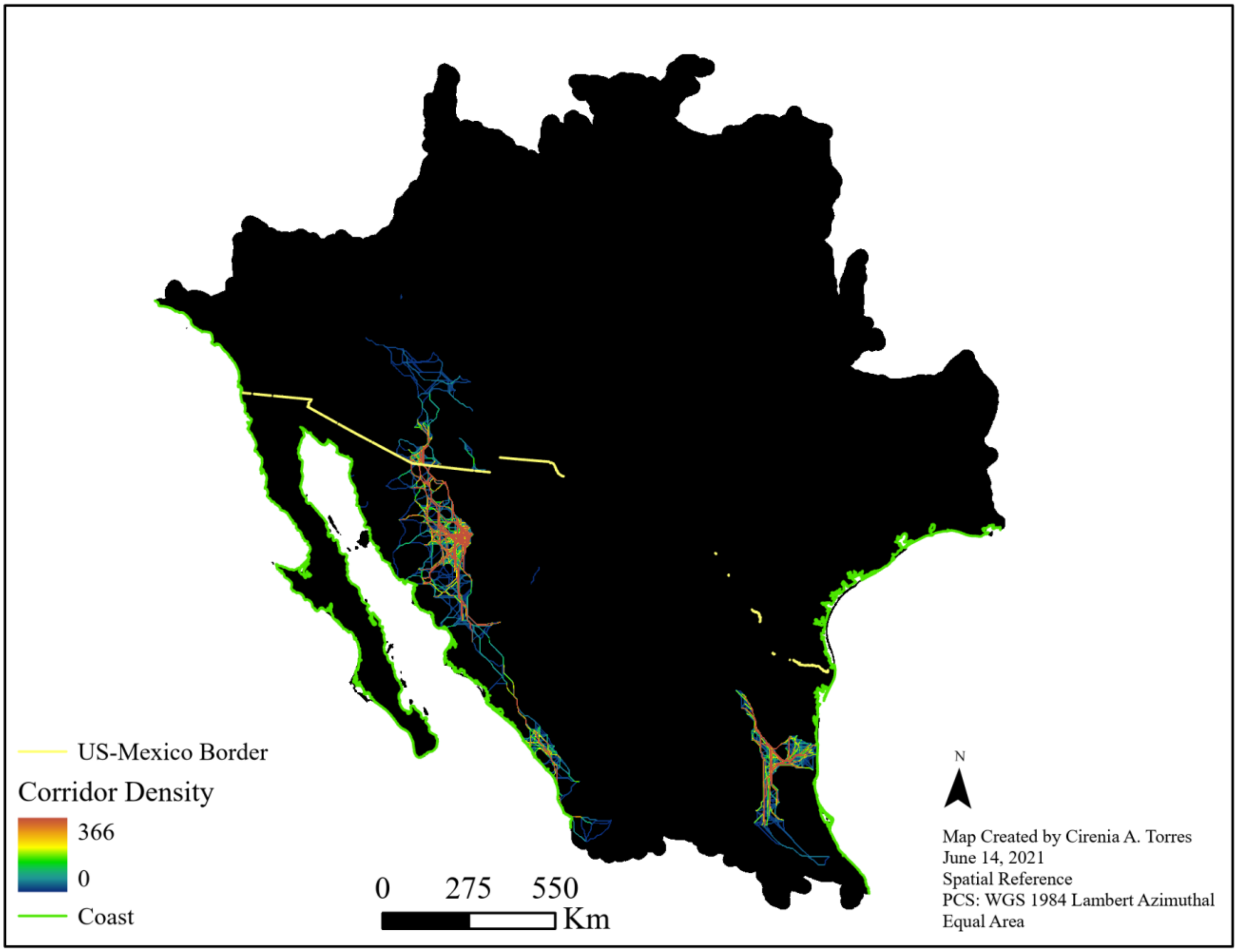

Figure 15 – Map of predicted corridors for the jaguar with the US-Mexico border wall and areas in Figures 16 and 17 overlaid on top of the corridor density map. The US and Mexico coastlines are shown in green.

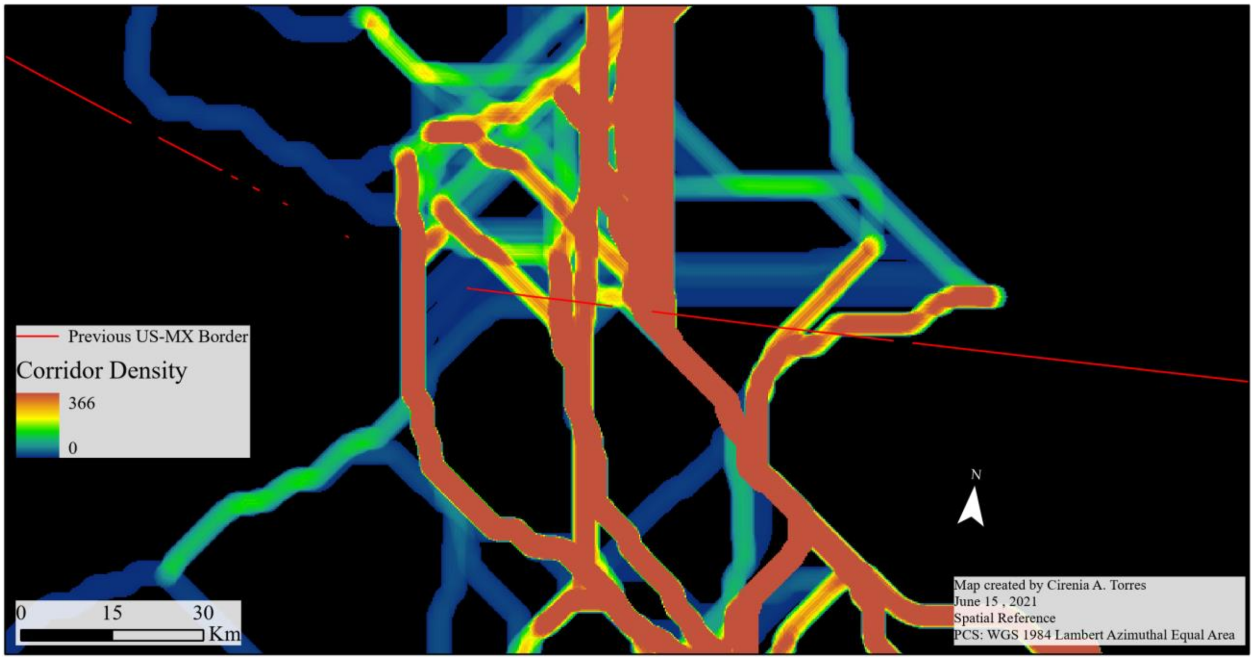

Figure 16 – The predicted corridors for the jaguar with the previous (old) US-Mexico border wall overlaid on top.

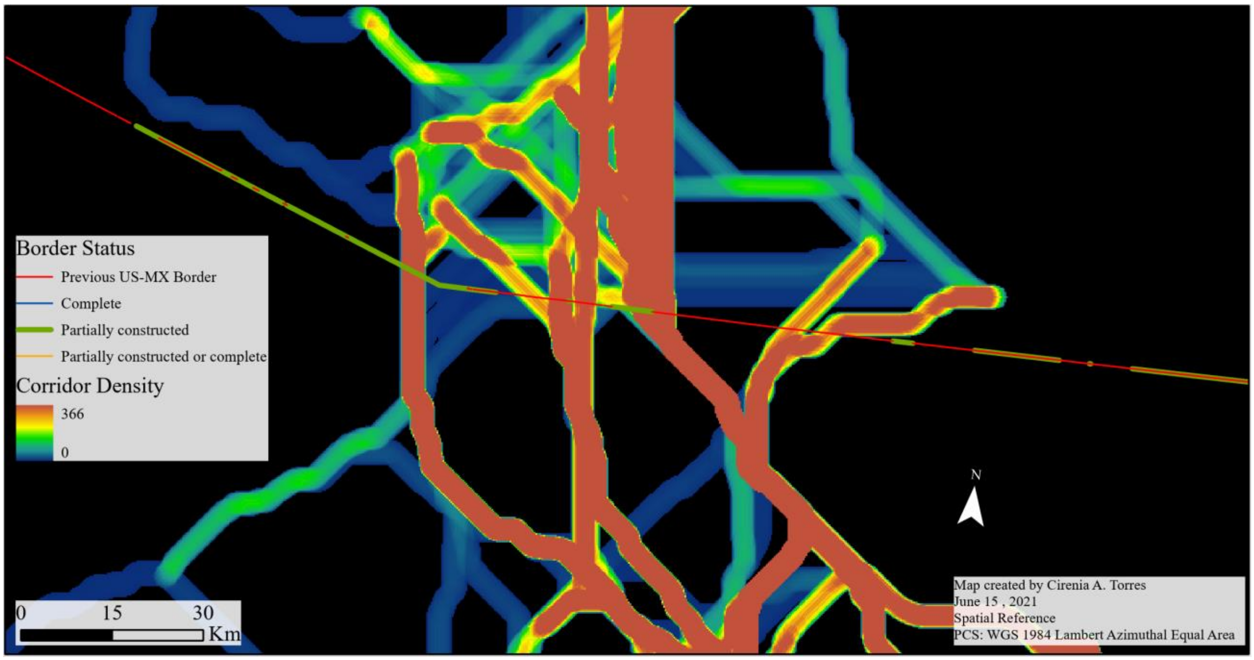

Figure 17 – The predicted corridors for the jaguar with the new US-Mexico border wall overlaid on top.

Chapter 5 : Conclusion

The potential habitat for the jaguar was mapped using the best available data for two countries, which included occurrence records and numerous covariate datasets. Covariate datasets and occurrence records were carefully selected from the best available sources to produce habitat maps, resistance surfaces, and potential corridors. The multiscale optimization and RF model provided the best possible jaguar model using this available data. The FLCP analysis incorporated the robust Dijkstra's (1959) algorithm to predict the best possible potential corridors for the jaguar.

The choice of covariate dataset sources was based on well documented covariates and metadata files were carefully inspected. However, country specific covariate datasets were a major challenge to use because some cannot be combined and others are not of comparable quality. This did limit the desire to include more covariates. However, future studies can certainly add more and compare model results. Open-Street-Map provides free global data and has excellent descriptions of numerous features including roads. However, there may still be differences in data collection when considering the US and Mexico. Sourcing covariate datasets that encompass two or more countries can be a challenging task, but future work can be done by choosing the best available datasets with the best possible descriptions and metadata files as is reasonably possible.

Several species occurrence records were carefully sourced and stored in geodatabases for future modeling. These other species records were found along other sections of the US-Mexico border and will likely produce corridors for one or more of these focal species along other sections of the US-Mexico border. Modeling for several species will provide more information about potential corridors which may be hindered by border walls and/-or roads throughout this study area. The effects of using different species datasets can be compared in future research. For example, if more datasets become available geodatabases can be updated and it would be interesting to see if the outcomes change.

Future research can be conducted to assess the core habitats for the jaguar as well as to include different modeling scenarios for this species. For example, scenarios with different dispersal barriers can be considered and modelled in future studies. In addition, different habitat models for the jaguar can be compared in future studies. For example, it would be interesting to compare a MaxEnt output as well as other model outputs to the results provided in this thesis project. Since covariate datasets were processed and completed to accommodate nine species, they could be used in future studies to model all nine species.

The potential corridor results for the jaguar in Figure 15 show their locations concentrated in parts of the western and eastern portions of this study area. With only one previously completely unobstructed corridor and one partially obstructed corridor available it becomes even more important to advocate for change. Jaguar corridors are found in only a small area relative to the entire US-Mexico border and require specialized habitats.

The Department of Homeland Security (DHS) was established June 2002 under the presidency of George W. Bush as a response to the September 11, 2001, terrorist attacks (https://dhs.gov). The DHS has an important and honorable duty to protect and defend the American people through established protocols that allow efficient information sharing between numerous federal departments and agencies. The U.S. Customs and Border Protection (CBP) is one of the many departments collaborating with the DHS. The CBP has worthy core values with the intention to make the US safer. However, keeping Americans safer could extend beyond identifying potential threats caused by humans to include strategies to enrich and improve the quality of human life by ensuring safe and resilient ecoregions within the US and across borders. There are numerous threatened species in many countries which affect the structure of ecological interactions needed for species long term persistence. Mexico has the third highest total number of threatened species and the US ranks six (IUCN 2021). The DHS and CBP have the potential to afford species in peril greater opportunities to overcome numerous threats by facilitating potentially important northern range expansions through potential corridors for endangered species such as the jaguar which is listed as endangered in both the US and Mexico. By doing so the connectivity of peripheral jaguar populations may improve and protect against future environmental conditions that may put them at a greater extinction risk. It is important to protect core areas, peripheral areas, and cross border connectivity to reduce a species probability of extinction, in this case the jaguar. The DHS and CBP could re-evaluate border barriers, including recently built ones because they can cut off corridors for jaguars and potentially other species. For endangered species it is even more critical to allow dispersal corridors given continuously shrinking habitats.

References

Blaszczynski, J.S. 1997 “Landform characterization with Geographic Information Systems.” *Photogrammetric Engineering and Remote Sensing, 63*(2), 183-191.

Boydston, E. E., and C. A. López González. 2005. “Sexual Differentiation in the Distribution Potential of Northern Jaguars (*Panthera onca*).” In G. J. Gottfried, B. S. Gebow, L. G. Eskew, and C. B. Edminster (Eds.), *Connecting mountain islands and desert seas: biodiversity and management of the Madrean Archipelago II*. (pp. 51-56). Proc. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Breiman, L. 2001. “Random Forests.” *Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324.

Breiman, L., A. Cutler, A. Liaw, and M. Wiener. 2018. “Breiman and Cutler's Random Forests for Classification and Regression.” Retrieved from https://cran.r-project.org/web/packages/randomForest/randomForest.pdf.

Berry, J. K. 2002. “Use surface area for realistic calculations.” *GeoWorld 15*(9), 20-21.

Brown D. E. 1983. “On the Status of the Jaguar in the Southwest.” *The Southwestern Naturalist, 28*(4), 459-460. doi:10.2307/3670828.

Brown, D. E., and C. A. López-González. 2000. “Notes on the Occurrences of Jaguars in Arizona and New Mexico.” *The Southwestern Naturalist, 45*(4), 537-542. https://www.jstor.org/stable/3672607.

Brown, D. E., and C. A. López-González. 2001. “*Borderland Jaguars: Tigres de la Frontera*.” Salt Lake City, UT: The University of Utah Press.

Cassaigne, I., R. A. Medellín, R. W. Thompson, M. Culver, A. Ochoa, K. Vargas, … A. Torres-Gómez. 2016. “Diet of pumas (*Puma concolor*) in Sonora, Mexico, as determined by GPS kill sites and molecular identified scat, with comments on jaguar (*Panthera onca*) diet.” *The Southwestern Naturalist, 61*(2) 125-132. https://doi.org/10.1894/0038-4909-61.2.125.

Cohen, J. 1960. “A Coefficient of Agreement for Nominal Scales.” *Educational and Psychological Measurement, 20*, 37-46. http://dx.doi.org/10.1177/001316446002000104.

Compton, B. W., K. McGarigal, S. A. Cushman, and L. R. Gamble. 2007. “A Resistant-Kernel Model of Connectivity for Amphibians That Breed in Vernal Pools.” *Conservation Biology, 21*(3), 788-799. https://doi.org/10.1111/j.1523-1739.2007.00674.x.

Cushman, S. A., B. McRae, F. Adriaensen, P. Beier, M. Shirley, and K. Zeller. 2013. “Biological corridors and connectivity.” *Key Topics in Conservation Biology, 2*, 384-404. https://doi.org/10.1002/9781118520178.ch21.

Cushman, S. A., J. S. Lewis, and E. L. Landguth. 2014. "Why Did the Bear Cross the Road? Comparing the Performance of Multiple Resistance Surfaces and Connectivity Modeling Methods." *Diversity,* 6(4), 844-854. https://doi.org/10.3390/d6040844.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., and Lawler, J. J. 2007. Random Forests for classification in ecology." *Ecology*, *88*(11), 2783–2792. https://doi.org/10.1890/07-0539.1.

Defenders of Wildlife. 2018. *"In the Shadow of the Wall: Executive Summary."* Retrieved from https://defenders.org/newsroom/shadow-of-wall.

DeMatteo, K. E., M. A. Rinas, J. P. Zurano, N. Selleski, R. G. Schneider, and C. F. Argüelles. 2017. "Using niche-modelling and species-specific cost analyses to determine a multispecies corridor in a fragmented landscape." *PLoS ONE, 12*(8), e0183648. https://doi.org/10.1371/journal.pone.0183648.

Dinerstein, E., D. M. Olson, D. J. Graham, A. L. Webster, S. A. Primm, M. P. Bookbinder, and G. Ledec. 1995. "*A Conservation Assessment of the Terrestrial Ecoregions of Latin America and the Caribbean*." Washington (DC): World Bank.

Dinerstein, E., D. Olson, A. Joshi, C. Vynne, N. D. Burgess, E. Wikramanayake, … M. Saleem. 2017. "An ecoregion-based approach to protecting half the terrestrial realm." *Bioscience, 67*, 534-545.

Dijkstra, E. W. 1959. "A Note on Two Problems in Connexion With Graphs." *Numerical Mathematics, 1*(1), 269-271. https://doi.org/10.1007/BF01386390.

Eizirik, E., J. H. Kim, M. Menotti-Raymond, P. G. Crawshaw, S. J. O' Brien, and W. E. Johnson. 2001. "Phylogeography, Population History and Conservation Genetics of jaguars (*Panthera onca*, Mammalia, Felidae)." *Molecular Ecology, 10*(1), 65-79. https://doi.org/10.1046/j.1365-294X.2001.01144.x.

Elith J., C. H. Graham, P. R. Anderson, M. Dudik, S. Ferrier, A. Guisan, … N. E. Zimmermann 2006. "Novel methods improve prediction of species' distributions from occurrence data." *Ecography (Copenhagen), 29*(2), 129-151. doi:10.1111/j.2006.0906-7590.04596.x.

Elliot, N. B., S. A. Cushman, D. W. Macdonald, and A. J. Loveridge. 2014. "The devil is in the dispersers: predictions of landscape connectivity change with demography." *Journal of Applied Ecology*, *51*(5), 1169–1178. https://doi.org/10.1111/1365-2664.12282.

Ernest, H. B., T. W. Vickers, S. A. Morrison, M. R. Buchalski, and W. M. Boyce. 2014. "Fractured Genetic Connectivity Threatens a Southern California Puma (*Puma Concolor*) Population." *PLoS ONE,* 9(10), e107985. https://doi.org/10.1371/journal.pone.0107985.

Ersoy, E., A. Jorgensen, and P. H. Warren. 2019. "Identifying multispecies connectivity corridors and the spatial pattern of the landscape." *Urban Forestry & Urban Greening, 40,* 308–322. https://doi.org/10.1016/j.ufug.2018.08.001

Evans, J. S., and S. A. Cushman. 2009. “Gradient modeling of conifer species using random forests.” *Landscape Ecology, 24*(5), 673-683. https://doi.org/10.1007/s10980-009-9341-0.

Evans J. S., J. Oakleaf, S. A. Cushman, and D. Theobald. 2014. “*An ArcGIS Toolbox for Surface Gradient and Geomorphometric Modeling (Version 2.0-0).”* Retrieved from http://evansmurphy.wix.com/evansspatial.

Fahrig, L. 2003. “Effects of Habitat Fragmentation on Biodiversity.” *Annual Review of Ecology, Evolution, and Systematics, 34*(1), 487–515. https://doi.org/10.1146/annurev.ecolsys.34.011802.132419.

Freund, Y., and R. E. Schapire. 1997. “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting.” *Journal of Computer and System Sciences, 55*(1), 119–139. https://doi.org/10.1006/jcss.1997.1504.

Friedman, J. H. 2001. “Greedy Function Approximation: A Gradient Boosting Machine.” *The Annals of Statistics, 29*(5), 1189-1232. https://doi.org/10.1214/aos/1013203451.

Friedman, J. H. 2002. “Stochastic Gradient Boosting.” *Computational Statistics and Data Analysis, 38*(4), 367-378. https://doi.org./10.1016/S0167-9473(01)00065-2.

Gaskill, M. 2011. *“United States Border Fence Threatens Wildlife.”* Retrieved from https://doi.org/10.1038/news.2011.452.

Groom, M. J., G. K. Meffe, and C. R. Carroll. 2006. “*Principles of Conservation Biology.”* Sunderland, MA: Sinauer Associates.

Haddad, N. M., L. A. Brudvig, J. Clobert, K. F. Davies, A. Gonzalez, R. D. Holt, … J. R. Townshend. 2015. “Habitat fragmentation and its lasting impact on Earth’s ecosystems.” *Science Advances, 1*(2), e1500052. https://doi.org/10.1126/sciadv.1500052.

Hatten, J. R., A. Averill-Murray, and W. E. van Pelt. 2005. “A spatial model of potential jaguar habitat in Arizona.” *Journal of Wildlife Management*, *69*(3): 1024–1033. https://doi.org/10.2193/0022-541X(2005)069[1024:ASMOPJ]2.0.CO;2.

Hijmans, R. J., and J. Elith. 2017. *“Species Distribution Modeling with R. (Packages. January 8, 2017).”* Retrieved from https://cran.r-project.org/web/packages/dismo/vignettes/sdm.pdf.

Hosch, W. L., ed. 2010. *“The Britannica Guide to Geometry.”* Chicago, IL: Britannica Educational Publishing.

IUCN. 2021. *“The IUCN Red List of Threatened Species (Version 2021).”* Retrieved from https://www.iucnredlist.org/statistics.

Kabacoff, R. I. 2017. *“Generalized Linear Models. (Quick-R: Data Types).”* Retrieved from https://www.statmethods.net/advstats/glm.html.

Keitt, T. H., A. Franklin, and D. L. Urban. 1995. *"Landscape analysis and metapopulation structure. (Chapter 3, Recovery Plan for the Mexican Spotted Owl; Volume II, Technical and supporting information)."* Albuquerque, NM: U.S. Department of the Interior Fish and Wildlife Service, Southwestern Region.

Khosravi, R., M.-R. Hemami, and S. A. Cushman. 2018. "Multispecies assessment of core areas and connectivity of desert carnivores in central Iran." *Diversity and Distributions*, *24*(2), 193–207. https://doi.org/10.1111/ddi.12672.

Landguth, E. L., B. K. Hand, J. Glassy, S. A. Cushman, and M. A. Sawaya. 2012. "UNICOR: A Species Connectivity and Corridor Network Simulator." *Ecography, 35*(1), 9–14. https://doi.org/10.1111/j.1600-0587.2011.07149.x.

Landguth, E. L., B. K. Hand, J. M. Glassy, and S. A. Cushman. 2016. "UNICOR User Manual (Version 2.0)." Retrieved from https://github.com/ComputationalEcologyLab.

Larson, S. E. 1997. "Taxonomic re-evaluation of the Jaguar." *Zoo Biology, 16*(2), 107-120.

Lasky, J. R., W. Jetz, and T. H. Keitt. 2011. "Conservation Biogeography of the US-Mexico Border: a Transcontinental Risk Assessment of Barriers to Animal Dispersal." *Diversity and Distributions 17*(4), 673–687. https://doi.org/10.1111/j.1472-4642.2011.00765.x.

McCain, E. B., and J. L. Childs. 2008. "Evidence of Resident jaguars (*Panthera onca*) in the southwestern United States and the Implications for Conservation." *Journal of Mammalogy 89*(1), 1–10. https://doi.org/10.1644/07-MAMM-F-268.1.

McGarigal, K., K. A. Zeller, and S. A. Cushman. 2016. "Multi-scale habitat selection modeling: Introduction to the special issue." *Landscape Ecology, 31*(6), 1157-1160. https://doi.org/10.1007/s10980-016-0388-4.

McGarigal, K., H. Y. Wan, K. A. Zeller, B. C. Timm, and S. A. Cushman. 2016. "Multi-scale Habitat Selection Modeling: A Review and Outlook." *Landscape Ecology, 31*(6), 1161-1175. https://doi.org/10.1007/s10980-016-0374-x.

Menke, K. 2008. *"Locating Potential Cougar (Puma concolor) Corridors in New Mexico Using a Least-Cost Path Corridor GIS Analysis."* Albuquerque, MN: Bird's Eye View.

Mi, C., F. Huettmann, Y. Guo, X. Han, and L. Wen. 2017. "Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ*, 5, e2849. https://doi.org/10.7717/peerj.2849.

Morato, R. G., K. M. Ferraz, R. C. de Paula, and C. B. Campos. 2014. Identification of Priority Conservation Areas and Potential Corridors for Jaguars in the Caatinga Biome, Brazil. *PLoS ONE, 9*(4), e92950. https://doi.org/10.1371/journal.pone.0092950.

Morcatty, T. Q., J. C. Bausch Macedo, K. A. I. Nekaris, Q. Ni, C. C. Durigan, M. S. Svensson, and V. Nijman. 2020. Illegal trade in wild cats and its link to Chinese-led development in Central and South America. *Conservation Biology, 34*(6), 1525–1535. https://doi.org/10.1111/cobi.13498.

Nelder, J. A., and R. W. M. Wedderburn. 1972. "Generalized Linear Models." *Journal of the Royal Statistical Society. Series A (General)*, 135(3), 370-384. https://www.jstor.org/stable/2344614.

O'Sullivan, D., and G. L. W. Perry. 2013. "*Spatial simulation: Exploring pattern and process*." New York, NY: John Wiley & Sons.

Pascual-Hortal, L., and S. Saura. 2006. "Comparison and Development of New Graph-Based Landscape Connectivity Indices: Towards the Priorization of Habitat Patches and Corridors for Conservation." *Landscape Ecology, 21*(7), 959-967. https://doi.org/10.1007/s10980-006-0013-z.

Pascual-Hortal, L., and S. Saura. 2008. "Integrating Landscape Connectivity in Broad-Scale Forest Planning through a New Graph-Based Habitat Availability Methodology: Application to Capercaillie (Tetrao Urogallus) in Catalonia (NE Spain)." *European Journal of Forest Research,* 127(1), 23–31. https://doi.org/10.1007/s10342-006-0165-z.

Peters, R., W. J. Ripple, C. Wolf, M. Moskwik, G. Carreón-Arroyo, G. Ceballos, … J. R. Miller. 2018. "Nature Divided, Scientists United: US–Mexico Border Wall Threatens Biodiversity and Binational Conservation." *BioScience, 68*(10), 740–743. https://doi.org/10.1093/biosci/biy063.

Phillips, S. J., R. P. Anderson, M. Dudík, R. E. Schapire, and M. E. Blair. 2017. "Opening the Black Box: an Open-Source Release of MaxEnt." *Ecography, 40*(7), 887–893. https://doi.org/10.1111/ecog.03049.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. "Maximum Entropy Modeling of Species Geographic Distributions." *Ecological Modeling, 190*(3-4), 231-259. https://doi.org/10.1016/j.ecolmodel.2005.03.026.

Phillips, S. J., and M. Dudík. 2008. "Modeling of Species Distributions with MaxEnt: New Extensions and a Comprehensive Evaluation." *Ecography, 31*(2), 161-175. https://doi.org/10.1111/j.0906-7590.2008.5203.x.

Phillips, S. J., M. Dudík, and R. E. Schapire. [nd] *"MaxEnt software for modeling species niches and distributions (Version 3.4.1)."* Retrieved from http://biodiversityinformatics.amnh.org/open_source/maxent/.

Phillips, S. J., M. Dudík, and R. E. Schapire. 2004. "A Maximum Entropy Approach to Species Distribution Modeling." *In Proceedings of the Twenty-First International Conference on Machine Learning,* Banff, Alberta, Canada (pp. 1-8). New York, NY: ACM. https://doi.org/10.1145/1015330.1015412.

Pimm S. L., C. N. Jenkins, R. Abell, T. M. Brooks, J. L. Gittleman, L. N. Joppa, … J. O. Sexton. 2014. "The biodiversity of species and their rates of extinction, distribution, and protection." *Science*, *344*(6187), 1246752. https://doi.org/10.1126/science.1246752.

Pocock, R. I. 1939. "The Races of Jaguar (*Panthera onca*)." *Novitates Zoologicae*, *41*(4), 406-422.

Proulx, S. R., D. E. L. Promislow, and P. C. Phillips. 2005. "Network Thinking in Ecology and Evolution." *Trends in Ecology and Evolution*, *20*(6), 345-353. https://doi.org/10.1016/j.tree.2005.04.004.

Quigley, H., R. Foster, L. Petracca, E. Payan, R. Salom, and B. Harmsen. 2017. "*Panthera onca (Errata version published in 2018) The IUCN Red List of Threatened Species 2017*)." Retrieved from http://dx.doi.org/10.2305/IUCN.UK.2017-3.RLTS.T15953A50658693.en.

R Core Team. 2018. *"R: A language and environment for statistical computing."* Vienna, Austria: R Foundation for Statistical Computing.

Rabinowitz, A., and K. A. Zeller. 2010. "A range-wide model of landscape connectivity and conservation for the jaguar, *Panthera onca*." *Biological Conservation*, *143*(4), 939–945. https://doi.org/10.1016/j.biocon.2010.01.002.

Ricketts, T. H. 2001. "The Matrix Matters: Effective Isolation in Fragmented Landscapes." *The American Naturalist*, *158*(1), 87-99. https://doi.org/10.2307/3078900.

Ridgeway, G. 2019. *"Generalized Boosted Models: A Guide to the gbm Package."* Retrieved from https://cran.r-project.org/web/packages/gbm/vignettes/gbm.pdf.

Riley, S. P. D., J. P. Pollinger, R. M. Sauvajot, E. C. York, C. Bromley, T. K. Fuller, and R. K. Wayne. 2006. "FAST-TRACK: A Southern California Freeway Is a Physical and Social Barrier to Gene Flow in Carnivores." *Molecular Ecology*, *15*(7), 1733–1741. https://doi.org/10.1111/j.1365-294x.2006.02907.x.

Riley, S. J., S. D. DeGloria and R. Elliot. 1999. "A terrain ruggedness index that quantifies topographic heterogeneity." *Intermountain Journal of Sciences*, *5*, 23-27.

Rodríguez-Soto, C., O. Monroy-Vilchis, L. Maiorano, L. Boitani, J. C. Faller, M. Á. Briones, … A. Falcucci. 2011. Predicting potential distribution of the jaguar (Panthera onca) in Mexico: identification of priority areas for conservation. *Diversity and Distributions*, *17*(2), 350–361. https://doi.org/10.1111/j.1472-4642.2010.00740.x.

Rodríguez-Soto, C., O. Monroy-Vilchis, and M. M. Zarco-González. 2013. Corridors for jaguar (Panthera onca) in Mexico: Conservation strategies. *Journal for Nature Conservation*, *21*(6), 438–443. https://doi.org/10.1016/j.jnc.2013.07.002.

Rosas-Rosas, O., L. Bender, and R. Valdez. 2008. "Jaguar and Puma Predation on Cattle Calves in Northeastern Sonora, Mexico." *Rangeland Ecology and Management*, *61*(5), 554-560.

Rosas-Rosas, O. and R. Valdez. 2010. The Role of Landowners in Jaguar Conservation in Sonora, Mexico. *Conservation Biology*, *24*(2), 366–371. https://doi.org/10.1111/j.1523-1739.2009.01441.x.

Rudnick, D. A., S. J. Ryan, P. Beier, S. A. Cushman, F. Dieffenbach, C. W. Epps, … S. C. Trombulak. 2012. "The Role of Landscape Connectivity in Planning and Implementing Conservation and Restoration Priorities." *Ecological Society of America* Issues in Ecology, *16*, 1-20.

Ruiz-Garcia, M., E. Payán, A. Murillo, and D. Alvarez. 2006. "DNA Microsatellite Characterization of the jaguar *(Panthera onca)* in Colombia." *Genes and Genetic Systems, 81*(2), 115-127. https://doi.org/10.1266/ggs.81.115.

Saura, S., and L. Pascual-Hortal. 2007. "A New Habitat Availability Index to Integrate Connectivity in Landscape Conservation Planning: Comparison with Existing Indices and Application to a Case Study." *Landscape and Urban Planning, 83*(2-3), 91–103. https://doi.org/10.1016/j.landurbplan.2007.03.005.

Saura, S., and J. Torne´. 2009. "Conefor Sensinode 2.2: A Software Package for Quantifying the Importance of Habitat Patches for Landscape Connectivity." *Environmental Modeling and Software, 24*(1), 135-139. https://doi.org/10.1016/j.envsoft.2008.05.005.

SEMARNAT (Secretary of Natural Resources of Mexico). 2002. *"Normas Oficiales Mexicanas (NOM-059-Ecol.) que determina las especies de flora y fauna silvestres terrestres y acuaticas en peligro de extincion, menazadas, raras y las sujetas a proteccion especial, que establece especificaciones para su proteccion."* Mexico City, Mexico: SEMARNAT.

Seymour, K. L. 1989. "*Panthera onca*." *Mammalian Species, 340*, 1-9. https://doi.org/10.2307/3504096.

Tôrres, N. M., P. De Marco, T. Santos, L. Silveira, A. T. de Almeida Jácomo, and J. A. Diniz-Filho. 2012. Can species distribution modelling provide estimates of population densities? A case study with jaguars in the Neotropics. *Diversity and Distributions*, *18*(6), 615–627. https://doi.org/10.1111/j.1472-4642.2012.00892.x.

Timm, B. C., K. McGarigal, S. A. Cushman, and J. L. Ganey. 2016. "Multi-scale Mexican spotted owl (*Strix occidentalis lucida*) nest/roost habitat selection in Arizona and a comparison with single-scale modeling results." *Landscape Ecology, 31*(6), 1209-1225. https://doi.org/10.1007/s10980-016-0371-0.

Upton, G., and I. Cook. 2014. *"A Dictionary of Statistics (3 ed.)."* Oxford, UK: Oxford University Press. https://doi.org/10.1093/acref/9780199679188.001.0001.

Urban, D., and T. Keitt. 2001. "Landscape connectivity: a graph-theoretic perspective." *Ecology, 82*(5), 1205-1218. https://doi.org/10.2307/2679983.

Wan, H. Y., S. A. Cushman, and J. L. Ganey. 2018. “Habitat Fragmentation Reduces Genetic Diversity and Connectivity of the Mexican Spotted Owl: A Simulation Study Using Empirical Resistance Models.” *Genes*, *9*(8), 403. https://doi.org/10.3390/genes9080403.

Wan, H. Y., K. McGarigal, J. L. Ganey, V. Lauret, B. C. Timm, and S. A. Cushman. 2017. “Meta-replication reveals nonstationarity in multi-scale habitat selection of Mexican Spotted Owl.” *The Condor: Ornithological Applications, 119*(4), 641–658. https://doi.org/10.1650/CONDOR-17-32.1.

Wiens, J. A. 1989. “Spatial Scaling in Ecology.” *Functional Ecology, 3*(4), 385-397. https://www.jstor.org/stable/2389612.

Worton, B. J. 1989. Kernel Methods for Estimating the Utilization Distribution in Home-Range Studies. *Ecology*, *70*(1), 164–168. https://doi.org/10.2307/1938423.

Zarco-González, M. M., O. Monroy-Vilchis and J. Alaníz. 2013. Spatial model of livestock predation by jaguar and puma in Mexico: Conservation planning. *Biological Conservation, 159*, 80–87. https://doi.org/10.1016/j.biocon.2012.11.007.

Zeller, K. A., S. Nijhawan, R. Salom-Pérez, S. H. Potosme, and J. E. Hines. 2011. “Integrating occupancy modeling and interview data for corridor identification: A case study for jaguars in Nicaragua.” *Biological Conservation, 144*(2), 892-901. https://doi.org/10.1016/j.biocon.2010.12.003.