Title
The influence of land use and habitat fragmentation on landscape connectivity

Permalink
https://escholarship.org/uc/item/43c5d073

Author
Gray, Morgan

Publication Date
2017

Peer reviewed|Thesis/dissertation
The influence of land use and habitat fragmentation on landscape connectivity

By

Morgan Gray

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Environmental Science, Policy, and Management
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:

Professor Adina M. Merenlender, Chair
Professor David Ackerly
Professor Matthew Potts

Spring 2017
The influence of land use and habitat fragmentation on landscape connectivity

Copyright 2017

by

Morgan Gray
Abstract

The influence of land use and habitat fragmentation on landscape connectivity

by

Morgan Gray

Doctor of Philosophy in Environmental Science, Policy, and Management

University of California, Berkeley

Professor Adina M. Merenlender, Chair

The built environment, especially roads, urban and suburban development, can reduce the ability for wildlife to move across landscapes. Maintaining landscape connectivity has become a central theme in ecology and conservation, as corridors of intact habitat help maintain ecosystem functionality and, in the face of climate change, may provide migration paths for species. However, the influence of the built environment on connectivity is rarely quantified using empirical data informed by species detection, movement, or genetic structure. Rather, structural connectivity, as opposed to functional connectivity, is estimated using land cover alone. Structural connectivity estimates offer a simple and potentially powerful approach with fewer data requirements for wildlife corridor planning; however, models of structural connectivity are rarely if ever evaluated with empirical species data, limiting our understanding of their reliability and utility. This dissertation fills this gap by investigating the influence of human land use and habitat fragmentation on landscape connectivity using a suite of quantitative modeling approaches and mammals as the focal species, including cross comparisons among these approaches. Specifically, three methods that vary in levels of biological information are used to evaluate how well structural connectivity models perform for individual species, as well as their relationship to functional connectivity.

To begin with, the utility of a structural connectivity model based on the distribution and intensity of land use is evaluated by comparing model predictions to observed land use by a generalist carnivore, the puma (Puma concolor). Findings from this study indicate that generic landscape permeability models can be used with confidence as a guide when prioritizing habitat corridors for biodiversity conservation across fragmented landscapes. Next, the utility of structural connectivity models is further evaluated by examining how the inclusion of specific human land use variables affects model accuracy in a species distribution model for gray fox (Urocyon cinereoargenteus). Findings from this study indicate that species distribution models generated in human-dominated landscapes have higher accuracy when informed by indices of land use. Finally, a combination of spatial and genetic methods is used to evaluate the influence of roads on the functional connectivity for a small mammal, California ground squirrels (Otospermophilus beecheyi). Findings from this study indicate that a combined spatial and genetic approach can be used to identify locations where roads act as barriers.

Given the importance of habitat fragmentation, there is a pressing need to rapidly develop and utilize connectivity assessment methods in conservation planning. Research findings
presented here have already impacted mammal conservation planning and management in California in the following specific ways. The structural connectivity model was used to identify priority habitat linkages for inclusion in the Conservation Blueprint for Santa Cruz County, and the combined genetic and spatial approach was used by the Santa Clara County Open Space Authority, CA to identify corridors and restore connectivity in an area with an existing road network. Beyond these regional impacts, future conservation planning worldwide can benefit from using readily available data collected by citizen scientists as input in predictive mapping to increase the sample size and spatial coverage for species distribution modeling.
# Table of Contents

Table of Contents  
Acknowledgements

## Introduction

- References  

### Chapter 1: Landscape feature-based permeability models relate to puma occurrence

- Abstract  
- Introduction  
- Methods
  - Study area  
  - Landscape permeability maps  
  - Puma GPS data  
  - Data analysis  
- Results  
- Discussion  
- Conclusion  
- Acknowledgements  
- References  
- Figures
  - Figure 1  
  - Figure 2  
  - Figure 3  

## Chapter 2: Including land use improves the predictive ability of a species distribution model

- Abstract  
- Introduction  
- Methods
  - Study area  
  - MaxEnt model construction  
  - MaxEnt model analysis  
- Results  
- Discussion  
- Conclusions  
- Acknowledgements  
- References  
- Figures
  - Table 1  
  - Figure 1  
  - Figure 2  
  - Table 2
Chapter 3: Human land use decreases genetic connectivity in a small mammal

Abstract
Introduction
Methods

Study area
Landscape permeability
Ground squirrel samples
Genetic diversity
Genetic structure

Results

Landscape permeability
Genetic diversity

Discussion
Conclusions
Acknowledgements
References

Figures

Figure 1
Figure 2
Figure 3
Figure 4
Table 1
Table 2

Conclusion
Multi-taxon approach to habitat connectivity
Evaluating and correcting for bias in presence only data
Roads as barriers to connectivity
References
Acknowledgements

I begin by thanking my advisor, Adina Merenlender, who is the type of mentor, colleague, and friend that any young researcher would be privileged to work with. I thank the faculty at the University of California, Berkeley, who have shared countless insights on many topics in ecology and conservation biology, especially my committee members Matthew Potts and David Ackerly.

I gratefully acknowledge financial support from the Department of Environmental Science, Policy, and Management as well as the National Science Foundation Graduate Research Fellowship Program (Grant No. DGE 1106400), the Santa Clara County Open Space Authority, Guadalupe-Coyote Resource Conservation District.

I thank my collaborators who contributed to the research presented here, especially Chris Wilmers, Shane Feirer, Sarah Reed, Allison Bidlack, Tanya Diamond, Ahiga Snyder, Dave Johnston. This research would not have been possible without the Land Trust of Santa Cruz County, the University of California Hopland Research and Extension Center, and iNaturalist.

For their friendship and advice over the years, I give special thanks to my colleagues Mary Matella, Justin Kitzes, Dylan Chapple, Andres Munoz, Paul Elsen, Annika Keeley, Chelsea Andreozzi, and David Hines of the Merenlender Lab, and Rachel Neurath, Alex Krohn, Kelly Iknayan, and Michelle Reilly. The friendship of Andro Hsu has been invaluable, and I thank him for believing in me. For his fount of love, warmth, and acceptance, I thank my dear Thomas Held. Last, I thank Winslow, Gwendel, Quentin, Joon, and Arlis for their love and support.
Introduction

Landscapes are increasingly urban (US: Alig et al. 2004; Europe: Antrop 2004), and becoming more so globally (McDonald 2008). More than half of the world’s inhabitants currently live in cities, and global urbanization forecasts predict the urban population to reach 70% by 2050 (United Nations 2012). Urbanization may have numerous effects on natural habitat and the wildlife it supports. Overt impacts include fragmentation, parcelization (Gobster & Rickenbach 2004), and exurban sprawl (Theobald 2005). Less obvious impacts include the increased size of the wildland-urban interface, and the corresponding threats of wildfires and invasive species (Radeloff et al. 2005; Theobald & Romme 2007).

A fragmented landscape is characterized by patches of natural habitat surrounded by a matrix of human-modified land cover (Mcintyre & Hobbs 1999). When a landscape is fragmented for agricultural, industrial, and/or residential uses, habitat connectivity decreases as a result of the altered landscape composition and development of road infrastructure. The built environment, especially roads, urban and suburban development, can reduce the ability for wildlife to move across landscapes (Fu et al. 2010; Tannier et al. 2012). As humanity’s footprint increases worldwide, so does habitat fragmentation (Nilsson et al. 2005; Ribeiro et al. 2009; Butchart et al. 2010). Given the variety of impacts urbanization imparts on landscapes, it is not surprising that human-induced habitat fragmentation is one of the primary threats to global biodiversity (Tilman et al. 2001; IUCN 2013). However, how populations fare in urbanizing landscapes is only partially understood (Collinge 2009; Collins & Kays 2011).

One way to lessen the harmful effects of urbanization on wildlife is to conserve corridors of land that link natural habitat patches across the landscape, thereby allowing species to move and hopefully persist around human development (Kindlmann & Burel 2008). Connectivity is a measure of the extent to which animals can move among patches of suitable habitat (Tischendorf & Fahrig 2000). It can be facilitated locally by corridors and at a larger scale by linkages. A corridor is a space that allows an organism to move between isolated patches of natural habitat over time (Lidicker 1999). A linkage refers to a broader extent of connectivity than that encompassed by a corridor, and thus can facilitate the movement of multiple species and flows of matter and energy across ecosystems ecological processes (Hilty et al. 2006).

Maintaining landscape connectivity has become a central theme in ecology and conservation (Rands et al. 2010; Fleishman et al. 2011; Sutherland et al. 2013). Landscape connectivity facilitates the wildlife movement by linking separated patches of suitable habitat (Hilty et al. 2006). These corridors of intact habitat help maintain current ecosystem functionality (Tewksbury et al. 2002; Damschen et al. 2006). Additionally, the resulting network of connected natural habitat may offer future ecological benefits by providing migration pathways to species in the face of climate change (Hannah 2011; Nuñez et al. 2013). To best conserve biodiversity we must identify and preserve core habitat patches that support the persistence of species assemblages and ecosystems. Further, connectivity among such patches must be maintained through a combination of habitat linkages and a permeable matrix (Noss 2001; Crooks et al. 2011).

Habitat connectivity is usually conceptualized as structural or functional. Structural connectivity is a metric of the contiguity of habitat patches in physical space. Estimates of
structural connectivity are derived from landscape attributes such as the shape, size, and configuration of habitat patches, and do not account for actual land use by wildlife (Calabrese & Fagan 2004; Theobald et al. 2012). Functional connectivity is a measure of the ability of organisms and their genes to move among patches of suitable habitat in a fragmented landscape (Taylor et al. 1993; Fahrig 2003; Hilty et al. 2006). Ideally, measures of functional connectivity are derived using empirical data about landscape composition, habitat use, and movement by wildlife.

Connectivity estimates differ in data requirements and informational yield across a spectrum that ranges from strictly structural connectivity at one extreme to biologically-informed functional connectivity at the other (Rayfield et al. 2011; Rudnick et al. 2012). Structural connectivity estimates require the least amount of input data (Calabrese and Fagan 2004), but are sensitive to input parameters like patch size and shape (Bender et al. 2003). Precise movement data for functional connectivity estimates are uncommon at the landscape level because they are costly to collect. Instead, species distribution models – based on landscape features, expert opinion, or species occurrence data – are often used to predict biodiversity distributions and aid in identifying suitable pathways for wildlife movement.

Protection of both structural and functional connectivity have been shown to be crucial for biodiversity conservation (Bennet 1999). Structurally isolated populations may have reduced access to resources like food or mates, and face barriers to dispersal when separated by inhospitable habitat (Moilanen & Hanski 1998). Functionally isolated wildlife populations can exhibit decreased genetic diversity (Epps et al. 2005), which could result in inbreeding depression (Roelke et al. 1993; Hedrick & Kalinowski 2000; Keller & Waller 2002), increased disease susceptibility (Spielman et al. 2004; Charlesworth et al. 2009), and greater risk of extinction, especially for species residing in small or isolated habitat patches (Fahrig & Merriam 1994).

At the heart of connectivity analyses are models derived using habitat resistance or landscape permeability. Beyond corridors and linkages, the matrix has a significant influence on connectivity among habitat fragments (Ricketts 2001). A number of models based on matrix connectivity have been developed that derive connectivity estimates from mathematical equations based on empirical studies of species’ abundance or occurrence among different land cover types, or on expert opinion of species’ habitat associations. Habitat resistance (Ray et al. 2002, Joly et al. 2003) and circuit theory (McRae et al. 2008) models are based in electric circuit theory, and create spatial maps that assign a value to each landscape or environmental feature that represents the degree to which that feature impedes or facilitates connectivity between two or more patches for an organism of interest. Because these models simultaneously consider all possible pathways that may connect patch pairs (McRae & Beier 2007b; McRae et al. 2008; Spear et al. 2010; Lamy et al. 2013), they can be thought of as hypothesized relationships between landscape variables and movement (O’Brien et al. 2006) or gene flow (Wang et al. 2008). Permeability models are an extension of the resistance concept (Ray et al. 2002); model output often is in the form of a grid-based map with a value assigned to each cell that represents its permeability to an organism’s movement (Merenlender 2011, Theobald et al. 2012).

Many corridor conservation plans are based on structural connectivity models, and are rarely tested with empirical field data. For example, The Land Trust of Santa Cruz County is a conservation organization in California, USA that is actively trying to conserve open space and
biodiversity (Press et al. 2002). Conservation and land use planners in the region face challenges commonly encountered in areas with sprawling development, including how to maintain wildlife movement across an increasingly developed landscape. To do so, a series of permeability models were developed for a linkage analysis by the Land Trust of Santa Cruz County (Merenlender and Feirer 2011) that were derived from an estimated linear relationship between specific landscape features related to the built environment and species detections from empirical studies (Forman 2000, Reed 2007, Merenlender et al. 2009).

Another gap related to connectivity models for conservation planning is the need to identify habitat characteristics required for species persistence, as well as the factors that limit species distributions. Species distribution models based on landscape features, expert opinion, or species occurrence data, are widely used to predict biodiversity distributions and aid the decision-making process for diverse conservation applications including connectivity planning (Rocchini et al. 2011). Despite the influence of the built environment on wildlife populations (Wilmers et al. 2013; Gray et al. 2016), species distribution models rarely incorporate human land use variables (Beninde et al. 2015) – a practice that may lead to inaccurate model estimates.

Structural and functional connectivity may be jointly evaluated, as exemplified by the interdisciplinary field of landscape genetics. The aim of landscape genetics research is to combine genetic, computational, and spatial analytic tools (Epps et al. 2007) to understand how landscape features affect spatial genetic structure (Manel 2003, Storfer 2007) which can be used as a surrogate measure of functional connectivity (Boulet et al. 2007). Specifically, landscape genetic approaches allow the quantitative estimation of effective separation that exists due to barriers in gene flow between individuals or populations (Vignieri 2005), and often reflects the geographic distance between populations as well as barriers to gene flow – whether natural or human made. While advances in spatial analysis and increasingly available molecular approaches have made it more accessible to investigate the impact of human land use, and specifically roads, on connectivity, little work has been done to quantify the impact of road density and the relative importance of road characteristics on connectivity (Balkenhol & Waits 2009) and evidence of fine scale genetic effects associated with fragmentation has been rare.

In my dissertation, I tackle some of these gaps in connectivity assessment methods for conservation planning. Specifically, in Chapter 1, I compare landscape permeability estimates, derived from simplified, biologically-informed connectivity models, with existing occurrence data for pumas (*Puma concolor*) across a gradient of land use. The results show that pumas were observed to readily use moderately disturbed habitats, and rarely were detected in the most heavily disturbed areas. This comparison of a more generic connectivity model estimate with animal field observations shows that generic models can be useful for corridor designs in highly disturbed environments. In Chapter 2, I evaluate the effect of including human land use variables in a species distribution model by comparing the predictive performance and distribution estimates of a MaxEnt model for gray fox (*Urocyon cinereoargenteus*) across a land use gradient in the San Francisco Bay Area, CA. The results show that including human land use variables improves model accuracy and decreases overprediction, indicating that species distribution models generated in human-dominated landscapes benefit when informed by land use. In Chapter 3, I combine molecular and spatial analyses to assess whether roads are a barrier to functional connectivity for ground squirrels (*Otospermophilus beecheyi*) separated by a major freeway in Coyote Valley, CA. The genetic similarity among populations on the same side of
roads was higher than among populations separated by major roads, in particular, a large highway. These results indicate that habitat fragmentation by roads is a barrier to gene flow in ground squirrels, resulting in quantitative differences in genetic structure and diversity between populations.

My research findings have implications that may be applied beyond the study sites in California. These findings may be used to prioritize habitat corridors for biodiversity conservation across fragmented landscapes by conservation organizations working on habitat connectivity enhancement on the ground. Land use and conservation planners often need connectivity assessment methods that can be rapidly developed and adapted into local and regional planning (Huber et al. 2012). However, the need for ecological information is greater than the resources available for this type of research, making the development of modeling techniques and tools important in estimating the impacts of habitat loss and fragmentation on wildlife (Fielding & Bell 1998). Thus, the methods employed here may be used to quickly and affordably identify regions that enhance connectivity and allocate limited resources toward land management or acquisition.

Identifying the social-ecological processes and land use patterns – such as patch size and distance to roads – that impact a variety of wildlife in a similar way can provide insight to managers and land use planners when prioritizing corridors in highly disturbed environments. Community-level habitat quality assessments can meet the need for generalized permeability maps of places lacking species data and modeling capacity. In evaluating how landscape permeability maps compared with existing occurrence data for a focal species of conservation concern in the area, Chapter 1 shows that biologically-informed, structural permeability models do reflect puma habitat use – and thus functional connectivity – on the ground to some extent. This evaluation supports the use of model that maps the level of landscape permeability that surrounds the built environment, as measured by distance to roads and housing density, as a spatially explicit way to identify areas important wildlife movement for conservation planning efforts. For example, this approach was used to identify potential habitat linkages for inclusion in the Conservation Blueprint for Santa Cruz County.

The impact of human land use – namely human population density, road density, and parcel size – can also inform species-specific distribution models, which are particularly needed in regions with rapid human population growth. Using a straightforward species modeling program, MaxEnt, and raster data for land use variables – both of which are freely available online – conservation professionals can develop species distribution models that accurately translate the impact of land use on wildlife into a quantifiable metric. Such species distribution models can be included when generating regional build-out plans for conservation-oriented development. Proposed build-out forecasts can then be compared with species distribution models to explore the overlap between future development plans and critical wildlife habitat.

Using a combination of spatial and genetic analyses provides unique insight into the impacts of human land use – specifically roads – on wildlife and can be used to inform conservation planning. The quantified estimates of the extent to which primary and secondary roads, road density, and low landscape permeability may act as barriers to gene flow from Chapter 3 are part of a conservation planning effort underway by public and private land trusts, California Department of Transportation, and California Department of Fish and Wildlife to assess connectivity across Coyote Valley for species found in the western Santa Cruz Mountains.
and the eastern Diablo Range. The findings of this research will be used to inform conservation planning across the valley are also useful for other regions with similar road networks.

Given the importance of habitat fragmentation, there is a pressing need to rapidly develop and adapt connectivity assessment methods for conservation planning. The methods presented here may provide conservation practitioners with tools to quantify connectivity for planning and management. For example, generic landscape permeability models can be used as a guide when prioritizing habitat corridors for biodiversity conservation across fragmented landscapes. When predictive mapping for conservation planning is needed in urban or exurban contexts, including human land use variables is important to reflect the distributions of many species. Finally, when efforts are made to restore connectivity through the addition of road crossing structures, a combined spatial and genetic approach can identify locations where roads are acting as barriers.
References


Chapter 1: Landscape feature-based permeability models relate to puma occurrence

Published in Landscape and Urban Planning 147 (2016): 50-58
Gray, M, CC Wilmers, SE Reed, and AM Merenlender.

Abstract

Habitat fragmentation in human-dominated landscapes is seen as a major threat to biodiversity persistence. Nearly all corridor conservation plans designed to restore habitat connectivity are based on modeled data, and are rarely tested with empirical field data. Here we describe landscape permeability models derived from an estimated linear relationship between specific landscape features related to human land use (e.g. traffic volume, housing density) and bird as well as mesocarnivore detection levels from empirical field studies. We compare these model estimates with existing occurrence data for pumas (*Puma concolor*), a generalist predator commonly used as a focal species for connectivity analysis, in the Santa Cruz Mountains. Our results show that pumas were observed to readily use moderately disturbed habitats, and rarely were detected in the most heavily disturbed areas. This comparison of a more generic connectivity model estimate with animal field observations shows that while generic models can be useful for corridor designs in highly disturbed environments they may be less useful in moderately impacted rural to semi-natural landscapes, where more detailed studies of species behavior may be required to delineate functional corridors. Mapping the level of landscape permeability that surrounds the built environment, as measured by distance to roads and housing density, offers a spatially explicit way to identify areas important wildlife movement. This approach provides a tool to help managers and land-use planners prioritize habitat corridors for biodiversity conservation across fragmented landscapes.

Introduction

One of the primary threats to biodiversity is human-induced habitat fragmentation (Tilman et al. 2001, IUCN 2004), which is on the rise worldwide (Nilsson et al. 2005, Ribeiro et al. 2009, Butchart et al. 2010). A fragmented landscape is characterized by patches of natural habitat surrounded by a matrix of human-modified land cover (McIntyre and Hobbs 1999). Protection of habitat connectivity is crucial for biodiversity conservation to facilitate movement through the matrix (Bennet 1999), especially for wide-ranging mammalian carnivores (Crooks et al. 2011, Hilty et al. 2006). Specifically, to conserve biodiversity we must identify and preserve core habitat patches supporting the persistence of species assemblages and ecosystems, and ensure connectivity among such patches with habitat linkages and/or a permeable matrix (Crooks et al. 2011, Noss et al. 2001).

Increasingly, habitat corridors are being planned and established to mitigate habitat fragmentation (Hilty et al. 2006) at multiple scales. For example, large-scale projects focusing on entire ecosystems are underway to connect forest communities from southern México into Panamá (Kaiser 2001) and linking the Yellowstone area in Wyoming north to Alaska (Walker
and Craighead 1997). Similarly, local-scale projects to protect wildlife movement are happening worldwide (Klar et al. 2012, Underwood et al. 2012). Connectivity endeavors are often custom projects that depend upon species- and landscape-specific information (La Rue and Nielsen 2008), a practice that is expensive and time-consuming. Yet, land use and conservation planners often need connectivity assessment methods that can be rapidly developed and adapted into local and regional planning (Huber et al. 2012).

Connectivity metrics for biodiversity conservation differ in data requirements and informational yield. For example, structural connectivity is derived from landscape attributes such as the shape, size, and configuration of habitat patches, but does not account for animal dispersal ability. Structural connectivity estimates require less input data and generate relatively crude estimates of connectivity (Calabrese and Fagan 2004). Similarly, simple estimates of naturalness levels have been used to coarsely model landscape permeability across the entire United States (Theobald et al. 2012). On the other hand, functional connectivity is a measure of the ability of organisms to move among patches of suitable habitat in a fragmented landscape (Hilty et al. 2006, Taylor et al. 1993, Fahrig 2003). Ideally, measures of functional connectivity are derived from actual data about landscape composition, habitat use, and movement by wildlife. Such detailed data is uncommon at the landscape level because it is costly to collect.

When empirical field data on species movement are unavailable, connectivity estimates can be derived from mathematical models. Models may be based on empirical studies of species’ abundance or occurrence among different land cover types, or on expert opinion of species’ habitat associations. Given the major influence the matrix has on connectivity among habitat fragments (Ricketts 2001), several models based on matrix connectivity have been developed including habitat resistance (friction; Ray et al. 2002, Joly et al. 2003), least-cost paths (Adriaensen et al. 2003), circuit theory (McRae et al. 2008), habitat permeability (Merenlender 2011, Theobald et al. 2012), and linkage designs (Beier and Brost 2010).

Here we describe landscape permeability models derived from an estimated statistical relationship between specific landscape features related to the built environment and species detections from empirical studies (Forman 2000, Reed 2007, Merenlender et al. 2009). Permeability models are an extension of the resistance concept (Ray et al. 2002); model output often is in the form of a grid-based map with a value assigned to each cell that represents its permeability to an organism’s movement. The permeability models were developed for linkage analysis by the Land Trust of Santa Cruz County (Merenlender and Feirer 2011) and designed to make biologically informed approximations of community assemblage response to habitat quality (Metzger and Décamps 1997). The built environment, especially roads, urban and suburban development can reduce the ability for wildlife to move across the landscape (Fu et al. 2010; Tannier et al. 2012). Santa Cruz County, California, harbors some of the world’s most majestic redwood and mixed conifer coastal forestlands; however, residential development is wide spread at urban to exurban densities. The Land Trust of Santa Cruz County and other environmental organizations in the area are actively trying to conserve open space and biodiversity (Press et al. 2002). Conservation and land use planners in the region face challenges commonly encountered in areas with sprawling development, including how to maintain wildlife movement across an increasingly developed landscape (Girvetz et al. 2008).

We compare these model estimates with occurrence data for pumas (Puma concolor), a generalist predator commonly used as a focal species for connectivity analysis (Terbough et al. 2001, Crooks 2002, Cardillo et al. 2005, Beier 2009), in the Santa Cruz Mountains. Pumas are
the largest predatory in the study area, and are known to travel long distances (Dickson & Beier 2002). The question guiding our analysis is: How well do our model estimates of landscape permeability derived from simplified, biologically-informed connectivity models compare with actual occurrence of a generalist predator across a gradient of land use?

Methods

Study area

The Santa Cruz Mountain range is in central California, adjacent to the San Andreas Fault (122° 7’ to 121° 50’ W, 37° 21’ to 36° 53’ N). Forming a ridge along the San Francisco Peninsula, the mountains separate the Pacific Ocean from the Santa Clara Valley. Our study area (217,375 ha) was an “island” of relatively undeveloped land within the Santa Cruz Mountains, situated between the Pacific Ocean on the west, and the metropolitan centers of San Francisco, San Jose, and Santa Cruz to the north, east, and south, respectively. The four primary land cover types within the study area were (1) forest and woodland, (2) shrubland and grassland, (3) agricultural land, and (4) land that is developed or otherwise of human use (US Geological Survey, Gap Analysis Program 2011). The study area was bounded to represent assumed puma occurrence in the region (Figure 1).

Our study area faces encroachment by development as the populations of surrounding San Francisco, Santa Clara, and Santa Cruz counties steadily increase. The annual population growth for each county between 1980 and 1990 ranged between 0.7% and 1.29% (U.S. Census Bureau, Population Division). The study area was also surrounded and intersected by highways, such as California State Route 17 that bisects the study area and connects the cities of San Jose and Santa Cruz.

Climate in the Santa Cruz Mountains is Mediterranean with mild, wet winters and cool, dry summers. The average summer and winter temperatures are 20° C and 10° C, respectively. Annual precipitation is approximately 114 cm, the majority of which falls between November and April. Heavy summer fog provides drizzle and fog drip to the western, ocean-facing part of the range, creating a cool coastal habitat supporting coast redwoods (Sequoia sempervirens) and Douglas fir (Pseudotsuga menziesii). At higher elevations and on sunny south slopes, the microclimate is warm and dry with drought-resistant chaparral vegetation including manzanita (Arctostaphylos spp.) and California scrub oak (Quercus berberidifolia).

Landscape permeability maps

We used regression models derived from mesocarnivore and bird assemblage response to human-modified land cover and landscape configuration as inputs to construct potential permeability maps (Figure 2). For each permeability map, we used as input a regression model derived from two indices of habitat fragmentation: distance to roads (yROADS; Forman 2000) and median patch size (yPATCH; Reed 2007). We calculated each permeability map with ArcGIS 9.3.1 software (ESRI, Redlands, CA, USA). We applied each regression model to create a map using both the permeability value and the geographical position and orientation of all relevant landscape elements in the study area (per Šafner et al. 2011). All permeability values ranged between 0.0 – 1.0 with a cell size of 30 m x 30 m (900 m2).

Distance to Roads

There is overwhelming evidence of the effects of roads on natural communities (Fahrig and Rytwinski 2009), and thus we used distance from road, scaled by traffic volume (yROADS), as
an index of animal response to transportation infrastructure. We calculated $y_{\text{ROADS}}$ based on empirical data from several prior studies that evaluated the impact of roads on wildlife (Forman 2000, Forman and Deblinger 1999, Reijnen et al. 1995, Reijnen et al. 1996). Forman (2000) described the correlation between the distance to a road and bird species abundance and diversity; the closer a location is to a road, and the greater the road’s traffic level, the larger the road effect, resulting in a corresponding decrease in abundance and diversity of urban avoiding birds. This approach assumes that the maximum magnitude of the road effect and effect-distance are proportional to the volume of traffic along the road.

We applied the equation derived by Forman (2000) to calculate the maximum effect-distance for each road in the study area as a function of mean traffic volume, measured as annual average daily traffic:

$$x_{\text{ED}} = 0.0126w_{TV} + 178.75$$

where $w_{TV}$ is the average traffic volume of the road, and $x_{\text{ED}}$ is the road effect-distance.

We then assumed that the magnitude of effect of any given road would be proportional to the maximum effect and would decline linearly with increasing distance from the road. Thus, the road effect of each cell was calculated using the following equation:

$$y_{\text{ROADS}} = -\left(\frac{1}{\max(x_{\text{ED}})}\right) z_{\text{ROADS}} + \frac{x_{\text{ED}} - \max(x_{\text{ED}})}{\max(x_{\text{ED}}) + 1}$$

where $z_{\text{ROADS}}$ is the Euclidean distance from the nearest road and $y_{\text{ROADS}}$ is the magnitude of the road effect. We calculated the permeability map for $y_{\text{ROADS}}$ with ArcGIS 9.3.1 software using road effect values from $y_{\text{ROADS}}$ and the geographical position and orientation of all relevant landscape elements in the study area (per Safner et al. 2011) (Figure 2B). The traffic volume data came from the California Department of Transportation (http://traffic-counts.dot.ca.gov). In our study area, the maximum effect-distance $\max(x_{\text{ED}})$ for all roads was 2812 m.

**Median patch size**

We used median patch size ($y_{\text{PATCH}}$) as a landscape-scale, area-informed index of habitat integrity calculated using the contiguity and relative size of proximate habitat patches. There is increasing recognition that area-informed metrics are useful to explain variation in wildlife abundances and movement capacity (Magle et al 2009) and perform well in analyses of landscape connectivity (Bender et al. 2003). We defined a patch as a contiguous area of habitat with natural vegetation cover and whose land use(s) were compatible with the establishment of mesocarnivore home ranges, based on information from prior empirical studies. The model for $y_{\text{PATCH}}$ was derived from a study (Reed 2007) investigating the correlation between patch size and mesocarnivore (e.g. coyote, bobcat, gray fox) occurrence in northern California, which found that the frequency of mesocarnivore detections increased with the size and contiguity of adjacent patches. $y_{\text{PATCH}}$ was calculated as the median area of habitat patches within a fixed buffer radius. In exploratory analyses, Reed (2007) found that $y_{\text{PATCH}}$ measured at a buffer distance of 2,500 m explained the most variation in detections of the greatest number of mammalian carnivores. This work also revealed ‘median patch size’ to be a better predictor than buffered radius indices or proximity metrics (Reed 2007).

Per Reed (2007), we calculated $y_{\text{PATCH}}$ using the equation:

$$y_{\text{PATCH}} = \frac{0.2356(x_{\text{PATCH}})^{1/2} + 1.385}{\max(y_{\text{PATCH}})}$$
where $x_{PATCH}$ is the median patch size in hectares (ha) within a 2,500 m radius buffer, and $y_{PATCH}$ is the effect of habitat integrity on landscape resistance, measured as the density of native mesocarnivore detections along a survey transect.

As input data for $y_{PATCH}$, we used a map of terrestrial vegetation cover from existing land cover data (Farming Mapping and Monitoring Program 2008). The land cover data was then modified to remove roads (Research and Innovative Technology Administration, Bureau of Transportation Statistics 2001), mines and quarries, water bodies, and all land parcels less than 2 ha. Selecting which patches should be analyzed as part of a habitat connectivity network is very difficult in a landscape where privately owned wild lands exist outside the protected areas, as is the case in Santa Cruz County. We began by selecting the larger patches in the landscape, which we defined to be any patch greater than 250 acres (101 ha). In addition to these larger patches, smaller patches found in the more fragmented parts of the study area were included if they were the largest patch within a fixed kernel distance ranging between 1 km from any given point in the landscape – a range of median dispersal distances expected for terrestrial vertebrates found in the area. We used the $y_{PATCH}$ equation to calculate the patch size effect for each grid cell in the permeability map (Figure 2A).

**Puma GPS data**

We used location point data for 30 (F = 15, M = 15) pumas (P. concolor) collected between October 2008 – January 2013 (Wilmers et al. 2013), for a total of 115,384 Geographic Positioning System (GPS) point locations. We visualized puma point location data in the ArcGIS 9.3.1 Geographic Information System (GIS) using the WGS1984 geographic coordinate system and an Albers projection. The Institutional Animal Care and Use Committee at UC Santa Cruz approved all animal-handling procedures (Wilmc1101). Roughly 53% of the puma occurrence locations in the study area were in forest or woodland, 20% in shrubland or grassland, 20% in developed land, and 4% in agricultural land (US Geological Survey, Gap Analysis Plan 2011).

**Data analysis**

To evaluate whether puma occurrence corresponded with landscape permeability, we compared expected land use with observed landscape use by pumas. We calculated the relative frequency of landscape permeability values across the entire study area and the relative frequency of landscape permeability values for only the puma points. We made these comparisons for $y_{ROADS}$, $y_{PATCH}$, and $y_{ROADS+PATCH}$ to assess the relative effect of each factor in influencing in habitat use. To determine the level of difference between the puma points and all the points in the study area we created a frequency distribution table for each data set, calculated the odds ratio between the two data sets, and used a Monte Carlo simulation method to generate a confidence interval estimate.

To determine the relative frequency of landscape permeability values, we extracted the cell values for the 2.37 million cells, created a frequency distribution table with 20 intervals (from 0.0 – 1.0 in 0.05 landscape permeability unit increments), and calculated the abundance of expected points ($A_E$) by interval. We repeated this calculation for each permeability map to create the relative distributions of landscape permeability for each regression model with which to compare with puma occurrence. We defined $A_E$ as:

$$A_E = \frac{\# \text{ points in interval}}{2,372,749}$$

To determine the relative frequency of the landscape permeability values for only the puma points, we extracted the permeability values at all puma point locations ($N = 115,384$),
created a frequency distribution table with 20 intervals (from 0.0 – 1.0 in 0.05 landscape permeability unit increments) for the 115,384 values for the puma, and calculated the abundance of observed points \((A_O)\) by interval. We repeated this calculation for each permeability map to summarize the relative distribution of permeability values for puma for each regression model. We defined \(A_O\) as:

\[
A_O = \frac{\# \text{ points in interval}}{115,384}
\]

To determine the level of difference between the puma points and all the points in the study area we calculated the odds ratio between the expected \((A_E)\) and observed \((A_O)\) abundance (Bland and Altman 2000). To calculate the odds ratio \((OR)\), we compared the distribution of landscape values \((E)\) with that of puma point locations \((O)\) for each regression model. We calculated odds \((ODDS)\) of occurrence by interval for \(A_E\) \((ODDS_E)\) and \(A_O\) \((ODDS_O)\) using the following two equations:

\[
ODDS_E = \frac{A_E}{1 - A_E}
\]

\[
ODDS_O = \frac{A_O}{1 - A_O}
\]

Next, we calculated the odds ratio \((OR)\) by interval between \(A_E\) and \(A_O\) using the following equation:

\[
OR = \frac{ODDS_O}{ODDS_E}
\]

Finally, we transformed \(OR\) using a log10 transformation and plotted the distribution. When there was no difference between \(A_E\) and \(A_O\), \(OR\) values were between 1 – 0; when \(A_E > A_O\), \(OR\) values were <1 (negative); and when \(A_E < A_O\), \(OR\) values were >1 (positive).

To approximate a confidence interval for distribution of permeability values in the study area we used a Monte Carlo simulation approach (Manly 2007) with 1000 iterations to extract landscape permeability values for 115,384 random grid locations from each permeability map. We separated the randomly subsampled points in a similar fashion (20 intervals from 0.0 to 1.0) and calculated the abundance of subsampled points \((A_S)\) by interval. We defined \(A_S\) as:

\[
A_S = \frac{\# \text{ points in interval}}{115,384}
\]

We visualized the results as a boxplot and used the minimum and maximum values from the boxplot as a confidence interval estimate.

Examining the location data at the population level could potentially mask sampling bias, such as the number of data points per puma or where pumas had been caught for collaring. Each puma had variable detection times – ranging from once per day to once every 15 minutes – and differing numbers of total collection days. The median collection time across all puma detections was once every 3.95 h (once every 237 minutes). The data were not corrected for autocorrelation, nor standardized for collection time in order to take advantage of all known detections. Behaviorally correlated features in the environment may be overrepresented in the distribution of puma point locations – a known issue for home range or utilization distribution estimates for individual animals (Swihart and Slade 1985). However, there is no a priori reason to expect a bias in levels of permeability and higher number of points collected. To explore the importance of study area delineation and potential false absences on the final results, we compared the study area at two different geographic extents: (1) the full study area and (2) a constrained study area cropped closely around the puma occurrence data.

**Results**

We compared the puma occurrence data with each of the three models independently as well as in various multi-model combinations. The odds ratio results showed distinct differences
between the availability of permeability values and the detection of pumas for all three models (Figure 3). Our results showed much of the landscape in the Santa Cruz Mountains is of intermediate permeability for $y_{\text{PATCH}}$ and high permeability for $y_{\text{ROADS}}$. Specifically, 88.50% of the study area had a landscape permeability value for $y_{\text{ROADS}}$ between 0.9 – 1.0. For $y_{\text{PATCH}}$, 88.50% of the study area had a landscape permeability value for $y_{\text{PATCH}}$ between 0.3 – 0.6, corresponding to a median patch size of 11.1 – 21.21 ha (Figure 3). The distribution of puma location points was similar, with 90.06% of puma locations found at points with a $y_{\text{PATCH}}$ permeability value between 0.3 – 0.6 (11.1 – 21.21 ha), and 97.03% of puma locations found at points with a $y_{\text{ROADS}}$ permeability value between 0.9 – 1.0 (Figure 3).

The logOR distribution for the lowest $y_{\text{PATCH}}$ permeability values showed a distinct difference between the availability of habitat with very low permeability values and the distribution of puma detections (interval 0 – 0.15; average logOR = -0.66). However, for habitat with intermediate and high permeability values, log OR was alternatively positive and negative (interval 0.2 – 1.0). Specifically, log OR was positive between 0.2 and 0.4, and negative between 0.4 and 0.6. There were no cells in the study area with a value between 0.65 and 0.7, nor between 0.85 and 0.95.

For $y_{\text{ROADS}}$, the logOR interval between the minimum landscape permeability value of 0.28 – 0.95 was negative (average logOR = -0.82). LogOR was smallest for habitat with permeability values between 0.28 – 0.35 (logOR = -2.40), corresponding to the most disturbed habitat. It was only for the least disturbed regions in the landscape that logOR values were positive (logOR = 0.39).

As expected of a combined model, the logOR distribution for the lowest $y_{\text{ROADS}+\text{PATCH}}$ permeability values was similar to the trend shown for $y_{\text{PATCH}}$ and $y_{\text{ROADS}}$. The logOR values were negative for low permeability habitat (interval 0.16 – 0.55; logOR = -1.17), and log OR was smallest for the most disturbed regions (interval 0.16 – 0.2; logOR = -2.67). As with the $y_{\text{PATCH}}$ model, logOR alternated between positive and negative values for habitats with intermediate and high permeability values (interval 0.55 – 1.0). Specifically, logOR was positive between 0.55 - 0.6 as well as 0.85 - 0.9, and negative between 0.75 – 0.8 and 0.95 – 1.0.

**Discussion**

By measuring landscape permeability associated with human development, this research offers an easy to calculate spatially explicit method to examine landscape permeability that can help land-use planners prioritize habitat corridors for improved wildlife movement.

Pumas in the Santa Cruz Mountains were not found in areas with low permeability scores as compared with the availability of this type of disturbed habitat. However, pumas were found in landscapes with moderate to high permeability in proportion to its availability. Pumas used areas with estimated moderate landscape disturbance, in the form of nearby low traffic volume roads, as well as land with intermediate median habitat patch size. There appeared to be a possible threshold below which pumas begin avoiding disturbed landscapes, equivalent to a permeability score of ≤ 0.2 for $y_{\text{PATCH}}$, and ≤ 0.90 for $y_{\text{ROADS}}$ (Figure 3).

Surprisingly, pumas did not preferentially occupy the least disturbed areas close to large core habitat patches or far from major roads. This finding might reflect an avoidance of habitat with steeper slopes. Pumas have been shown to avoid land with steep slopes (Beier 1995), possibly for energetic reasons (Dickson et al. 2005), or to preferentially utilize bottom canyons...
and gentle terrain where available (Dickson and Beier 2006). Because human activity tends to decrease on steep slopes, when pumas do travel on steep slopes they travel more closely to houses present there (Wilmers et al. 2013). Slope tends to correlate with high permeability for sites within the Santa Cruz Mountains. For example, 5.6% of $y_{\text{ROADS+PATCH}}$ values over 0.7 in our study area have a slope > 50%. Because our study area includes very steep slopes and sheer cliffs, these locations could be the least disturbed because they are topographically unsuitable for development.

Across all models, pumas appear to avoid areas with very low permeability. Species avoidance of habitat with low permeability values may be even stronger for other carnivores and bird species than we observed for pumas. To explore the importance of taxonomic differences between the species under consideration and the modeled species assemblage – such as dissimilar sensitivity (Crooks 2002) or scale of response to environmental variables – the model predictions could be compared to occurrence data for a mesocarnivore like the gray fox ($U. \text{cinereoargenteus}$) or a migratory songbird. Such a comparison would likely show a different habitat use pattern than that of a large, generalist carnivore like the puma, and perhaps result in clearer trends between moderate and undisturbed landscapes.

For our analysis, we pooled all data points and examined the location data set as a whole at the population level. All puma observations were taken from contiguous home ranges within the study area and are assumed to represent the entire resident population during the sampling period, although there could have been migrating non-resident individuals in the area that were not collared. Imperfect detection is a known issue with sampling an elusive carnivore like the puma that can negatively bias occupancy estimates (Karanth et al. 2011) and in this case could mean that pumas may be present in the study area that were not detected or collared. The number of individual pumas within the study area was limited and there is inherent selection bias as to the home range location of each individual. For this reason, we pooled the observations rather than analyzing each puma’s location points separately. We found no difference between the results from our comparison of the study area delineation, and so we based our conclusions on the entire study area.

Because landscapes are increasingly urban (US: Alig et al. 2004; Europe: Antrop 2004), and becoming more so globally (McDonald 2008), it is important to identify social-ecological processes and land use patterns – such as patch size and distance to roads – that impact a variety of wildlife in a similar way. The permeability maps were designed to make a general, community-level habitat quality assessment based on linear regression models derived from species assemblages in northern California, developed for linkage analysis as requested by the Land Trust of Santa Cruz County (Merenlender and Feirer 2011). The original linkage exercise for the Land Trust of Santa Cruz County was coarse, completed in response to managers’ desire for generalized permeability maps of places lacking species data and modeling capacity. In this analysis, we evaluated the how community-level habitat permeability maps compared with existing occurrence data for a focal species of conservation concern in the area. To that end, this analysis presents a simple comparison of permeability maps based on land cover and noted responses by entire assemblages of species rather than detailed research on puma habitat selection. See Wilmers et al. (2013) for details about behavioral responses of pumas to human development using mixed effects modeling. Others have shown how the behavioral features of the focal species can change the dynamic permeability of the studied landscape (Dickson et al. 2013, Stoner et al. 2013) and therefore warrants further study.
Our work shows that biologically-informed, structural permeability models do reflect puma habitat use – and thus functional connectivity – on the ground to some extent. While the puma is an important species of conservation concern, the use of any sole focal species to assess landscape adequacy cannot fully capture the needs of other integral species in the landscape. Instead, this evaluation supports the use of a structural permeability model for conservation planning efforts. For example, this approach was used to identify potential habitat linkages for inclusion in the Conservation Blueprint for Santa Cruz County.

Our models were derived from mesocarnivore and bird species detections from empirical studies, but compared to data set from a large-bodied apex predator often used for connectivity planning (Beier et al. 2006). While our roads model was originally developed for bird community response to traffic volume in a geographically distant environment from our study area, it showed a reasonably strong relationship with the empirical puma data. However, the model’s utility may vary by location, environment, and species of interest. We included this response to roads model based on bird community assemblage data, despite obvious differences between the environments and study animals, because this was the best data on traffic impacts on wildlife available when we applied these models.

Although both mesocarnivores and carnivores are predators, their diets differ in terms of prey size and composition (Elbroch et al. 2013; Robinette et al. 1959). Thus, an evaluation of functional connectivity based on data from a smaller-bodied organism with a correspondingly reduced home range size and distinct diet composition may show different results. Future landscape-level evaluations of habitat permeability could be enhanced by including organisms of differing sizes and from multiple trophic levels (e.g. mesocarnivores, migratory birds)

**Conclusion**

The comparison of the permeability model described and the animal field observations shows that a model constructed using information about animal response to human land use can be an informative component for reserve design, land management, and conservation planning. Specifically, landscape planners and conservation organizations should focus on conserving the most permeable core areas and linkages while taking care to avoid landscapes with steep slopes. Steeper areas are also at lower risk for development, and, hence, do not require conservation action.

Land use planners can employ our approach to conservation planning in fragmented landscapes even when species data are unavailable. For example, this research was used to identify potential habitat linkages for inclusion in the Conservation Blueprint for Santa Cruz County. The priority linkages, comprised of the most natural land cover types that facilitate wildlife movement among core natural areas, were identified in the Santa Cruz Mountains Connectivity Report, which is available from the Santa Cruz Land Trust.

**Acknowledgements**

This project was made possible thanks the Land Trust of Santa Cruz County and our gratitude goes to A. Mackenzie, M. Freeman, J. McGraw, and R. Bancroft. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1106400. S. Feirer’s help with GIS analysis at the University of California Hopland
Research and Extension Center was invaluable. The help of J. Kitzes and A. Hsu for statistical analysis ideas are much appreciated.
References


IUCN. 2004. IUCN Red list of threatened species. IUCN, Gland, Switzerland.


Figures

Figure 1
A map of the study area depicting the Santa Cruz Mountains in California, USA showing the 115,384 puma points we sampled from 2008 – 2013 and major roads through the region. Our study area was a 217,375.26 ha “island” of relatively undeveloped land within the Santa Cruz Mountains, situated between the Pacific Ocean on the west, and metropolitan centers of San Francisco, San Jose, and Santa Cruz to the north, east, and south, respectively.
Figure 2
Distribution of permeability values for landscape (above), and with puma points overlaid (below), across the study area. A) Median patch size layer ($y_{\text{PATCH}}$). B) Distance to roads layer ($y_{\text{ROADS}}$). C) Combined distance to roads and median patch size layer ($y_{\text{PATCH+ROADS}}$).
Figure 3
Relative abundance of permeability values for landscape and puma points across the study area, with Monte Carlo results visualized as a confidence interval estimate above, and log OR results below. A) Median patch size layer ($y_{\text{PATCH}}$). B) Distance to roads layer ($y_{\text{ROADS}}$), with inset showing detailed abundance for landscape permeability values < 0.95. C) Combined distance to roads and median patch size layer ($y_{\text{PATCH+ROADS}}$).
Chapter 2: Including land use improves the predictive ability of a species distribution model

Abstract

Landscapes are increasingly urban, resulting in habitat fragmentation, road infrastructure, and greater population densities from humanity’s expanding footprint. Despite the influence of the built environment on wildlife populations (size and range), species distribution models rarely incorporate human land use variables – a practice that may lead to inaccurate model estimates. Further, because of sampling biases inherent to most presence only data, failing to address bias may compromise the accuracy and reliability of model predictions. Here, I assessed whether the inclusion of human land use information improved the predictive performance of a model of gray fox (Urocyon cinereoargenteus) distribution at an urban-wildland interface. Two sources of gray fox location data were included in the analysis: 100 presence-only observations collected by citizen scientists and 209 presence/absence observations collected by academic researchers. All gray fox locations were collected in California between 2006 and 2016. To account for sampling bias, detections of non-gray fox species in the order Carnivora were used as background pseudo-absence points. A combination of 9 environmental and land use variables were used as input into maximum entropy (MaxEnt) models to predict the distribution of gray foxes. I analyzed each predictor in terms of its effect on model accuracy, variable importance score, and overall spatial prediction. Including human land use variables improved the accuracy of the MaxEnt model, as measured by the area under the curve (AUC) value. These results show that omitting human land use likely over predicts the distribution of suitable habitat for the gray fox. Additionally, using citizen science data can be informative in distribution models, improving spatial coverage and providing better insight into the important drivers of species distributions.

Introduction

Landscapes across the globe are increasingly urban (Alig et al. 2004; Antrop 2004; McDonald et al. 2008; Butchart et al. 2010). More than half of the world’s inhabitants currently live in cities, and global urbanization forecasts predict the urban population to reach 70% by 2050 (United Nations 2012). Urbanization has numerous effects on natural habitat and is one of the primary threats to biodiversity and its conservation (Tilman 2001). Overt impacts of urbanization, like fragmentation of natural habitat, parcelization (Gobster & Rickenbach 2004), and exurban sprawl (Theobald 2005) may isolate wild species populations, limit dispersal, and lead to demographic and genetic problems (Magle et al. 2008). Proximity to human activity is associated with pervasive wildlife mortality in the form of animal-vehicle collisions (Forman et al. 2013) and human activities like hunting, poisoning, and predation (Collins & Kays 2011). Despite the influence of human land use on the distribution of species (Wilmers et al. 2013; Gray et al. 2016), land use variables remain underutilized as input (Beninde et al. 2015) in models of species distributions. A major challenge in conservation biology is predicting how wildlife populations respond to urbanization and using this knowledge to maintain biodiversity by conserving natural habitat in human-dominated landscapes.
The ability to identify habitat characteristics required for species persistence, as well as the factors that limit species distributions, is essential for conservation planning. Species distribution models based on landscape features, expert opinion, or species occurrence data, are widely used to predict biodiversity distributions and aid the decision-making process for diverse conservation applications (Rocchini et al. 2011). The species distribution modeling program MaxEnt (Phillips & Dudik 2008), predicts distributions across space using georeferenced presence-only occurrence records and environmental variables, and has consistently fared well in model comparison studies (Elith et al. 2006; Wisz et al. 2008; Elith & Leathwick 2009; Thibaud et al. 2014). MaxEnt has been used to assess habitat use and species distributions for numerous vertebrate taxa, (geckos, Pearson et al. 2008; American black bears, Baldwin and Bender 2008; bush dogs, DeMatteo and Loiselle 2008; mule deer and gemsbok, Hoenes and Bender 2010; sage grouse, Yost et al. 2008), and has also been used to evaluate explicit conservation threats, including forecasting range shifts with climate change (Franklin 2013) and the potential distribution of invasive species (Cuban tree frogs, Rodder & Weinsheimer 2009). While these applications have seen wide use, few MaxEnt models incorporate land use variables, despite clear empirical evidence that habitat modification affects species ranges.

An advantage of using MaxEnt to create species distribution models is the ability to use as input presence-only data, such as museum and herbaria samples that span decades in time and a diverse geographic range (Graham et al. 2004). However, these data are often collected opportunistically, resulting in a bias toward more readily accessible locations near roads, hiking trails, or human settlements (Reddy & Dávalos 2003). Several authors have raised the issue of sampling bias in MaxEnt models (Kadmon et al. 2004; Costa et al. 2010; Leitão et al. 2011; Syfert et al. 2013), yet the potential effect of data bias is rarely quantified or corrected for. Methods proposed to address bias in models include spatial filtering of the presence locations (Kramer-Schadt et al. 2013), providing weights for species records using bias kernels (Elith et al. 2010), or using a set of background pseudo-absence points selected outside a minimum distance to the presence locations (Engler et al. 2004; Lobo et al. 2010). Using as input pseudo-absence background locations collected in a manner similar to that of the target species presence-only locations, may normalize bias across the model and can provide better model results (Phillips et al. 2009). Pooling survey and collection data for multiple species allows the model to borrow strength across species to more efficiently estimate the bias and improve inference from presence-only data (Fithian et al. 2015).

In this study, I evaluated the impact of incorporating land use in a species distribution model predicting the habitat suitability of the gray fox in northern California, USA. Gray foxes are mid-sized carnivores with a home range compatible with landscape-scale analysis (Kitzes & Merenlender 2013), and are geographically widespread across the diverse habitat types found in the study area including forest, brush chaparral, and urban margins (Hall 1981, Eisenberg 1989, Cooper et al. 2012, Haroldson & Fritzell 1984, Sawyer & Fendley 1994, Chamberlain & Leopold 2000, Temple 2007). The gray fox was used as a focal species because of its life history traits and demonstrated sensitivity to human presence. The species is also of high conservation value in California (Myers et al. 2000). Georeferenced presence only observation records for gray foxes from iNaturalist and presence absence locations from an earlier field study conducted by academic researchers (Bidlack 2007) were used as input data. The MaxEnt models were built using nine environmental and human land use variables that are heterogeneous across the landscape. To account for sampling bias, detections of non-gray fox species in the order
Carnivora were used as background pseudo-absence points. Overall the three objectives of the study were to (1) test whether including human land use variables in the two models improved their performance, (2) compare the relative importance of habitat variables between the models for gray foxes, and (3) compare models built using presence only and presence-absence data.

**Methods**

**Study area**

The study area included six counties in California’s San Francisco Bay Area: Alameda, Contra Costa, Marin, Napa, Solano, and Sonoma (Figure 1; centroid: 95989 m, 44551 m; north: 96969 m; south: -60413 m; west: -306523.424 m; east: -129740.089982 m; NAD 1983 California Teale Albers projection). These six counties cover 13,434 km² and support a total population of 4,106,875 (U.S. Census Bureau 2015); while covering only 3% of California’s land area, they constitute 1% of the human population (U.S. Census Bureau 2015). The population has been steadily rising, with an average increase of 7% over 5 years. The study area is also within the California Floristic Province, one of 25 global biodiversity hotspots that provides critical habitat for 584 vertebrates and 2,125 endemic plants (Myers et al. 2000). The large number of habitat types encompassed in the study area with high geodiversity and the resulting microclimates provides an opportunity to model a relatively small space with numerous heterogeneous environmental variables. This diversity in environmental variables allows an analysis across a region large enough to encompass numerous gray fox home ranges, which average 1.29 km² in California (Hallberg and Trapp 1982), and thus discern which variables are important. Additionally, the region encompasses a range of human land uses including large cities and metropolitan areas, along with smaller urban and rural areas.

**MaxEnt model construction**

MaxEnt was used to generate two pairs of species distribution models for gray fox occurrence in six Bay Area counties, for a total of four models. Each pair consisted of a reduced model that included six environmental input variables, and a full model that included the environmental variables as well as three land use variables. The first model pair used georeferenced presence only observation records for gray foxes from iNaturalist collected by the general public in 2010–2016 (iNat observations = 100). iNaturalist is an online database of species observations submitted by citizen scientists. The detections in iNaturalist are considered opportunistically collected, and tend to be associated with areas that are easier for the public to access. “Research grade” records for the species “Urocyon cinereoargenteus”, within “California”, and with a photograph to confirm species identity (December 17, 2016, [www.inaturalist.org](http://www.inaturalist.org)) were used as input data. To be designated as “research grade”, an iNaturalist observation must be georeferenced, have a photo and/or sound, have a date, and the taxon identification must be agreed upon by more than 2/3 of the identifiers. Species observations in iNaturalist are automatically georeferenced by a smartphone app at the time of collection, so the presence locations were assumed to be accurate. Each gray fox detection was visually confirmed prior to inclusion in the analysis. The second model pair used presence locations from an earlier field study (Bidlack 2007) (Transect observations = 209). Presence locations were collected in 2005-2006 using a systematic sampling method employing transect searches for scats using scat detection dogs; scats were identified to species using genetic sequencing (Bidlack 2007). Detections of non-gray fox species in the order Carnivora from the
corresponding data source were used as background points for these model pairs (iNat observations = 578, Transect observations = 732).

Model variables of potential biological relevance for the distribution of gray foxes were selected for inclusion in the MaxEnt models: climate, topography, and biotic variables (Table 1). Minimum winter temperature (°C), maximum summer temperature (°C), average cloud cover (days/year), precipitation (mm/year) and cumulative water deficit (mm/year; CWD) were included to represent the microclimate heterogeneity across the study area. Elevation (m) was included because topographical heterogeneity influences the microclimate, local hiding place availability, and terrain inaccessibility (human presence and predators) and may be positively associated with species presence (e.g. Kuemmerle et al., 2010, Fritzell & Haroldson 1982, Grinnell et al, 1937; Johnson et al 1948). Finally, percent canopy cover was included as cover may facilitate movement and access to food resources, and be positively associated with species presence as gray foxes are arboreal (Hall 1981, Eisenberg 1989, Cooper et al. 2012, Haroldson & Fritzell 1984, Sawyer & Fendley 1994, Chamberlain & Leopold 2000, Temple et al. 2010, Temple 2007, Farias et al. 2012). Detailed description and sources of data for all variables are listed in Table 1.

The full model included the six aforementioned environmental variables as well as three human land use variables: (1) human population density (number of people/acre), (2) road density (average number of roads/500m²), and (3) mean parcel size. The road density and mean parcel size rasters were created in ArcGIS 10.4.1. The road density raster was created by taking the average density of streets, excluding freeways and major roads, using a 500m² radius. The mean parcel size raster was created using 2014 California statewide parcel data, with a kernel window of 4. Because highly correlated variables can impair the accuracy of model predictions, all variables were confirmed to be minimally correlated using the software R (R Development Core Team, 2010) (Pearson correlation coefficient | r | < 0.53 for all pairs).

All environmental layers were rescaled in ArcGIS 10.4.1 with a spatial resolution of 30 arc-seconds (~1 km). All species occurrences and data layers were projected to the NAD 1983 Albers coordinate system, and the geographic extent was fixed to north: 96969.965977 m, south: -60413.282925 m, west: -306523.424331 m, east: -129740.089982 m. Each variable was exported as an ASCII file, as were the gray fox and Carnivora occurrence points in sample with data format.

**MaxEnt model analysis**

All models were configured to run a training data set of 80% gray fox locations and a test set of 20% locations using default parameter settings and bootstrapped replication (n=20). The default non-threshold model settings for cumulative output were used, with the exception of increasing the iteration maximum from 500 to 5000.

The models were compared using three approaches. Model specificity was evaluated by comparing the area under the curve (AUC) estimates for each model, one of the most widely used threshold-independent evaluators of model discriminatory power (Fielding & Bell, 1997). The AUC is equal to the probability that the model will rank a randomly chosen presence site higher than a randomly chosen absence site. A random ranking has on average an AUC of 0.5, and a perfect ranking achieves the best possible AUC of 1.0. Models with values above 0.75 are considered potentially useful (Elith 2002), and those with AUC > 0.9 are considered very successful models (Swets 1988).
The relative ranking of, and the response curves for, each variable was compared across all model pairs. Variable importance was quantified using the percent permutation importance, a measure of the fraction of total variance explained by the model that resulted from iterations of the algorithm based on the variable in question, which has been shown to be reliable in variable evaluation (Halvorsen 2012). The response curves were visually examined to qualitatively understand how predicted probability of gray fox occurrence related to each variable, noting the direction of the response, the range of suitability, and the local maximum.

Finally, the difference between the paired model predictions was evaluated by comparing the mean output from each model using a 0.5 threshold. The spatial distribution of each model prediction was mapped in ArcGIS 10.4.1, and the percent of suitable habitat for gray foxes shared between the two models was calculated for each model pair.

**Results**

MaxEnt was used to generate two pairs of species distribution models for gray fox occurrence, for a total of four models. Each pair consisted of a reduced model that included six environmental input variables, and a full model that included the environmental variables as well as three land use variables. The analysis of the MaxEnt model performance using AUC showed that including indices of land use improved the sensitivity and precision of the MaxEnt model for gray foxes, as measured by AUC (Table 2). The AUCs for the full models were both over 0.8 (iNat = 0.8, transect = 0.856), which would classify the models as “good” (Araújo et al. 2005). The AUCs for the reduced models were 2-4% lower than those of the corresponding full models (iNat = 0.773, transect = 0.837).

While the probability of gray fox distribution predicted by the two model pairs showed some differences in the exact locations of areas of highest suitability, two general trends across both models resulting from the two data sets emerged. First, both full models predicted a concentrated gray fox distribution, whereas the reduced models predicted a more dispersed distribution, across the study area (Figure 2). The average size of overlapping habitat with predicted suitability >0.5 was 54.2% between the full and reduced model pairs (iNat = 52.4%, transect = 56.0%). However, a sizeable amount of the suitable habitat was identified only by the full models (iNat = 16.8%, transect = 18.4%). In a comparison between the two model pairs, the iNaturalist models predicted more land was suitable for gray foxes overall. Specifically, the iNaturalist models predicted 29.4% of the study area, whereas the transect models predicted 18.9%.

In the full models, the three land use variables explained between 15.7% and 24.8% permutation importance. While the exact ranking of land use variable importance differed between the full models, parcel size and road density were of intermediate importance in both. Specifically, average permutation importance for parcel size was 10.3% for parcel size (iNat = 11.7%, transect = 8.9%) and 6.25% for road density (iNat = 9.4%, transect = 3.1%). Human population density was ranked last in both full models, with an average permutation importance of 3.7%. While the contribution of individual land use variables to the full models varied, the responses of gray fox distribution to land use was consistent. Gray fox distribution was correlated with locations on undeveloped parcels that were between 125-150 hectares, with low road density, and in areas with low human population density. The response figures illustrating the shape of the impact of each variable for the two model pairs are shown in Figures 3 and 4.
As with the land use variables, the exact ranking of the environmental variables across the full and reduced models differed. Overall, precipitation was the single most important predictor of gray fox presence across all four models. The average permutation importance of precipitation was 18.9% for the full models (iNat = 11.1%, transect = 26.6%), and 26.2% for the reduced models (iNat = 21.9%, transect = 30.5%). Gray fox occurrences were correlated with locations with warmer winters, cooler summers, intermediate cloud cover, higher elevation, and low levels of precipitation.

**Discussion**

The results highlight the importance of including indices of human land use in delineating species distributions, and suggest that land use variables are important for land managers to include in models to forecast conservation planning. Including human land use variables improved the sensitivity and precision of the MaxEnt model for gray foxes, as measured by AUC. Additionally, omitting land use resulted in a model that over-predicted suitable habitat in the study area, likely overestimating the current distribution of this species.

By using pseudo-absence background points collected using similar methods to those employed for the gray fox presence only data, I used one method to address sampling bias. However, as others have noted (Araújo & Guisan 2006; Loiselle et al. 2008), the inability to validate predictions from models created using presence only data remains a limitation. For rare or data-deficient species, the best alternative may be to evaluate model predictions using data simulations that generate a “known” distribution against which the MaxEnt results may be tested (Fithian et al. 2015). Despite these limitations, using data from other species as pseudo-absence background data provides a first step to address sampling bias in presence only data like that found in databases of citizen science collected observations like iNaturalist, as well as biocollection records from museums and herbaria.

While gray foxes have been documented using human-disturbed areas (Kapfer & Kirk 2012), margins of urban areas (Harrison 1997), and locations close to human habitation (Rountree 2004; Temple et al. 2010), they prefer natural habitats with some element of vegetative cover (Rountree 2004; Temple et al. 2010; Riley 2006) and select against open habitats (Farias et al. 2012). Specifically, gray foxes prefer tree cover (Cooper et al. 2012, Chamberlain & Leopold 2000, Wood et al., 1958; Progulske and Labisky, 1997), particularly when establishing core areas (Temple et al. 2010). Gray foxes were present primarily in developed areas that also contained adequate canopy cover, based on the model, which is consistent with the finding that gray fox distribution is dependent upon the presence of mature woody plants (Kapfer & Kirk 2012).

While individual climatic variables did not rank as highly as precipitation, as a group, the indices of climate (January minimum temperature, July maximum temperature, and cloud cover) comprised an average of 49.5% permutation importance for the two reduced models and 37.2% for the two full models. The results are consistent with those from a global study that suggests climate and anthropogenic pressures are more influential than biological traits in predicting terrestrial mammal range size (Di Marco & Santini 2015). Because a species’ range size is commonly used as a proxy for extinction risk (Gaston & Blackburn, 1996; Purvis et al., 2000; Ceballos & Ehrlich, 2002; Diniz-Filho et al., 2005), the importance of climatic variables highlights the potential impact future climate change may have on species distributions and, ultimately, extinction risk. The high explanatory power of land-use in the model suggests that
predicted distributions from strict climate-envelope models could be improved with the incorporation of future land use change projections to better inform likely changes in mesocarnivore distributions.

When compared with the full model results, those of the reduced models overpredicted the distribution of gray foxes in the study area. Management decisions for biodiversity conservation often rely on quantitative estimates of organisms (Gorrod & Keith 2009). Thus, accurate species distribution estimates, such as those produced by including indices of land use in these Maxent models, are useful for prioritizing land conservation and future predictions given expected changes in land use. It is worth noting that unlike the data layers used for climatic variables, those for indices of land use can be the outcome of planning decisions including conservation and open space. In principle, this kind of work could inform the density of land use, in addition to prioritization forecasts.

In a comparison between the two model pairs, the iNaturalist models predicted more land was suitable for gray foxes overall. Specifically, the iNaturalist models predicted 29.4% of the study area, whereas the transect models predicted 18.9%. This finding illustrates the impact the input data set, given the same suite of model variables, can have on MaxEnt predictions. The differences between the models derived using iNaturalist and transect input data were largest in Sonoma and Contra Costa Counties.

The research also demonstrates the value of including citizen science data into species distribution models as a way to increase spatial coverage and provide insight into the important drivers of species distributions in the absence of systematic survey data. In urbanized settings a large number of people engage with the study and conservation of local wildlife, and the role of volunteer citizen scientists in biological monitoring and conservation management is growing (Schmeller et al. 2009, Lukyanenko et al. 2016). Accounting for location accuracy set by only using highly resolved data (e.g., from smartphones), can alleviate concerns about the quality of spatial data obtained from citizen scientist observations. Additionally, many citizen science volunteers already have topical expertise or education in biology (Foster-Smith & Evans 2003; Cohn 2008), characteristics that suggest they may be as qualified and capable as professionals. In a systematic review of the peer-reviewed literature on the quality of data collected by volunteers, Lewandowski and Specht (Lewandowski & Specht 2015) found no strong evidence to support the belief that volunteer data are consistently less precise than professional data. On the contrary, citizen science data accuracy has been shown to be comparable to that of expert observations (Crall et al. 2011a; Nagy et al. 2012; Jordan et al. 2012; Zapponi et al. 2015). Still, issues in citizen science observations exist, such as the potential unknown sampling biases associated with observations being preferentially located proximally to roads and human habitation. While I could not completely quantify nor correct for this bias, this analysis used data from taxonomically related species, collected using the same survey techniques employed for each gray fox data, as pseudo-absence background data to normalize the bias across the models.

Although citizen science observations like those found in iNaturalist are not yet distributed evenly worldwide, they should be considered as input for species distribution models requiring presence only data depending on the study location and species of interest. iNaturalist has a total of 1,292,712 observations and 82,229 species across all 7 continents, of which 1,098,653 (57%) and 27,266 (61%) are verifiable and research grade (VRG), with birds (354,737 VRG, 91%) and mammals (51,551, 74%) having the highest percent VRG (Accessed 31 March 31, 2016). What this means is verifiable and research grade observation data for birds and
mammals (n=406,288 VRG) are high quality and suitable for input into species distribution models.

Several important caveats in the study merit attention. Most importantly, I used a single species for the analysis, specifically a habitat generalist sensitive to the built environment. The effect of human land use will be taxon specific, resulting in models with these variables showing differing amounts of influence and importance. While it is understood that the importance of biophysical variables changes with the taxa used in analysis (bird, insect, plant diversity and species richness: Beninde et al. 2015), we must extend our understanding of a species’ realized niche to include human land use. Subsequent species distribution model analyses that incorporate land use could be particularly important for species that show a range in their tolerance of the human development footprint; land use may still be an important predictor for mesocarnivores known to be more sensitive to (bobcat, *Lynx rufus*) or more tolerant of (raccoon, *Procyon lotor*) human presence.

It is also worth noting that the importance of climatic variables may change with study location and size. The study area included habitat near the ocean and the resultant maritime airflow, as well as land with topographic diversity leading to patches of extreme microclimates (Daly 2006) that may not be present in other study areas. Projecting the model into different temporal or spatial extents could alter the ranking of individual variables, although I expect land use will remain a significant contributor to the model.

**Conclusions**

The results of this model evaluation show the importance of considering human land use as a predictor when creating species distribution models. Specifically, including land use improved the accuracy of the MaxEnt model for a mesocarnivore sensitive to development, the gray fox. For animals sensitive to development, omitting land use variables in species distribution models may over-predict suitable habitat and lead to inaccurate prioritization for conservation action. Creating models that accurately translate the impact of land use on wildlife into a quantifiable metric for species distribution models is particularly needed in regions with rapid human population growth seeking conservation-oriented development. Additionally, this study acknowledges the potential of including presence-only citizen scientist observations with MaxEnt models to increase the sample size and spatial coverage for species distribution modeling.

**Acknowledgements**

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1106400. The author thanks Sarah Reed and Alison Bidlack for sharing their gray fox data, and iNaturalist for their ongoing citizen science efforts that contributed to this analysis.
References


Borchert, M. I. 2012. Mammalian carnivore use of a high-severity burn in conifer forests in the San Bernardino mountains of Southern California, USA. Hystrix, the Italian Journal of Mammalogy.


### Figures

#### Table 1
Environmental input variables for the full and reduced modeling scenarios.

<table>
<thead>
<tr>
<th>Variable</th>
<th>F</th>
<th>R</th>
<th>Data source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canopy cover</td>
<td>+</td>
<td>+</td>
<td>National Land Cover Database 2011, USFS Tree Canopy cartographic</td>
<td>Percent tree canopy cover, with file pixel values ranging from 0 to 100 percent, with each individual value representing the area or proportion of that 30m cell covered by tree canopy.</td>
</tr>
<tr>
<td>Cloud cover</td>
<td>+</td>
<td>+</td>
<td>Compiled using the MODIS Satellite, processed by Park Williams and Chris Still at the University of California Santa Barbara.</td>
<td>The fraction of days with cloud cover at 10:30am from July 8 - Sept. 30th 2000-2006.</td>
</tr>
<tr>
<td>Cumulative water deficit</td>
<td>+</td>
<td>+</td>
<td>2014 California Basin Characterization Model (BCM) Downscaled Climate and Hydrology 30-year Summaries</td>
<td>Climatic water deficit quantifies evaporative demand exceeding available soil moisture, and is quantified as the amount of water by which potential evapotranspiration exceeds actual evapotranspiration.</td>
</tr>
<tr>
<td>Elevation</td>
<td>+</td>
<td>+</td>
<td>National Elevation Dataset (NED)</td>
<td>The NED is the best available raster elevation data of the conterminous United States at 30m resolution.</td>
</tr>
<tr>
<td>Human population density</td>
<td>+</td>
<td>-</td>
<td>2000 Dasymetric Data, San Francisco Bay Area</td>
<td>Areal interpolation, using boundaries that divide the area into zones of relative homogeneity with the purpose of better portraying the population distribution, of aggregated census block-group data to a human population density surface.</td>
</tr>
<tr>
<td>January minimum temperature</td>
<td>+</td>
<td>+</td>
<td>PRISM high-resolution spatial climate data for the United States</td>
<td>January minimum temperature (°C) averaged over 1986 – 2005.</td>
</tr>
<tr>
<td>July maximum temperature</td>
<td>+</td>
<td>+</td>
<td>PRISM high-resolution spatial climate data for the United States</td>
<td>July maximum temperature (°C) averaged over 1986 – 2005.</td>
</tr>
<tr>
<td>Parcel size</td>
<td>+</td>
<td>-</td>
<td>2014 California statewide parcel boundaries</td>
<td>All of parcel boundary datasets collected by the Information Center for the Environment under funding.</td>
</tr>
</tbody>
</table>
from the California Strategic Growth Council.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Road density</td>
<td>+</td>
<td>-</td>
<td>2015 TIGER/Line</td>
<td>Primary and secondary roads described by Topologically Integrated Geographic Encoding and Referencing (TIGER)/Line for the U.S. Census Bureau.</td>
</tr>
</tbody>
</table>
Figure 1
A map of the 13,434 km$^2$ study area depicting six counties in the San Francisco Bay Area: Alameda, Contra Costa, Marin, Napa, Solano, and Sonoma in California, USA showing scat transect (white) and iNaturalist (black) gray fox detections recorded between 2005 and 2016. General groups of vegetation and land use types, as well as major roads and highways, are indicated.
Figure 2
Maps of the study area showing the input data, and reduced and full model predictions for the iNaturalist and scat transect data sets. The input data (left) show the gray fox locations (large circles) overlaid on the locations of other species in the order Carnivora (small gray circles). The reduced (center) and full (right) model prediction maps span four breaks: 0.2, 0.4, 0.6, 0.8, and 1.0, with probability of detection highest in warm colors.
Table 2
Performance summary for the two model pairs, showing AUC and ranked variable importance (percent permutation importance). Human land use variables are indicated in bold. In both pairs, the AUC of the full model was greater than that of the reduced model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>AUC</th>
<th>Importance</th>
<th>Variable</th>
<th>AUC</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full model</strong></td>
<td></td>
<td></td>
<td><strong>Scat transects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cloud cover</td>
<td>0.8</td>
<td>26.7</td>
<td>Precipitation</td>
<td>0.856</td>
<td>26.6</td>
</tr>
<tr>
<td>Precipitation</td>
<td>11.1</td>
<td>11.1</td>
<td>January minimum</td>
<td>17.9</td>
<td></td>
</tr>
<tr>
<td>Elevation</td>
<td>10.7</td>
<td>14.5</td>
<td>Elevation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>July maximum</td>
<td>9.4</td>
<td>13</td>
<td>Canopy cover</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Road density</strong></td>
<td>9.4</td>
<td></td>
<td>Parcel size</td>
<td>11.7</td>
<td></td>
</tr>
<tr>
<td>Canopy cover</td>
<td>9.2</td>
<td></td>
<td>July maximum</td>
<td>10.3</td>
<td></td>
</tr>
<tr>
<td><strong>Parcel size</strong></td>
<td>8.9</td>
<td></td>
<td>Road density</td>
<td>3.1</td>
<td></td>
</tr>
<tr>
<td>January minimum</td>
<td>8.1</td>
<td></td>
<td>Cloud cover</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td><strong>Human population density</strong></td>
<td>6.5</td>
<td></td>
<td><strong>Human population density</strong></td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td><strong>Reduced model</strong></td>
<td>0.773</td>
<td>29.5</td>
<td>Precipitation</td>
<td>0.837</td>
<td>30.5</td>
</tr>
<tr>
<td>July maximum</td>
<td>12.6</td>
<td>14.3</td>
<td>Elevation</td>
<td>14.3</td>
<td></td>
</tr>
<tr>
<td>Canopy cover</td>
<td>11.2</td>
<td>11.5</td>
<td>January minimum</td>
<td>11.5</td>
<td></td>
</tr>
<tr>
<td>January minimum</td>
<td>4</td>
<td>6.8</td>
<td>Cloud cover</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 3
Modeled responses for each variable in the pair of models using gray fox transect data.

Transects: Reduced model

Transects: Full model
Figure 4
Modeled responses for each variable in the pair of models using gray fox transect data.

**iNaturalist: Reduced model**

![Graphs showing modeled responses for canopy cover, cloud cover, elevation, January minimum temp, July maximum temp, and precipitation for the reduced model.]

**iNaturalist: Full model**

![Graphs showing modeled responses for canopy cover, cloud cover, elevation, January minimum temp, July maximum temp, precipitation, human population density, parcel size, and road density for the full model.]

43
Chapter 3: Human land use decreases genetic connectivity in a small mammal

Abstract

Combining spatial and genetic data can improve our understanding of ecosystems modified by human land use. Molecular genetic approaches can provide a direct measure of the influence of human land use, and specifically roads, on gene flow among populations segregated by roads. However, few studies have quantified the impact of road density and the relative importance of road characteristics on gene flow or provide evidence of population isolation associated with fragmentation. To assess whether the presence of roads is a barrier to connectivity, I evaluated the genetic diversity and structure of 61 ground squirrels (*Otospermophilus beecheyi*) from population separated by a major freeway in Coyote Valley, Santa Clara Co., CA using the mitochondrial gene cytochrome b. The spatial extent of genetic structuring was assessed using interpolation analysis, and the effect of landscape permeability and road density were evaluated with Mantel tests and resistance surfaces. Based on haplotypes across 615 base pairs, the genetic similarity among populations on the same side of roads is higher than among populations separated by major roads, in particular, a large highway. These results indicate that habitat fragmentation by roads is a barrier to gene flow in ground squirrels, resulting in quantitative differences in genetic structure and diversity between populations. Using a combination of spatial and molecular methods provide unique insight into the impacts of human land use on wildlife and can be used to inform conservation planning.

Introduction

Ongoing human actions continue to threaten biodiversity worldwide, particularly the processes of habitat loss and fragmentation (Tittensor et al. 2014; Newbold et al. 2015). Urbanization may be the most severe form of fragmentation, with impacts on wildlife that include restricting animal movement (Riley et al. 2006), reducing habitat quality and quantity (Prugh et al. 2008; Öckinger et al. 2009), or increasing human disturbance (Merenlender et al. 2009). Traffic infrastructure (hereafter “roads”) is a form of fragmentation that negatively effects wildlife, both directly, through mortality and habitat loss, and indirectly, by population division and isolation (Forman & Alexander 1998; Forman 2003; Shepard et al. 2008). Roads are unique as barriers in their disproportionate impact relative to their size. For example, the 6.6M km of roads in the United States cover less than 1% of the land (CIA 2012), yet they effect an estimated 20% of the landscape (Forman 2000).

One way to lessen the harmful effects of urbanization on wildlife is to conserve the pieces of land that link natural habitat patches across the landscape, thereby allowing species to persist around human development (Kindlmann & Burel 2008). Maintaining this landscape connectivity has become a central theme in ecology and conservation (Rands et al.
as corridors of intact habitat facilitate current ecosystem functionality (Tewksbury et al. 2002; Damschen et al. 2006) and may provide access to future climate refugia (Hannah 2011; Nuñez et al. 2013). Habitat connectivity can be conceptualized as structural or functional. Structural connectivity is derived from landscape attributes such as the shape, size, and configuration of habitat patches, but does not account for actual land use by wildlife (Calabrese & Fagan 2004; Theobald et al. 2012). Structurally isolated populations may have reduced access to resources like food or mates, and face barriers to dispersal when separated by inhospitable habitat (Moilanen & Hanski 1998).

The complement to structural connectivity is functional connectivity, which is a measure of the ability of organisms and their genes to move among patches of suitable habitat in a fragmented landscape (Taylor et al. 1993; Fahrig 2003; Hilty et al. 2006). Ideally, measures of functional connectivity are derived using empirical data about landscape composition, habitat use, and movement by wildlife. Functionally isolated wildlife populations can exhibit decreased genetic diversity (Epps et al. 2005), which could result in inbreeding depression (Roelke et al. 1993; Hedrick & Kalinowski 2000; Keller & Waller 2002), increased disease susceptibility (Spielman et al. 2004; Charlesworth et al. 2009), and greater risk of extinction, especially for species residing in small or isolated habitat patches (Fahrig & Merriam 1994).

Combining spatial and genetic data can improve our understanding of ecosystems modified by human land use (Levy et al. 2013; Anderson et al. 2015). The interdisciplinary field of landscape genetics combines genetic, computational, and spatial analytic tools (Epps et al. 2007) to understand how landscape features affect spatial genetic structure (Manel 2003, Storfer 2007), which can be used as surrogate measures of functional connectivity (Boulet et al. 2007). Specifically, landscape genetic approaches allow the quantitative estimation of effective separation that exists due to barriers in gene flow between individuals or populations (Vignieri 2005), and often reflect the geographic distance between populations as well as barriers to gene flow – whether natural or human made. These geographic distances can be measured as the crow flies (Euclidean distance) between two sampled populations, or take into account estimates of matrix permeability between populations (Taylor et al. 1993, With et al. 1997).

There are several ways to evaluate landscape connectivity and predict optimal corridors to connect habitat patches, such as least cost paths (Adriaensen et al., 2003) and linkage designs (Beier & Brost, 2010). However, gene flow among real populations is not restricted to a single, optimum corridor. Resistance surfaces, based in electric circuit theory, have been applied in genetic analyses of connectivity because they simultaneously consider all possible pathways that may connect population pairs (McRae & Beier 2007b; McRae et al. 2008; Spear et al. 2010; Lamy et al. 2013). Thus, resistance surfaces can be thought of as hypothesized relationships between landscape variables and movement (O’Brien et al. 2006) or gene flow (Wang et al. 2008).

While advances in spatial analysis and increasingly available molecular approaches have made it more accessible to investigate the impact of human land use, and specifically roads, on connectivity, several gaps remain. Some work has been done using landscape resistance models to investigate the effect of linear features on population genetic structure (Lee et al. 2012; Marrotte et al. 2014), but little work has been done to quantify the impact of road density and the relative importance of road characteristics on connectivity (Balkenhol & Waits 2009). Additionally, evidence of fine scale (within 5-10 km) genetic effects associated with fragmentation has been rare.
To assess the impact of roads on functional connectivity, I evaluated the genetic diversity and structure of ground squirrels (Otospermophilus beecheyi) across a land use gradient in Coyote Valley, CA. Although Coyote Valley has been identified as a critical linkage (Penrod et al. 2013), little is known about actual wildlife movement or genetic diversity across the valley floor. Ground squirrels are widespread throughout almost all habitats across California, and can commonly be found along roadsides and in croplands (Fitch 1948), making them an easy species to sample. Using a landscape permeability model, I identified a series of 15 transects with permitted access across a gradient of landscape permeability and on both sides of an eight lane freeway, U.S. Highway 101 (hereafter “Highway 101”). Ground squirrel DNA was obtained from scats collected along transects, sequenced using the mitochondrial gene cytochrome b, and analyzed for indices of genetic diversity and structure. The specific goal of this research was to quantify the extent to which primary and secondary roads, road density, and low landscape permeability may present a barrier to ground squirrel gene flow. Here, molecular analyses were combined with landscape analysis of potential movement pathways across the matrix to evaluate the influence of anthropogenic and environmental variables on the genetic patterns observed in ground squirrels. The results reported here form part of a conservation planning effort underway by public and private land trusts, California Department of Transportation, and California Department of Fish and Wildlife to assess connectivity across Coyote Valley for species found in the western Santa Cruz Mountains and the eastern Diablo Range. The findings of this research will be used to inform conservation planning across the valley are also useful for other regions with similar road networks.

Methods

Study area

Coyote Valley is a floodplain in central California, located at the southern portion of Santa Clara County, CA (centroid: 37°11’ N, 121°42’W), in a rapidly developing region between San Jose to the north and Morgan Hill to the south. The study area (2,914 ha) is a narrow valley that connects two large regions of undeveloped land, the Santa Cruz Mountains to the west and Diablo Range to the east. The four primary land cover types within the study area were (1) shrubland and grassland (36.5%), (2) forest and woodland (34.4%), (3) land that is developed or otherwise of human use (23.6%), and (4) agricultural land (4.7%) (US Geological Survey, Gap Analysis Program, 2011) (Figure 1). The study area faces encroachment by development as the populations of Santa Clara County and the surrounding San Francisco Bay Area grow. The population in Santa Clara County is steadily increasing, with a change of +7.7% between 2010 and 2015, from 1.78 million to 1.92 million residents in the 5-year time period (U.S. Census Bureau, Population Division, 2015). The study area is bisected by a large highway, United States Highway 101, and is intersected by two arterial roads and several secondary roads. Recent modifications to the section of Highway 101 that passes through the study region have included adding additional lanes and enlarging the median barrier.

Climate in the study area is Mediterranean with mild, wet winters and warm, dry summers. The average summer and winter temperatures were 25°C and 10°C, respectively. Coyote Valley spans a spectrum of land use and road infrastructure intensity. Habitat on the west
side of Highway 101 is a mosaic of agriculture, residential, and native grassland and woodland. Roads on the west side of the study area include two arterial roads (Monterey Highway and Santa Theresa Avenue), several secondary roads, and numerous narrow residential streets. Habitat on the east side of Highway 101 is more topographically diverse, and is comprised of privately owned, undeveloped grassland with few roads. The entire study area provides habitat for numerous native species, including predators of ground squirrels such as red-tailed hawks (*Buteo jamaicensis*), horned owls (*Bubo virginianus*), Cooper’s hawks (*Accipiter cooperii*), rattlesnakes (*Crotalus oreganus*), gopher snakes (*Pituophis catenifer*), bobcats (*Lynx rufus*), and coyotes (*Canis latrans*).

**Landscape permeability**

An existing landscape permeability model (Gray et al. 2016) was extrapolated across the study area to estimate structural connectivity. The resulting structural connectivity estimate for the study area was used in two ways. First, the permeability model results were used inform transect sampling. Scat sampling efforts were stratified such that samples were collected across a land use gradient. The permeability model estimates were also used to evaluate connectivity predictions using circuit theory analysis.

The landscape permeability models were derived from an estimated linear relationship between specific landscape features related to human land use (e.g. traffic volume, housing density) and bird and mesocarnivore detection levels from empirical field studies. These models were designed to make a general, community-level habitat quality assessment based on linear regression models derived from species assemblages in northern California (Forman 2000; Reed 2007; Merenlender et al. 2009). When compared with animal field observations, the permeability estimates of these biologically informed, structural models were shown to reflect animal habitat use on the ground (Gray et al. 2016). Thus, habitat permeability models constructed using information about animal response to human land use activities can be an informative component for land management and conservation planning in fragmented landscapes even when species data are unavailable.

**Ground squirrel samples**

Ground squirrels are an important prey source for aerial and terrestrial predators, and create burrows that are used by several amphibian species to prevent desiccation during dry months (Fitch 1948). Because the average ground squirrel home range size is approximately 120 m around a burrow (Fitch 1948), numerous individuals can be sampled within the study region.

Scat samples were collected at 15 transects located at sites along a permeability gradient using the permeability map described above (Gray et al. 2016) across Coyote Valley to identify areas of predicted high, medium, and low permeability, and used these locations to inform the location of scat surveys. Eleven transects were to the west of Highway 101 and 4 transects were to the east (Figure 1). Transects were walked on multiple separate occasions between the dates August 14, 2015 and June 30, 2016. For each sampling event, all available ground squirrel scats were collected outside burrows using sterile tweezers and placed in a 2 ml screw-cap tube containing DET solution (20% DMSO, 0.25 m EDTA, 100 mm Tris, pH 7.5 and NaCl to saturation) per Seutin et al. Tubes were kept frozen at -20°C until processing.
Ground squirrel DNA was extracted from 83 scat samples using Qiagen’s QIAamp DNA Stool Mini Kit. The resultant DNA was amplified and sequenced using primers for cytochrome b (815 base pairs; L14723: ACCAATGACATGAAAAATCATCGTT and H14896: TAGTTGTCCGGGTCTCTCTA; Ducroz et al. 2001). Cytochrome b is commonly used as a region of mitochondrial DNA for determining phylogenetic relationships in animals, due to its internal sequence variability. While the primer regions of the cytochrome b locus are broadly conserved across vertebrates, the coding region within the gene varying considerably between taxa. Cytochrome b has been extensively used in intra- and inter-specific molecular systematic studies.

The following conditions were used for initial double-stranded amplifications: 2 µl template (10 ng), 7.25 µl sterile H2O, 0.35 µl of each primer (10 µM), 0.25 µl dNTPs (10mM), 0.75 µl MgCl₂ (25mM), 1.2 µl 10X buffer, 0.25 µL Bovine Serum Albumin (20 µg/µL), and 0.1 µl Taq polymerase (Invitrogen, Carlsbad, CA) to a final volume of 12.5 µl. Amplification conditions consisted of the following sequence: an initial 3 minutes denaturation at 95 °C followed by 35 cycles of the following sequence: 30 seconds denaturation at 95 °C, 1 minute annealing at 60 °C, and 1.5 minutes extension at 72 °C. The amplification concluded with a final extension of 2 minutes at 72 °C. The presence of amplified product at 815 bp was confirmed on a 1.5% agarose gel electrophoresed for 40 minutes at 120V. Amplified products were purified by adding 0.2µL ExoSAP to each tube and holding the samples at 37 °C for 30 minutes to digest remaining PCR reagents, followed by 15 minutes at 80 °C to denature the ExoSAP enzyme. The purified cytochrome b products were cycle-sequenced using the following conditions: 6.96 µL water, 1.63 µL 5X BigDye buffer, 0.75 µL BigDye, 0.2 µL L14723 (10 µM) or 0.2 µL H14896, and 0.5 µL PCR product. Cycle sequencing conditions consisted of the following sequence: an initial 1 minute denaturation at 96 °C followed by 25 cycles of the following sequence: 10 seconds denaturation at 96 °C, 5 seconds annealing at 50 °C, and 4 minutes extension at 60 °C. Cycle-sequenced products were purified through a Sephadex column and analyzed on an ABI 377 automated sequencer. Nucleotide sequences were visually inspected using the Geneious Pro software v. 5.1.7 (Kearse et al., 2012), and aligned using MEGA7. Haplotypes of 615 bp were aligned using ClustalW. Representative haplotypes generated for this study were deposited in GenBank (submission ID 1968012, accession numbers pending).

**Genetic diversity**

To explore how land-use, and roads in particular, may affect genetic diversity in ground squirrels, genetic diversity metrics within and between the east and west populations were calculated. The number of haplotypes, gene diversity (Nei & Genetics 1973), nucleotide diversity (Tajima 1983) were calculated for all samples in the program Arlequin 3.5.2.2 (Excoffier & Lischer 2010). Population subdivision was quantified using F-statistics. Global and pairwise estimates of genetic differentiation were examined using F<sub>ST</sub>. Uncorrected pairwise differences were used to calculate relatedness. Significance was assessed with 10,000 randomizations of the AMOVA test statistic.

A network, based on all haplotypes, was created using median-joining in the program PopART v5 (http://popart.otago.ac.nz). This method organizes haplotypes into networks according to genetic distance. Haplotypes are connected when the number of differences between them does not exceed a 95% parsimony probability threshold (Templeton et al. 1992). Haplotype networks are useful in illustrating genetic divergence at the intraspecific level,
especially in cases of multiple haplotypes that are derived from a single ancestral sequence (Templeton et al. 1992).

Genetic structure

Three approaches were used to characterize the patterns of the spatial distribution of genetic variation in the samples. Initially, the pattern of isolation by distance (IBD) across the fifteen populations was calculated by correlating a matrix of pairwise genetic distances among all populations with Euclidean distances between sampled locations using a Mantel matrix correspondence test. The centroid of each population cluster was measured in ArcGIS 10.4.1 (ESRI). Pairwise matrices of geographic distance and F$_{ST}$ were compared using Mantel tests for matrix correlation (Mantel 1967), with significance assessed by 10,000 randomizations of the genetic distance matrix using IBDWS 2.3 (Jensen et al. 2005). Genetic distances and geographic distances were log-transformed.

The second way genetic structure was evaluated was by using circuit theory to map the functional distances between sampling locations to determine least cost paths between the 15 populations. Circuit theory treats landscapes as conductive surfaces, and predicts connectivity using resistance, voltage, and current. Locations with high current flow are predicted to be more connected, and, conversely, locations where current is funneled through a narrow area are considered “pinch-points” of constrained connectivity. Using the Circuitscape tool (McRae & Shah 2009) in ArcGIS 10.4.1, I generated maps of predicted animal movement (“current”) across the study area using the landscape permeability raster described above (Gray et al. 2016) as a resistance surface. To convert permeability to resistance, the permeability raster was loaded as conductance (= 1 / resistance). The resulting current map was then analyzed to generate least cost paths between each population using the Linkage Mapper toolkit version 1.1 (McRae & Cavanagh 2011) run in ArcCatalog. Linkage Mapper uses maps of core habitat areas and raster maps of resistance to movement to identify and map least-cost linkages between core areas. Each cell in a resistance map is attributed with a value reflecting the difficulty of moving across that cell. Cost-weighted distance analyses create maps of least-cost corridors between core areas that are then normalized and mosaicked to create a single composite corridor map that shows the relative value of each grid cell in providing connectivity between core areas. For this analysis, the fifteen populations were used as “core habitat”, and the cumulative current raster generated with Circuitscape was used as the resistance surface. Pairwise matrices of genetic distance and least-cost paths were then compared using Mantel tests for matrix correlation (Mantel 1967), with significance assessed by 10,000 randomizations of the matrix using IBDWS 2.3 (Jensen et al. 2005).

Lastly, the spatial pattern of genetic diversity between the sample locations was calculated using landscape shape interpolation in the program Alleles in Space (Miller 2005). Interpolation produces a three-dimensional surface plot where X and Y axes correspond to geographical locations and surface heights (Z-axes) represent genetic diversity that contains peaks in areas where there are large genetic distances. The analysis begins by first constructing a connectivity network among all of the sampling locations in the data set, then generates a connectivity network based on all pairwise locations. After the connectivity network is created, raw genetic distances (“surface heights”) between observations connected in the network are placed at the midpoints of each connection on the network using an inverse distance weighted
interpolation procedure (Watson 1992, Watson and Phillips 1985). Genetic distances are calculated using mismatch distances between sequences. Specifically, for aligned sequences of length n, the genetic distance between individuals i and j will be:

\[ D_{ij} = \frac{1}{n} \sum_{k=1}^{n} d_k \]

Where \( d_k \) takes on a value of 1 if individuals i and j contain different nucleotides at site k, and is 0 if the nucleotides are identical. Grid dimensions adjust the resolution of the resulting interpolation surface, with a 10x10 grid creating a coarse surface with little detail, and a 100x100 grid creating a surface with substantial detail and more apparent localized peaks. A distance weighting parameter of 0 will assign equal weight to all points regardless of their distance from a point on the grid, resulting in a meaningless flat surface. From this analysis, a landscape shape plot was generated with a 50 x 50 grid and a distance weight parameter of 2.

**Results**

**Landscape permeability**

The application of the structural connectivity model across Coyote Valley showed a broad range of permeability values, with much of the region with relatively high permeability. The distribution of permeability values ranged between 0.14 and 0.71 (Figure 2), with an average of 0.47 (standard deviation = 0.14). 66% of land in Coyote Valley has a permeability value between 0.45-0.75, a habitat permeability range that is preferentially used by wildlife, as demonstrated for pumas (*Puma concolor*) by Gray et al. (2016). This distribution indicated that there is land of relatively high permeability within Coyote Valley, and such habitat is not rare. Further, the permeability of that habitat adjacent to U.S. Highway 101 was shown to be the lowest in the study region, with an average of 0.15.

**Genetic diversity**

For this analysis, 83 scat samples were collected at 15 groups. 61 samples were successfully sequenced (73.5% success rate for amplification), with 32 west samples and 29 east samples. Nucleotide diversity (\( \pi \)) in the population west of Hwy 101 was 0.322, which was greater than that in the east population (0.219). For the 615 bp of the cytochrome b mtDNA, I detected 10 haplotypes. Six haplotypes were detected in the West population, and eight were detected in the East population. Two haplotypes were unique to the West populations (h2 and h4), and four were unique to the East populations (h1, h7, h9, and h10) (Table 2).

Genetic structure The pairwise \( F_{ST} \) value was 0.4, indicating significant genetic differentiation between the east and west ground squirrel populations. The Mantel test for matrix correlation of genetic distance and Euclidean distance (log-transformed) between populations showed very little evidence of isolation by distance, resulting in \( r = 0.1292 \), which was significant (one sided \( p < 0.05 \)). The Mantel test result for correlation between genetic distance and length of least-cost paths also showed very little correlation between this adjusted distance measure and genetic distance, with \( r = 0.0613 \), which was significant (one sided \( p < 0.05 \)).
The software program Alleles in Space was used to create a visualization of interpolated genetic divergence over geographic space in the form of a heat map. The largest area of genetic divergence was located in the area surrounding and including Highway 101 (Figure 3). There was also an area of high divergence on the west side of the study region near arterial and secondary roads, and agricultural land.

The Circuitscape resistance map created using the landscape permeability raster was remarkably similar to the interpolated genetic divergence map (Figure 3). The area with the greatest resistance was located surrounding and including Highway 101. However, the resistance map considered streets to have much higher resistance than the interpolation results derived from genetic distance.

**Discussion**

Using two different methods, traditional pair-wise genetic distance analysis (FST; Table 1), and landscape genetic analysis (Figure 3), there is strong evidence of genetic differences between sample locations on the east and west sides of Highway 101. Furthermore, the two methods showed similar and strong genetic effects of fragmentation, with the largest genetic divergence over the widest expanse of urban areas surrounding and including Highway 101.

Pairwise FST showed that the genetic divergence was significant. The genetic diversity of populations on the same side of roads is more alike than between populations separated by roads, and genetic structure showed the populations on the same side of the highway share more haplotypes, indicating roads are a barrier that isolates ground squirrels resulting in quantitative differences in genetic structure and diversity between populations separated by roads. These results are consistent with other findings that suggest the roads with high traffic volume are relatively impenetrable for many species. For example, Delaney et al. (2010) found significant genetic divergence and loss of genetic diversity in four vertebrate species between neighboring populations bifurcated by roads.

Both genetic interpolation and landscape resistance analyses showed populations separated by Highway 101 to be the most distinct. The permeability map estimated low levels of permeability across areas with small streets in addition to the highway. The road effect for this estimate of permeability was derived using published data on how mesocarnivore and bird detections decline close to roads, which may be different than the impact of roads for ground squirrel movement. While the resistance map provides an estimate of landscape permeability, the genetic interpolation map is informed by species-specific data providing an estimate of functional landscape connectivity for ground squirrels.

The results reveal a weak significant relationship between genetic distance and Euclidean distance \((r = 0.1292, \text{ one sided } p\text{-value} < 0.05)\), as well as with the length of the least cost paths between sampled populations \((r = 0.0613, \text{ one sided } p\text{-value} < 0.05)\). These Mantel r values are much less than what other studies have shown to substantiate isolation by distance relationships. For example, in an evaluation of a resistance-based connectivity model, McRae and Beier (2007a) demonstrated a strong positive relationship, with Mantel \(r > 0.79\), between isolate by resistance predictions and genetic distance for two endangered species (mahogany: \(r > 0.82, p\text{-value} < 0.0005\); wolverines: \(r >0.79, p\text{-value} <0.0001\) (McRae & Beier 2007a)).

It would be useful to examine how highways influence ground squirrel population structuring at other sites to determine how generalizable these findings are. Also, while
comparable numbers of samples were collected from the east and west sides of Highway 101, there were more sampling locations on the west side. This was because of land ownership on the east side, resulting in difficulty accessing the area for a diversity of sampling locations. Additionally, habitat permeability at the boundaries of this analysis would be affected by neighboring landscapes and their use. For example, the presence of roads and residential development to the north of Coyote Valley would likely reduce landscape permeability, whereas the open habitat to the west would not. Expanding this landscape permeability analysis to include the wider planning area would provide additional information about the matrix within which the study area is situated.

In this analysis, the current connectivity of ground squirrel populations was evaluated with the assumption that genetic structure and diversity are related to extant landscape influences. However, historical influences can be important drivers of currently observed patterns in landscape genetics (Epps & Keyghobadi 2015). Because Coyote Valley is a floodplain, historical water flows could contribute to the connectivity results shown here. One way to test the influence of historical events on current genetic connectivity is to include a comparison of genetic sequences from historical ground samples. No samples have as yet been found from this study site before the highway was built. For example, the Museum of Vertebrate Zoology at UC Berkeley has an expansive collection of historical specimens, the collection did not contain ground squirrel specimens collected from within the study region (accessed: November 15, 2016).

Wildlife use of roads varies based on many factors such as animal type, body size, and mobility; and road width, composition, traffic volume, and traffic speed. Thus, a seemingly low permeability value of 0.144 as seen along Highway 101, Monterey Highway or Santa Theresa Avenue may not indicate that the road is a complete barrier to all varieties of birds or terrestrial animals. Future research that investigates functional connectivity through the use of animal movement or genetic data for birds or larger bodied mammals would evaluate the effects of habitat fragmentation at a landscape scale. To this end, camera traps have been installed at road crossing structures throughout the study area, which will yield detection rates for a suite of wildlife and provide data for a quantitative evaluation of animal response to the built environment.

Conclusions

This analysis shows the importance of considering both structural connectivity associated with human development and functional connectivity at the molecular level. This dual approach could inform conservation planning and wildlife management by identifying critical linkages for wildlife movement. Using permeability models to inform sampling design and including molecular analysis to evaluate functional connectivity is valuable.

In addition to the contribution these results make to the fields of landscape genetics and road ecology, the outcomes of this study have practical utility by informing conservation planning through collaboration with regional land management to prioritize habitat for acquisition. For example, the results could be used to create a connectivity design for Coyote Valley that would reduce animal-vehicle collisions by facilitating wildlife movement underneath roads via bridges and culverts. High speed rail is being planned that would run parallel to highway 101, which would further isolate populations on the east and west sides of Coyote Valley.
Valley, but may also provide an injection of funding for conservation mitigation projects. Given the impacts of roads on wildlife populations, efforts to maintain connectivity, by installing directional fencing to guide animals to crossing structures or identifying key crossing locations and culverts for maintenance, are valuable management actions that may be undertaken by regional stakeholders.

Acknowledgements

I thank T. Diamond and A. Snyder of Pathways for Wildlife for access to samples, as well as D. Johnston of California Department of Fish & Wildlife, and the Santa Clara County Open Space Authority for both financial and logistical support. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1106400.
References
Elith, Leathwick J. 2009. Species distribution models: ecological explanation and prediction
McRae B, Dickson B, Keitt T, Shah V. 2008. Using circuit theory to model connectivity in
McRae, B, Kavanagh D. 2011. Linkage Mapper Connectivity Analysis Software. The Nature
Conservancy, Seattle WA.
Merenlender A, Reed S, Heise K. 2009. Exurban development influences woodland bird
Miller MP. 2005. Alleles In Space (AIS): Computer software for the joint analysis of
Newbold T et al. 2015. Global effects of land use on local terrestrial biodiversity. Nature
520:45–50.
Nilsson C, Reidy CA, Dynesius M, Revenga C. 2005. Large River Systems Fragmentation and
Nuñez T, Lawler JJ, McRae BH, Pierce DJ, Krosby MB, Kavanagh DM, Singleton PH,
Tewksbury JJ. 2013. Connectivity planning to address climate change. Conservation Biology
27:407–16.
O’Brien D, Manseau M, Fall A, Fortin MJ. 2006. Testing the importance of spatial configuration
of winter habitat for woodland caribou: An application of graph theory. Biological
Conservation 130:70–83.
Öckinger E, Dannestam Á, Smith H. 2009. The importance of fragmentation and habitat quality
Penrod, K., P. E. Garding, C. Paulman, P. Beier, S. Weiss, N. Schaefer, R. Branciforte and K.
Prugh L, Hodges KE, Sinclair AR, Brashares JS. 2008. Effect of habitat area and isolation on
fragmented animal populations. Proceedings of the National Academy of Sciences of the
United States of America 105:20770–20775.
Radeloff VC, Hammer RB, Stewart SI, Fried JS, Holcomb SS, McKeefry JF. 2005. The
1303.


Figures

Figure 1
A map of the study area and the location of the 15 transects sampled in Coyote Valley California, USA. The study area was a 2,914 ha corridor of relatively undeveloped land between the Santa Cruz Mountains to the west and the Diablo Mountains to the east, with the rapidly developing San Jose Metropolitan area encroaching from the north. (B) The study area overlaid with 15 transects sampled for ground squirrel scats from 2014 to 2015. Coordinates for the center of study area are: 37°11’ N, 121°42’W.
Figure 2
Distribution of landscape permeability values across the study area.
Figure 3

(A) Interpolated genetic distance between transect locations, overlaid with primary and secondary roads, as well as smaller streets. Positive genetic distances (in warmer colors) are indicative of increased genetic separation between transect locations. (B) Circuitscape resistance between transect locations, overlaid with primary and secondary roads, as well as smaller streets. Greater resistance (in warmer colors) is indicative of decreased connectivity between transect locations. (C) Circuitscape resistance overlaid with least cost paths, scaled by path length.
Figure 4
Haplotype (h1 – h10) diversity and structure based on 615 base pairs of *O. beecheyi* cytochrome b mtDNA (A) Haplotype diversity at the 15 transects. The size of the transect markers is proportional to number of individuals at each transect. Ten unique haplotypes were detected in the study area. Orange haplotypes were detected in east populations only, teal haplotypes were detected in west populations only, and gray haplotypes were detected in both east and west populations. (B) Haplotype network for east and west populations. The size of the haplotype circles is proportional to their relative frequency in the sample. Orange area represents individuals found in the east population, and teal area represents individuals of the haplotype found in the west population.
Table 1
Genetic structuring for 15 populations (transects) of *Otospermophilus beecheyi*, estimated using 615 base pairs of cytochrome b mtDNA.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Sum of squares</th>
<th>Degrees of freedom</th>
<th>Variance</th>
<th>% Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among populations</td>
<td>1925.53</td>
<td>1</td>
<td>59.64207</td>
<td>41.43</td>
</tr>
<tr>
<td>Within populations</td>
<td>5059.293</td>
<td>60</td>
<td>84.32154</td>
<td>58.57</td>
</tr>
<tr>
<td>Total</td>
<td>6984.823</td>
<td>61</td>
<td>143.96361</td>
<td></td>
</tr>
<tr>
<td>Fixation index FST</td>
<td></td>
<td></td>
<td>0.41429</td>
<td></td>
</tr>
</tbody>
</table>

Table 2
Genetic diversity parameters for *Otospermophilus beecheyi*, estimated using 615 base pairs of cytochrome b mtDNA.

<table>
<thead>
<tr>
<th>Genetic diversity index</th>
<th>West (n=32)</th>
<th>East (n=29)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Number of sequences</em> N</td>
<td>32</td>
<td>29</td>
</tr>
<tr>
<td><em>Number of haplotypes</em> H</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td><em>Nucleotide diversity</em> (\pi)</td>
<td>0.322447</td>
<td>0.219096</td>
</tr>
<tr>
<td><em>Tajima's D</em> D</td>
<td>2.95546</td>
<td>0.6387</td>
</tr>
</tbody>
</table>
Conclusion

This dissertation presents three studies of habitat connectivity that quantify the influence of land use and habitat fragmentation on structural and functional connectivity. The results provide guidance for conservation practitioners working to maintain and restore habitat connectivity for species conservation. Habitat corridors are commonly identified based on landscape permeability as estimated by levels of naturalness across the landscape. This approach was compared to detailed occurrence data for pumas (Puma concolor) to examine how well these more general approaches reflect how pumas, as species widely used for corridor planning, move through the landscape. Pumas were observed to readily use moderately permeable landscapes as estimated by the more general model, and rarely were detected in the most heavily disturbed areas. However, the general model over predicted habitat connectivity for the pumas which were not frequently detected in the most natural areas mostly likely to due to the steep terrain associated with these undisturbed sites. Mapping the level of landscape permeability that surrounds the built environment, as measured by distance to roads and housing density, offers a spatially explicit way to identify areas important wildlife movement, and provides a tool to help prioritize habitat corridors for biodiversity conservation across fragmented landscapes.

Species distribution modeling is another readily used method for identifying priority linkages for wildlife conservation. However, some of the more common approaches do not consider the influence of land use type on habitat suitability which can vary among taxa. I provide strong evidence for the inclusion of specific human land use variables affecting species distribution model accuracy for gray fox (Urocyon cinereoargenteus). By comparing the predictive performance and distribution estimates of a Maxent model, it was shown that including human land use variables improved model accuracy, as measured by AUC. Further, omitting human land use likely over predicts the distribution of suitable habitat for the gray fox. The model evaluation included citizen scientist collected observations as input, showing that citizen science data can be informative in distribution models, improving spatial coverage and providing better insight into the important drivers of species distributions. The results of this model evaluation show the importance of considering human land use as a predictor when creating species distribution models.

Finally, a combination of molecular and spatial analyses was used to assess whether roads were a barrier to functional connectivity for ground squirrels (Otospermophilus beecheyi). There is strong evidence of roads impacting functional connectivity across Coyote Valley, as demonstrated by the greater genetic similarity among populations on the same side of roads than between populations separated by major roads, in particular, a large highway. In addition to showing the similarity in structural and functional connectivity estimates, these results indicate that habitat fragmentation by roads is a barrier to gene flow in ground squirrels, resulting in quantitative differences in genetic structure and diversity between populations. This analysis shows the unique insights that can result from considering both structural connectivity associated with human development and functional connectivity at the molecular level.

The research also exposed three critical gaps that merit further investigation. First, future connectivity analyses could benefit from using a multi-taxa approach to habitat connectivity on multiple trophic levels. Second, given the increasing availability of presence only location data, further work is needed to evaluate and correct for bias in these data; including comparing model predictions that rely on presence/absence data with presence only data. And finally, including
information on roads as barriers to connectivity is informative. More information is needed to estimate the level of movement across and under roads by multiple taxa and to what extent improving road permeability (e.g. overpasses and underpasses) might mitigate the effects of roads on functional connectivity.

Multi-taxa approach to habitat connectivity

The three studies presented here evaluated structural and functional connectivity using a single species. The response of a sole focal species to habitat fragmentation cannot fully capture the needs of other integral species in the landscape because the effects of human land use are taxa specific, with the specific aspects of land use and their relative importance varying by species (Nogeire et al. 2013; Torres et al. 2016). While pumas (Wilmers et al. 2013; Smith et al. 2015; Benson et al. 2016) and gray foxes (Temple et al. 2010; Cooper et al. 2012; Kapfer & Kirk 2012) may change their overall activity levels and hunting patterns when moving through human dominated habitats, avoidance of development use may differ for other carnivores, herbivores, or more vagile species like birds. To explore the importance of taxonomic differences between the species under consideration – such as dissimilar sensitivity or scale of response to environmental variables – the model predictions from the landscape permeability and MaxEnt species distribution models could be compared to occurrence data for other taxa. Specifically, subsequent species distribution model analyses that incorporate land use could be particularly important for species showing a range in tolerance of the human development footprint; land use may still be an important predictor for mesocarnivores known to be more sensitive to (bobcat, Lynx rufus) or more tolerant of (raccoon, Procyon lotor) human presence. In addition to evaluating model predictions for different species, another next step would be to look at multiple trophic levels. For example, in Coyote Valley, a novel permutation would be to include a functional connectivity analysis for species that prey upon ground squirrels, like mesocarnivores, snakes, or raptors.

Evaluating and correcting for bias in presence only data

There is a growing body of verified, research grade presence only data available that have the potential to advance biodiversity research. Specifically, species locations recorded by citizen scientists in the form of observations on public websites like iNaturalist and eBird. Citizen scientist collected data have been shown to be of comparable accuracy to that of expert observations (Crall et al. 2011b; Nagy et al. 2012; Jordan et al. 2012; Zapponi et al. 2015), likely because many citizen science volunteers already have topical expertise or education in biology (Silvertown 2009), characteristics that suggest they may be as qualified and capable as professionals. Despite the ability of citizen scientists to collect and accurately identify species observations, presence only data will likely continue to suffer from sampling biases associated with observations being preferentially located proximally to roads and human habitation. Despite these limitations, because the use of citizen scientist collected data can increase spatial coverage and provide insight into important drivers of species distributions in the absence of systematic survey data, further investigation that could provide best practices in accounting for sampling bias is warranted.
Roads as barriers to connectivity

Road ecology is a relatively new field, with steady growth in the number of journal articles, books, conferences, and “best practice” guidelines since the publication of *Road ecology: science and solutions* in 2003 (Forman 2003). Since 2003, numerous studies have shown the negative effects of traffic infrastructure on a variety of wildlife. A systematic review of published literature showed that in results from 79 studies, covering 131 species and 30 species groups, negative effects of roads on animal abundance outnumbered the positive effects by a factor of 5 (114 negative, 22 positive, 56 no effect) (Fahrig & Rytwinski 2009).

Recent research has begun investigating the utility of road crossing structures, in the form of overpasses, culverts, and bridges, in facilitating animal movement across roads. However, crossing structures only allow wildlife movement when they accommodate the needs of the animals that use them. For example, very large mammals like grizzly and black bears require large overpasses that span roads and freeways for successful movement across highways (Sawaya et al. 2014). Culverts and below-road passages can be sufficient for animals of a variety of body sizes including coyotes and bobcats (Alonso et al. 2014). Future research could quantify the environmental and human land use variables correlated with wildlife use or avoidance of specific crossing structure types. Because crossing structures may also span waterways, it is possible that their use by wildlife could be associated with riparian zones, water availability, and, thus, exhibit seasonal trends.

A related avenue that would advance road ecology research is to refine the way roads are modeled in connectivity analyses. Many connectivity networks are being mapped using traffic infrastructure databases (Theobald et al. 2012; Theobald 2013, 2014), yet few (if any) of these mapping projects include the connectivity provided by crossing structures. Roads are typically defined as linear features with a fixed barrier effect. However, given the connectivity provided by road crossing structures, this simplistic definition is not accurate. One way to approach this gap is to conduct a connectivity analysis that uses a road network with known, functioning crossing structures. The roads could be modeled using the traditional “linear barrier” method and as a linear feature that affords connectivity at crossing structures, and model predictions then tested using camera traps set up to record crossing structure use. I anticipate the connectivity of the crossing structure would vary by species, as well as by structural features like dimension or substrate type.
References

68

