2013

Modeling connectivity in landscape genetics: applications, optimization and assessing uncertainty

Brian Kevin Hand
University of Montana, Missoula

MODELING CONNECTIVITY IN LANDSCAPE GENETICS: APPLICATIONS, OPTIMIZATION AND ASSESSING UNCERTAINITY

By

BRIAN KEVIN HAND

B.A., Astronomy-Physics, University of Montana, Missoula, MT, 2004
M.S., Computer Science, University of Montana, Missoula, MT, 2010

Dissertation

presented in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in Interdisciplinary Studies

The University of Montana
Missoula, MT

Official Graduation Date October 2013

Approved by:

Sandy Ross, Dean of The Graduate School
Graduate School

Dr. Douglas W. Raiford, Co-Chair
Computer Science

Dr. Erin Landguth, Co-Chair
Division of Biological Sciences

Dr. Gordon Luikart
Flathead Lake Biological Station and Division of Biological Sciences

Dr. Winsor Lowe
Wildlife Biology

Dr. Jon Graham
Mathematical Sciences

Mr. Joseph Glassy
Member-at-Large

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Hand, Brian K, Ph.D., Fall 2013 Computational Biology

Modeling Connectivity in Landscape Genetics: Applications, Optimization and Assessing Uncertainty

Co-Chairperson: Dr. Douglas W. Raiford

Co-Chairperson: Dr. Erin Landguth

Connectivity modeling and corridor identification are an essential part of landscape genetics and important tools for the future of conservation biology. The previous decade has shown a steadily increasing interest and rise in publications in landscape genetics. This enthusiasm has led to advances in the methods and theoretical background of the field; however, there remain important, yet unresolved, challenges. Many of these are related to validation and uncertainty testing for resistance surfaces (hypotheses of connectivity). These fundamental issues need to be addressed before landscape genetics can gain the full recognition of a scientific discipline such as population genetics or landscape ecology. The results herein not only describe the application of traditional landscape genetic techniques to empirical data, but also explore two new major approaches to improving connectivity modeling and corridor identification. In the first new approach, general theory is advanced using resistant kernel modeling by assessing a wide range of potential resistance surfaces to broadly model species distribution, connectivity, and response to habitat fragmentation and loss. Resistant kernel models allow generality across several species based on abiotic (human footprint) and life-history traits (dispersal ability and population size) for the entire Western United States. The second approach is to introduce a genetic algorithm for optimizing the process of resistance map fitting to empirical data. Optimization has three benefits. The first is removing the potential bias of expert opinion. The second is making possible multi-method evaluations of model uncertainty using different statistical tests, genetic distance metrics, and connectivity models. Lastly, optimization allows one to compare a large number of models enabling sensitivity analysis testing (e.g. leave-one-out populations, loci, or individuals). Together optimization and sensitivity analysis provide better, and more consistent, identification of landscape corridors and illustrate where models fail due to sensitivity to noisy genetic data. Described herein is a more rigorous framework of resistance map fitting and testing to help alleviate drawing faulty inferences in landscape genetic studies.

Contents

List of Figures

List of Tables

Acknowledgements

Thanks to my committee co-chairs Dr. Doug W. Raiford and Dr. Erin Landguth, and my entire committee. Dr. Raiford was a great mentor and advisor during the process, and always with great enthusiasm, and an extraordinary eye for detail and an example for the benefits of hard work. I'm grateful to Dr. Landguth for introducing me into the field of Landscape Genetics, also with great enthusiasm, and a love of science that only can be described as infectious. Thanks to Dr. Gordon Luikart for his patience and mentorship in writing and rewriting and rewriting. He worked to set a great example of hard work and passion for learning. Dr. Winsor Lowe was an exemplary teacher and I thank him for his willingness to help and his great patience. Dr. Jon Graham is perhaps the only professor capable of making statistics a "fun" endeavor, though it was an endeavor made no less difficult, but very rewarding. Mr. Joe Glassy had unsurpassed knowledge in everything to do with software engineering and scientific modeling. He was a great help and example in my effort to be a better software engineer.

Thanks to my good friends: Marnie Rout, Julie Betsch, Ted Cosart, Nick McClure, and Tarun Gupta, and my collogues in the MEID program: Meredith Berthelson, Marty Kardos, Clark Kogan, Ellen Lark, and Joshua Marceau.

The duration of a Ph.D. can feel like a lifetime and coming out the other side, I have surmised that often one can barely recognize their life except through the extraordinary events that happened along the way. Let alone just finishing a Ph.D. In the process of getting my Ph.D. the changes in my life were profound, and because of them I am a better student, researcher, scientist and person. I thank those mentioned above for having a large part in shaping my life and experiences. Most of all I thank with my love, my beautiful wife Melissa, my son Gavin and my daughter Emerson for making the greatest and best changes of all in my life.

PREFACE

This dissertation contains a number of chapters that have been included in a series of submitted or in-preparation manuscripts. Chapters 2-5 have co-authors therefore the pronoun ‘we’ is used throughout those chapters. At the time of writing the material in chapter 2 has been accepted by the Journal of Fish and Wildlife Management, Chapter 3 is under review at the Journal of Biodiversity and Conservation, Chapter 4 has been submitted to Bioinformatics, and Chapter 5 will be submitted to Molecular Ecology. The beginning and ending chapters serve as an introduction and concluding thoughts on the field of landscape genetics and connectivity modeling.

Chapter 1 Literature Review and Background

1.1 Introduction

Landscape genetics was introduced in Manel et al. (2003) as a field encompassing several other disciplines including spatial statistics, landscape ecology, population genetics, and molecular biology. Much of the theoretical background of landscape genetics has emerged over the previous two decades from ideas developed largely from metapopulation theory (Levins 1969; Hanski 1998; Manel *et al.* 2010). A major goal of landscape genetics is to study the influence of ecological processes on genetic variation. This goal is often accomplished by quantifying the relationship between landscape variables, population genetic structure, and genetic variation (Storfer *et al.* 2007). This relationship is of importance because it can be used to describe or quantify the amount and impact of connectivity. Connectivity can have multiple definitions depending upon the context of use.

Taylor et al. (1993) defined landscape connectivity as the degree to which the landscape facilitates or impedes movement between resource patches. Landscape connectivity has two major components: functional and structural. Structural connectivity is the size, shape and relative location of structures on the landscape (such as hedgerows, stepping stones, etc.), and these structures are present regardless of functional connectivity (Stevens *et al.* 2006; Manel & Holderegger 2013). Functional connectivity is the response of individuals to structural connectivity (dispersal behavior, deferred costs, and mortality risks) and the patterns of dispersal resulting from those responses (Stevens *et al.* 2006). Successful dispersal (the movement of individuals between spatially discrete populations where the individual takes up long-term or permanent residence and reproduces) leads to gene flow (Lowe & Allendorf 2010). The majority of the time, one studies genetic connectivity, the impact of gene flow on evolutionary processes such as drift, mutation and selection, based on the absolute number of dispersers (Lowe & Allendorf 2010). It is important not to confuse genetic connectivity with demographic connectivity. Demographic connectivity is the impact of migration (emigration, immigration) on population growth (lambda, r) and vital rates (births and deaths). Demographic connectivity can also be thought of as a function of total immigration rates vs. local recruitment that is dependent upon local population dynamics (Lowe & Allendorf 2010).

For the purposes of landscape genetics, functional connectivity is often measured using genetic connectivity, or inferred measures of gene flow, such as genetic distances. The assumed counterpart to genetic distances (and therefore another proxy for functional connectivity) is the effective distance (Spear *et al.* 2010). Effective distance is the Euclidean distance weighted by the cumulative cost of all landscape traversed (Adriaensen *et al.* 2003; Sawyer *et al.* 2011). Effective distances are often the desired outcome of a landscape genetics study in the hopes that they best describe observed genetic structure or some other form of empirical data describing dispersal. Broadly, understanding the relationship between effective distance and spatial population genetic structure allows researchers and conservation professionals to develop methods for promoting and protecting functional connectivity.

The loss of functional connectivity, often due to habitat fragmentation, is of major concern in conservation biology because of the role it plays in increasing extinction risk (Crooks & Sanjayan 2006a; Frankham *et al.* 2010). Major genetic factors that increase extinction risk are inbreeding (and outbreeding) depression, loss of genetic diversity, the reduced ability to adapt to climate change, and mutation accumulation (Saccheri *et al.* 1998; Frankham *et al.* 2010; Allendorf *et al.* 2013). Habitat fragmentation related to anthropogenic activities is of major global concern as human populations increase (Riitters *et al.* 2000). Future climate change is also expected to greatly alter ecological conditions due to shifting vegetation types and climes (Dale *et al.* 2000). Climate change also has the potential to have fragmentation effects on climate-sensitive species that already occupy narrow climatic niches (e.g., wolverines in the Northern Rockies; Schwartz et al. 2009).

Also, functional connectivity can be important in mapping routes of disease spread. Rees et al. (2008) studied rabies transmission in raccoons (*Procyon lotor*) near the Niagara River. An individual-based spatially-explicit model was used to simulate the barrier effect of the river on gene flow, and hence, disease spread. Using genetic distances and Mantel tests (Mantel 1967) they studied different permeability thresholds to best fit modeled data to field data to predict the percentage of crossing attempts prevented by the river. Disease spread can be understood in the context of gene flow, with the implication that natural landscape barriers may be important in stopping the spread of disease in wild populations when managed properly**.**

The following will lay out the common steps in a typical landscape genetics study with an overview of how methods vary throughout. The common methods discussed herein are greatly concerned with resistance map creation. Resistance maps are hypotheses of species dispersal patterns based on weighted landscape variables suspected to be important to gene flow. The common steps include: 1) creating resistance maps from empirical data or expert opinion (Section 1.2), 2) feeding the resistance map through a connectivity model to calculate effective distances (Section 1.3), and 3) identifying the best corridor model using statistical tests like the Mantel and partial Mantel tests (Section 1.4).

1.2 Resistance Map Creation

1.2.1 Using environmental data to create resistance maps

In landscape genetic studies, variables associated with features that are hypothesized to have an impact on individual movement are often represented in rasterized maps/grids. Each raster map represents some landscape or environmental variable (e.g., elevation or land use data; Figure 1-1) that can be continuous or categorical data. Every pixel (environmental attribute) in a raster image can then be assigned a weight depending upon the underlying variable (e.g., elevation could be classified via low, medium and high elevation). Weights for each variable layer are adjusted and based on underlying gradient effects of movement, survival, abundance and reproduction (Spear *et al.* 2010). These weights represent the relative cost of animal movement from one vertex to the next or from one pixel on a raster image to the next. Finally, a resistance map is built by summing the weights associated with each variable layer for that pixel on the map. Though there has been some consideration given to functions related to path weightings,

there has not been much consideration given to the differential contribution of each environmental variable weighting when creating resistance maps (Parks *et al.* 2013).

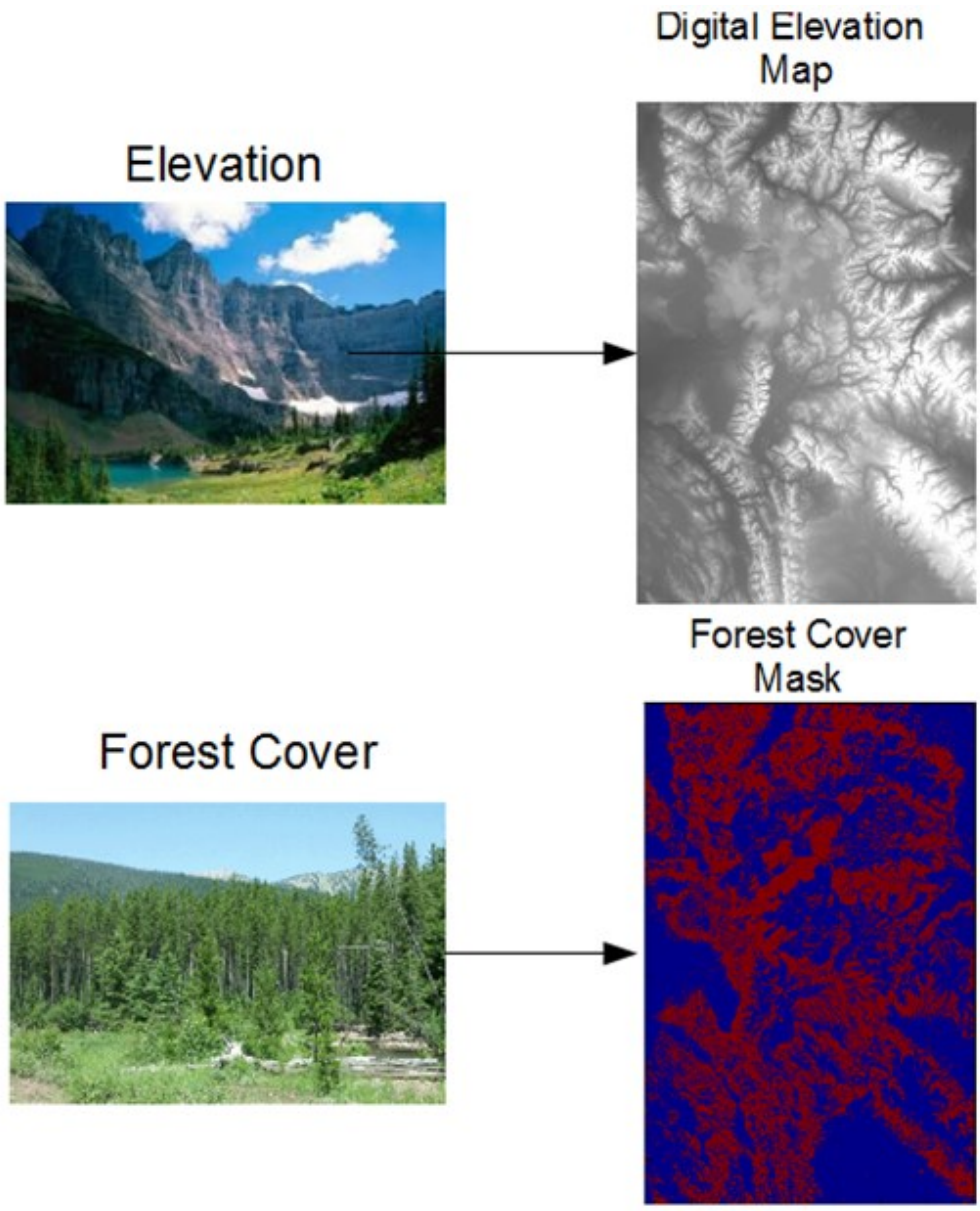

Figure 1-1 Individual landscape variables classified for hypothesized dispersal scenarios. Elevation is an example of a continuous variable type, while forest cover is a simple binary mask of forest and non-forest, reclassified from a categorical land use data set.

1.2.2 Determining Resistance Map Weights

The assignment of weights to multivariate resistance maps presents one of the great challenges currently for functional connectivity modeling (Spear *et al.* 2010; Zeller *et al.* 2012). Zeller et al. (2012) studied a collection of 96 published studies spread over several journals to explore resistance map creation. They grouped studies into three categories: 'one-stage expert', 'one-stage empirical', and 'two-stage empirical'. In the 'two-stage empirical' approach, resistance maps are based on expert opinion and confronted with empirical data. According to Zeller et al. (2012), only three studies validated the created resistance map using independent empirical data sets (e.g. the data to create a resistance map and validating it with demographic data). Following serves as an overview of the four main approaches to determining resistance map weight values: expert opinion, field data, information-theoretic, and model optimization.

1.2.3 Expert Opinion

The most common approach to determining resistance map weight values is expert opinion, nearly half of the 96 papers studied in Zeller et al. (2012) used this approach. In well studied systems or with well-known organisms, expert opinion was desired due to the lack of or difficulty in collecting field data. When expert opinion is used, a researcher might consider several different weight configurations and weight ranges in order to test several competing hypotheses (Cushman et al. 2006). Sensitivity analysis and resistance cost schemes, and the optimization of these with empirical data is often neglected (Cushman *et al.* 2006; Spear *et al.* 2010; Zeller *et al.* 2012). The expert opinion approach is biased by the assumptions of the researcher and the relative relationships between different landscape variable resistances can be greatly skewed. Weights assigned in this way are often arbitrary and can be highly non-specific in the practical information they provided for system management. As stated above, however, there is often good reason for this approach as data collection substantial enough to support resistance map creation is costly both in effort and materials.

1.2.4 Field data

The second approach to determining resistance map weight values is through the use of field data. Often these data are generated from tracking devices, such as GPS, radio-telemetry or satellites (Boyce *et al.* 2003; Epps *et al.* 2007; Driezen *et al.* 2007; Cushman & Lewis 2010), or presence-absence data (Laiolo & Tella 2006; Wang *et al.* 2008). For example, Boyce et al. (2003) used GPS data from 93 radio collared adult female elk to parameterize a resource selection function model for elk in Yellowstone National Park. During winter elk moved to lower elevations, related to snow depth, and selected landscapes with a mix of forest and open vegetation. In summer, elk traveled through forests with recent burns (12-14 years earlier), but did not use the same areas during winter. Boyce et al. (2003) also studied 4 different spatial scales, park-wide, 6-km circular plots with a buffer radius, variable buffers with mean diameter of ~8 km, and home range scale with diameters of 5.9 km. Over these scales, the relative influence of different landscape variables on habitat selection varied greatly. The conclusion of Boyce et al. (2003) was that no single scale was preferred for modeling habitat use by elk and

was dependent on the research question or management issue at hand. Scale and temporal influences can greatly vary for habitat selection, making it difficult to create resistance maps that also truthfully reflect empirical genetic structure. In addition, tracking data is expensive in both time and materials especially when the study area is large, and data can be highly variably when sample numbers are low. Due to this difficulty, there are very few studies to date that have used tracking data. For example, Zeller et al. (2012) only found 10 instances in 96 studies that use relocation data in the form of sequential points or pathway data to parameterize resistance maps.

1.2.5 Information-theoretic

Spear et al. (2010) suggested the use of network based approaches, such as information-theoretic techniques as a third approach to determining resistance map weight values. Recent work in this area has used an information-theoretic framework to study fishers (*Martes pennanti*) in Ontario, Canada (Garroway *et al.* 2011). This approach builds on previous work in Garroway et al. (2008) using a minimal edge set to describe genetic covariance. This, in turn, was used to create a network of genetic connectivity. Genetic distances were calculated as shortest-paths between vertices. Models combined variables hypothesized *a priori* to impact fisher gene flow in raster layers for each variable. Combined, the layers were run through the CIRCUITSCAPE program that calculates mean effective distances for each raster surface, and between each set of sample sites. The standardized (using a z-transformation) mean effective distances were predictor variables in linear regressions, with the response variable being the pairwise estimates of genetic distance. This model provides a substantial gain in the resistance map creation process. This method, however, does have lower explanatory power than other landscape genetic methods as it looks to achieve "parsimony over complexity" (Garroway *et al.* 2011). In this context there is good reason to avoid complexity, such as model over-fitting.

1.2.6 Model optimization

A final approach determining resistance map weight values is through optimization (constrained and full). This approach consists of varying resistance weight values then selecting resistance surfaces based on Mantel and partial Mantel *r* values (Epps *et al.* 2007; Pérez-Espona *et al.* 2008), causal modeling (Cushman *et al.* 2006) or Akaike's information criterion (AIC; Spear *et al.* 2005; Garroway *et al.* 2011). Cushman et al. (2006) used genetic data from 146 American black bear and a causal modeling framework to test 110 competing resistance map hypotheses. The causal modeling approach uses Mantel tests to test between competing hypothesis landscape models (e.g; isolation-by-distance, resistance, isolation-by-barrier). In the study, there were four landscape variables chosen and combined through expert opinion: land cover, slope, roads and elevation. In total, 110 landscape resistance hypotheses were produced with relative weight ranges from 1-10.

Cushman et al. (2006) considered relative resistances for 4 landscape variables (2 weight classes for roads and forest cover, and 1 each for elevation and slope) allowed to vary from 1-10 in increments of 1. It was also assumed that the 1-10 weight range is a sufficient range to capture the relative impact of each landscape variable. Considering

there are 1,000,000 possible combinations based on 6 variable weight classes, only 110, or roughly 0.01 % of the total possible combinations were tested. Even the most computationally efficient algorithms like Dijkstra's (1959) shortest-path algorithm can take weeks or even months to create all possible resistance maps for even a small number of landscape features. This illustrates the computationally intensive nature of creating even a few thousand resistance maps.

Two previous studies used constrained optimized methods to create resistance surfaces: Wang et al. (2009) and Shirk et al. (2010). Wang et al. (2009) applied constrained optimization to create resistance maps for the California tiger salamander. This was accomplished by choosing a very simple set of three variables and weight ranges from 1-10. They held one weight at 1 while letting the other 2 variable weights vary from 1-10 in 0.1 steps. This is satisfactory because it is the relative weight that is most important and not the absolute weight. In total, they produced 24,843 models. Gene flow measures were estimated in BayesAss (a genetic assignment tool; Wilson & Rannala 2003) to identify recent immigrant ancestry, and asymmetrical rates of gene flow between populations with confidence intervals. Assignment tests use genotypic information to identify individuals who did not originate in the subpopulation in which they were sampled in a way that is analogous to non-genetic approaches for estimating immigrants among populations (e.g. marking individuals; Lowe and Allendorf 2010). Wang et al. (2009) kept all resistance maps that produced effective distances that fell within the 95% confidence interval for all paths. They assumed all kept least-cost paths were of biological significance.

Shirk et al. (2010) studied mountain goats in the Cascade Range, Washington. They used genetic distances as the response variable to first optimize landscape variables with univariately to reach unimodal peaks. These peak values were summed and used in the multivariate optimization model. Shirk et al. (2010) held all variables constant except one. This single weight was allowed to change and then the model was re-tested with all variables to detect if model fit improved. This process was done iteratively for all variables until reaching stability for all variables. Shirk et al. (2010) used circuit theory, summarized in section 1.3.2, connectivity model in CIRCUITSCAPE (McRae 2006; Shah and McRae 2008) to calculate effective distances. They fitted models using the correlation between effective distances to three different individual measures of genetic distance. The three distances were a PCA based genetic matrix, shared alleles (D_{ps}; Bowcock *et al.*, 1994) , and Rousset's *a* (Rousset 2000).

The only full optimization of the resistance surface search procedure was implemented in the recent Graves et al. (2013) study using two non-linear search algorithms in R (R Core Team 2013). The two optimization algorithms include the Nelder Mead optimization implemented in *optim* (Nelder & Mead 1965) and the Newton-type line search algorithm implemented in *nlm* (Dennis Jr. & Schnabel 1983; Schnabel *et al.* 1986). They applied the optimization procedure to a simple surface and a wide range of simulated genetic data to test the ability of the Mantel r and causal modeling to predict the correct process for a simulated genetic pattern.

1.3 Connectivity Models

Once created, a hypothetical resistance surface must be run through a connectivity model to calculate effective distances. There are three major connectivity models used throughout the landscape genetics literature: least-cost paths, circuit theory and resistant kernel (Adriaensen *et al.* 2003; McRae 2006; Compton *et al.* 2007). The two most popular forms of connectivity models are least-cost path and circuit theory (McRae 2006). While connectivity models differ in methodology and use, it is important to point out one caveat they all share; there remains the need for *a priori* knowledge of the landscape and resistance values for each landscape feature (section 1.4). Thus, improvements at the resistance map creation level (e.g., the goal of Chapter 4) will help to improve all functional connectivity studies.

1.3.1 Least-cost paths

A least-cost path is calculated by finding the minimum effective distance between source and destination points, (e.g., using Dijkstra's (1959) single-source shortest path algorithm; Figure 1-2, panel A). Least-cost modeling works partially off the general assumption that animals have perfect, or near perfect knowledge of the landscape and therefore take the shortest path when dispersing (Cushman *et al.* 2006, 2009). For this same reason, least-cost models have been criticized for being overly simplistic (Sawyer *et al.* 2011). Recent work has suggested that more often landscape genetic studies consider least-cost corridors. Least-cost corridors are an extension of least-cost paths to take into account alternate routes that may have the same effective distance, or very similar based on some percentage difference (e.g., all paths < 10% more in cost relative to the least-cost path; Parks et al. 2013; Pinto and Keitt 2008). An example program for calculating the shortest path is the UNICOR multi-path simulator that uses parallel processing to efficiently calculate multiple least-cost paths on input resistance maps (Landguth et al. 2012).

1.3.2 Circuit Theory

By using circuit theory, McRae (2006) termed ‘isolation-by-resistance’ as the resistance distance measure that considers all possible pathways connecting population pairs (i.e., multiple pathways instead of a single pathway as in least-cost path; Figure 1-2, panel B). Circuit theory relates organism movement to the path of a random walker on a resistance landscape and considers all potential paths to contribute to gene flow (McRae 2006). The effective (resistance) distance is related to the commute time between nodes, using all possible pathways in the distance calculation (McRae *et al.* 2008). For example, in a simple system where two nodes share identical and independent pathways, the resistance distance will be half the distance of the least-cost path. Therefore, the relationship between least-cost distance and resistance distance can be related to path redundancy (McRae *et al.* 2008):

path redundancy = least-cost distance / resistance distance.

The major program for calculating circuit theory related connectivity models is CIRCUITSCAPE (http://www.circuitscape.org/; McRae 2006; Shah and McRae 2008)

1.3.3 Resistant Kernel

Resistance kernel modeling uses a dispersal kernel to predict dispersal probability distributions (Compton et al. 2007; Figure 1-2, panel C). This is done by calculating the expected density of individuals for each pixel around a source pixel (e.g., a shortest path from each source point; Cushman et al. 2010; Hand et al. 2013). These probability densities are then scaled to one using a transform function, (e.g., linear or Gaussian) so that probabilities of one are the points of origin and values close to one are neighboring vertices. For example, with a cost distance threshold of 40,000 cost units, the returned resistant kernel would be 1 at the source, 0.5 at 20,000 cost units and 0 at 40,000 cost units when using a linear scaling function. In addition, all the points of a kernel can be scaled based on a constant volume across all kernel dispersal thresholds. This volume can be thought of as a cone centered at each starting location. The scaling constant is then the value needed to keep the height of the cone equal to one based on different dispersal threshold radii. This is used to maintain constant population sizes regardless of dispersal ability. Once the expected density is calculated around each source pixel, all dispersal kernels are summed to produce a probability distribution function for the dispersal of organisms on the landscape.

A resistant kernel can be used to assess areas of high dispersal probability or connectivity. Cushman et al. (2010) used this approach to study roads and land-use scenarios using different functions of dispersal ability and breeding population size to study the impacts of habitat fragmentation on population connectivity. Resistant kernel modeling is available in version (v 2.0) of the UNICOR software (cel.dbs.umt.edu/UNICOR).

1.3.4 Preferred connectivity model

Least-cost path and circuit theory connectivity models are on the opposite sides of the spectrum in terms of model behavior and methodology (Figure 1-2; Cushman et al. 2013; Spear et al. 2010; Zeller et al. 2012). Zeller et al. (2012) found 23 instances of least-cost path and four of circuit theory used in the previous literature. They stress, however, that there is no preferred connectivity model, and McRae and Beier (2007) concluded that circuit theory performed better than least-cost path, while Schwartz et al. (2009) concluded the opposite. Further, Cushman et al. (2013) suggest that circuit theory and least-cost connectivity models are complimentary rather than opposing. Combined, least-cost and circuit theory modeling can give greater insight into paths of importance and "pinch points", areas where gene flow is constricted and easily severed.

The resistant kernel approach provides a more probabilistic assessment of functional connectivity using the basics of least-cost path and is more inclusive of all paths like circuit theory. While it is calculated using least-cost path, and covers much larger regions of connectivity like CIRCUITSCAPE, the intended use of a resistant kernel differs slightly. In previous literature and in the research presented herein, it has

been used as an exploration tool in order to discover the probability distribution for animals occupying different ranges of habitat. It can, therefore, assign areas of high habitat occupation or the reverse by identifying likely areas of habitat fragmentation and loss. This is in contrast to the intent of least-cost path and circuit theory models which is to predict corridors of functional connectivity.

1.4 Identifying the Best Corridor Model

The final step in many landscape genetics studies is to identify a preferred resistance model or to rank a range of varied models accordingly. Commonly, this is done by correlating a cost-distance matrix composed of effective distances (for a set of populations or individuals) to a genetic distance matrix. There can be several ways of calculating genetic distances for individuals (e.g. shared alleles [D_{ps}]; Bowcock *et al.*, 1994) or for populations (e.g. F_{ST}, $G_{ST,}$ $G'_{ST,}$ or Jost's D; Hedrick 2005; Jost 2008; Nei 1972; Wright 1951). Alternatively, genetic distances can be avoided by using allele frequencies and using associated statistical tests like Redundancy Analysis (RDA; (Legendre & Legendre 2012)) or Canonical Correspondence Analysis (CCA; Legendre and Legendre 2012). Balkenhol *et al.* (2009) gives a much more thorough treatment and analysis of various statistical tests. It is worth discussing the current literature on two very popular statistical tests (Mantel and partial Mantel tests) used widely in landscape genetics and other disciplines where spatial analysis is important.

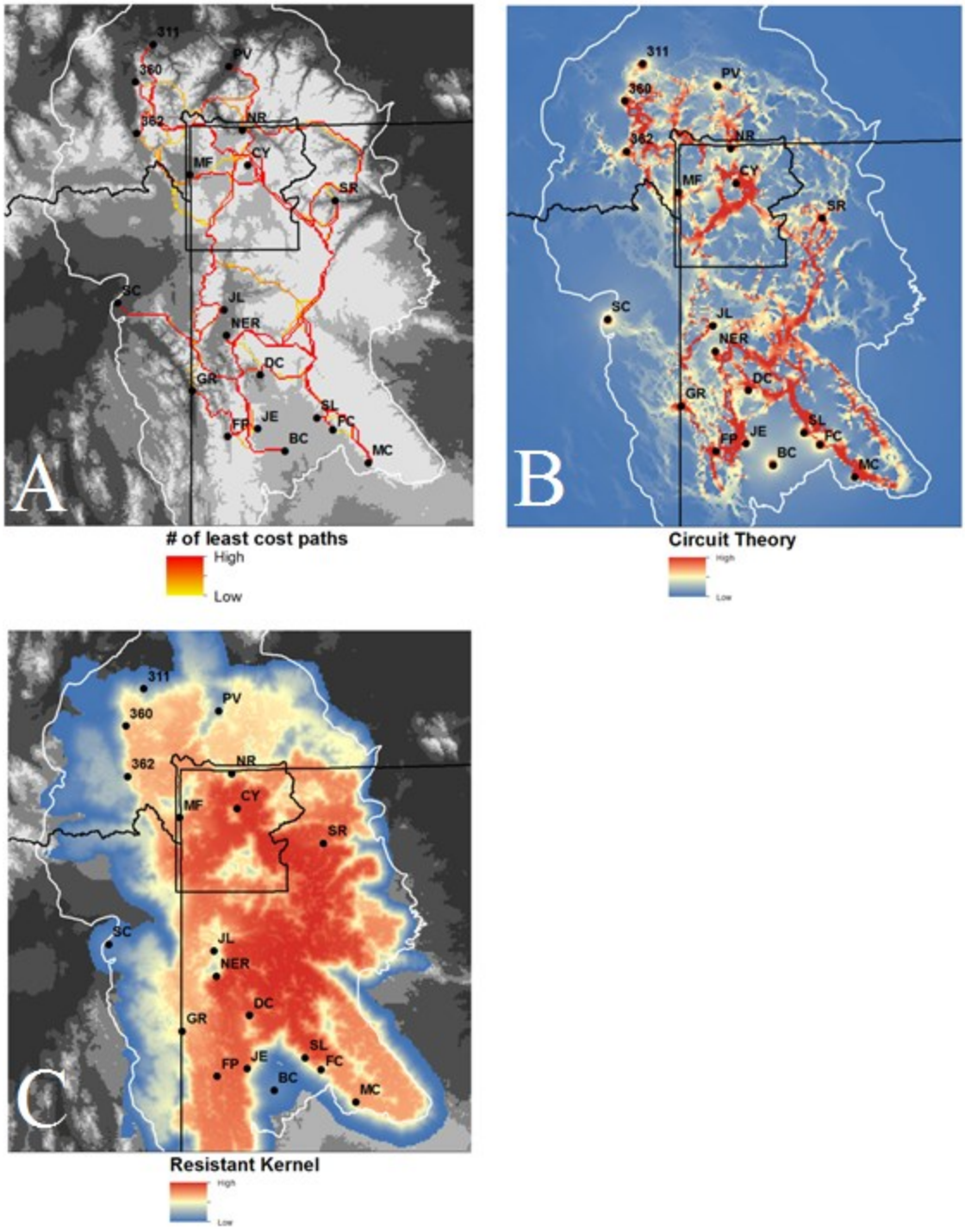

Figure 1-2 The three major types of connectivity models. Examples of the three major connectivity models, panel A) is the least-cost path, panel B) uses circuit theory and panel C) uses resistant kernel. The identical resistance map was used to create all maps. The resistant kernel was limited using a hypothetical dispersal distance for the purposes of illustration.

1.4.1 Mantel and partial Mantel Tests

A mainstay of the landscape genetics toolbox has been the Mantel and partial Mantel tests when correlating a matrix of genetic distances to one of ecological distances (Mantel 1967; Smouse et al. 1986). The partial Mantel test is an extension of the Mantel test and is often used as a means to alleviate the problems of spatial auto-correlation that is the prevalent in landscape genetics (Cushman et al. 2006; Balkenhol et al. 2009). The longstanding appeal of the Mantel and partial Mantel tests is the ease of use in comparing two distance matrices, and a permutation-based test of significance.

Several studies have recommended against using Mantel and partial Mantel tests in the recent literature (Balkenhol et al. 2009; Graves et al. 2013; Guillot and Rousset 2013). Much of the controversy stems from the fact that Mantel tests have been shown to have high type I error, often due to even moderate amounts of spatial autocorrelation (Legendre & Fortin 2010; Guillot & Rousset 2013). Unfortunately, there is confusion over what methods are more appropriate if the Mantel tests are to be avoided. Guillot and Rousset (2013) offer alternatives that are based on site-specific measures rather than distance matrices (e.g. hierachical Bayesian models and modified t-tests). Legendre and Fortin (2010) make the same recommendation because of the reductive nature of distance matrices and their tendency toward loss of information and statistical power. Balkenhol et al. (2009) suggests that canonical correspondence analysis (CCA) and multiple regression on distance matrices (MRDM) might perform better. This suggestion is problematic, nonetheless, if one were to use only MRDM, based on the later recommendation of Legendre and Fortin (2010), and Guillot and Rousset (2013) against the use of distance matrices, in general, and that MRDM is simply multivariate regression using Mantel tests. Balkenhol et al. (2009) and Guillot and Rousset (2013) have suggested alternatives to partial Mantel tests, but little has been done to implement or further explore these alternative methods. It is worth noting that Balkenhol et al. (2009) used population based simulations, whereas many landscape genetic studies employ individual genetic data.

The causal modeling framework, an extension of Mantel tests, has been debated in papers such as Graves et al. (2013) and Cushman et al. (2013). Both papers used spatially-explicit individual-based simulations of genetic structure within the same framework scenarios of cost distance. While both studies agree that casual modeling using significance tests do not perform well, Cushman et al. (2013) were able to improve results by either using a smaller alpha value (0.005) or by examining correlation values and the relative difference in rank models (prediction rates were > 75% using correlation to rank models). Cushman et al. (2013) also makes the observation that much of the problem in model selection is related to high correlation between effective distance models.

For now, there remains much confusion and debate about which tests are most appropriate. This uncertainty is likely to remain for the foreseeable future. Mantel and partial Mantel tests remain useful for the purposes of comparison and initial analysis. A majority of landscape genetics studies have used these tests. The Mantel and partial Mantel tests are likely to be useful for some situations, but must always be used with caution.

1.5 Summary

While connectivity approaches vary greatly, most landscape genetics studies strive to measure functional connectivity by identifying landscape features impacting gene flow. Landscapes can facilitate or impede gene flow and movement through natural (e.g., elevation, bodies of water, or barren climatic areas) and anthropogenic (e.g., roads, human development, or climate change) influences. In terms of conservation biology, another central goal is to identify corridors beneficial to dispersal (Crooks & Sanjayan 2006a).

The work included here explores the use of the three major types of connectivity models (least-cost path, resistant kernel and circuit theory) and their applications. The research conducted in chapter 2 uses traditional methods to investigate sex-biased gene flow in elk (*Cervus canadensis*) in the GYE. Chapter 3 extends analysis from an empirical, single species system to a theoretical approach considering several hypothetical species over the extent of the entire Western United States. For this purpose, a resistant kernel connectivity model is used to illustrate the interaction of dispersal ability and population size with habitat fragmentation and loss under varying scenarios of human impact. From the theoretical based approach described in chapter 3, chapter 4 addresses the need for a connectivity model optimization tool (GARM) that utilizes available empirical genetic data. The GARM tool uses a novel optimization technique, whereby optimal weights are discovered through the use of a genetic algorithm to relate empirical genetic data to landscape variables. Based on the results found in chapter 2, chapter 5 expands on previous work and provides an application of the GARM tool for elk in the GYE. Chapter 5 considers several methods used (both current and new) in landscape connectivity modeling to provide a more rigorous examination of the impact of variable weights on corridor model identification. Chapter 5 serves as the basis for a new and rigorous framework for the improved assessment of uncertainty in connectivity modeling.

Chapter 2 Sex-Biased Gene Flow among Elk in the Greater Yellowstone Ecosystem

2.1 Chapter Summary

Patterns of population genetic structure were quantified to help understand gene flow among elk populations across the Greater Yellowstone Ecosystem. We sequenced 596 base pairs of the mitochondrial (mt)DNA control region of 380 elk from eight populations. Analysis revealed high mtDNA variation within populations, averaging 13.0 haplotypes with high mean gene diversity (0.85). The F_{ST} from mtDNA was relatively high (0.161; P = 0.001) compared to F_{ST} for nuclear microsatellite data (0.002; P = 0.332), which suggested relatively low female gene flow among populations. The estimated ratio of male to female gene flow (m_m/m_f = 46) was among the highest reported for large mammals. Genetic distance (for mtDNA pair-wise F_{ST}) was not significantly correlated with geographic (Euclidean) distance between populations (Mantel's r = 0.274, P = 0.168). Large mtDNA genetic distances between some geographically close populations (<65 km) suggested landscape features serving as partial barriers may shape female gene flow patterns. Future research and conservation should consider the sexes separately when modeling corridors of gene flow or predicting spread of maternally transmitted diseases.

2.2 Introduction

The Greater Yellowstone Ecosystem (GYE) supports world-renowned populations of elk (*Cervus canadensis*) that provide significant visitor enjoyment and benefits to local economies through guiding, hunting, and ecotourism. Elk are the most numerous large mammal in the GYE (N ~ 50,000) and have strong effects on other species including predators and scavengers. Elk influence ecosystem characteristics and processes such as soil fertility, and vegetation production and diversity (Toweill *et al.* 2002). Elk maternal gene flow (where gene flow is defined as the exchange of alleles between populations) is important to understand because females strongly influence colonization rates, demographic vital rates, and the spread of certain diseases (Thorne *et al.* 1979; Martin *et al.* 2000).

The GYE stretches approximately 400 kilometers north-south and 300 kilometers east-west, spanning portions of Idaho, Montana, and Wyoming with elevation ranges from 1,200 to 4,200 meters for our study area (Figure 2-1). The GYE is one of a few areas where elk were not extirpated in North America by the early 1900s due to over harvest, competition with livestock, and perhaps disease (Houston 1982). Elk have not been translocated into or within the GYE. These facts make the GYE among the best (and few remaining) locations to study natural population genetic structure and patterns of gene flow (Boyce & Hayden-Wing 1979; Houston 1982; Polziehn & Strobeck 1998).

Recent work from Hedrick et al. (2013) presents an equation to estimate the ratio of male to female gene flow. The equation is useful for studies where markers for both

overall (including males and females) and maternal gene flow are available. Additionally, the equation is derived from Wright's (1951) original equations of gene flow using an island model where the major measure of genetic differentiation is F_{ST}. Using a traditional measure of gene flow makes between study comparison possible for other studies that report the correct values of overall and maternal F_{ST}.

Mitochondrial (mt)DNA is a useful marker for resolving maternal population structure and gene flow because it is a maternally inherited haploid marker (a single chromosome coming only from the mother). With relatively high rates of mutation and genetic drift, mtDNA is also more highly differentiated between populations than nuclear DNA and useful for detecting fine scale geographic structure among populations (Allendorf *et al.* 2013). Mitochondrial DNA was sequenced from eight elk populations in the GYE. Populations were defined as large groups, or collections, of individuals from a location where elk congregate, such as winter ranges with hundreds to thousands of elk. For the purpose of estimating overall gene flow nuclear microsatellite DNA markers were analyzed, with one allele inherited from each parent, for a subset of three elk populations in the GYE.

A recent equation from Hedrick et al. (2013) was used to calculate the ratio of male to female gene flow using F_{ST} values calculated from microsatellite markers and mtDNA. This was then used to compare male-to-female elk gene flow relative to other large mammals using published studies reporting the appropriate measures of gene flow for maternal and overall gene flow. Finally, we tested the hypothesis that maternal population genetic structure was related to geographic structure (distance between populations) based on the idea that genetic isolation (differentiation) was generated by geographic distance. This phenomenon of genetic structure, termed isolation-by-distance (Wright 1943) has been supported in studies of Scottish Highland red deer (*Cervus canadensis*; Pérez-Espona et al. 2009) and West Canadian white-tailed deer (*Odocoileus virginianus*; Cullingham et al. 2011). Results from the research herein suggest new and important directions for future research in maternal elk genetic structure in the GYE.

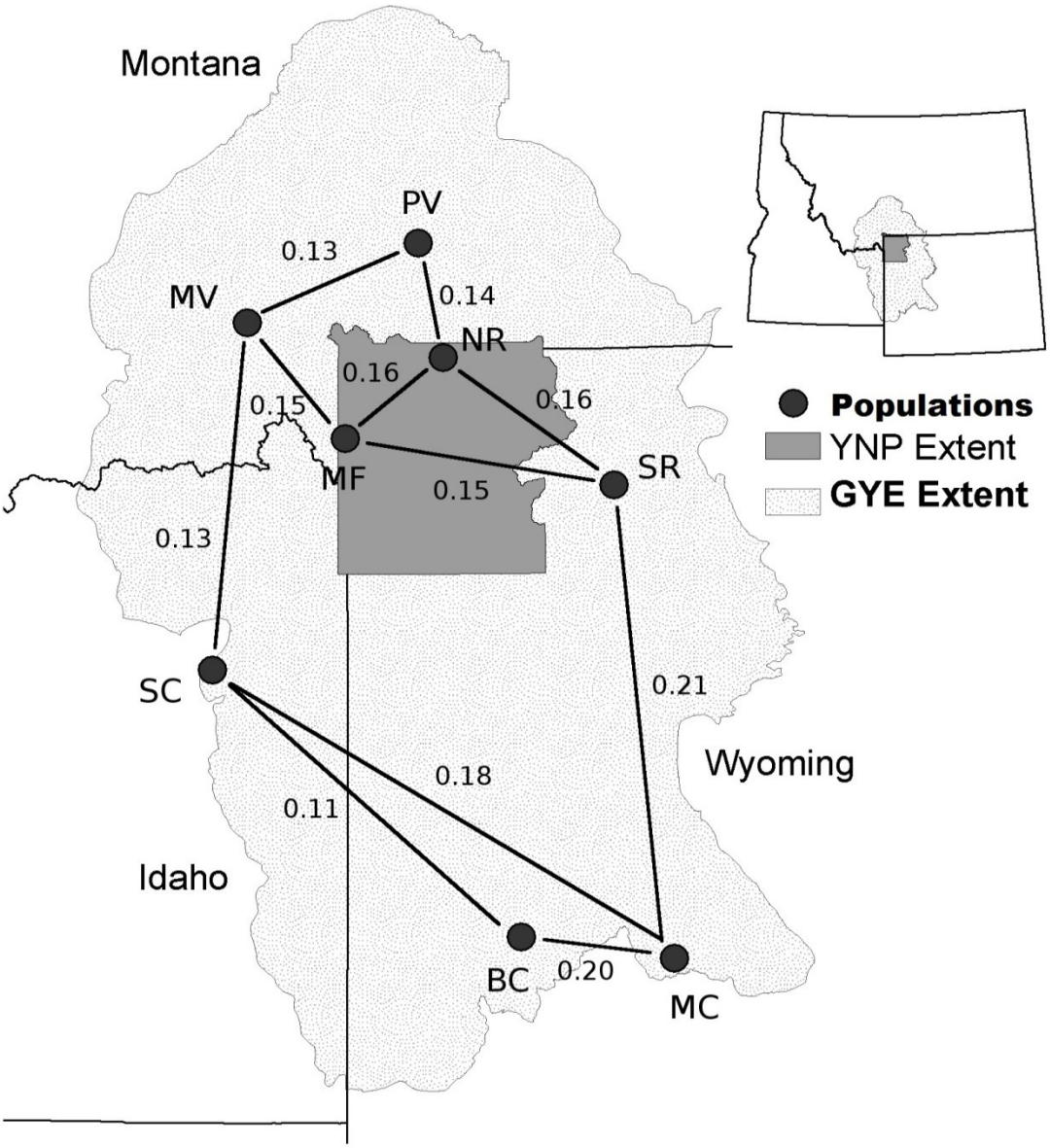

Figure 2-1 Map of the eight elk (*Cervus canadensis*) populations sampled from the Greater Yellowstone Ecosystem (GYE). Populations here are defined as large groups of individuals from a location where elk congregate, such as winter ranges with hundreds to thousands of elk. For example, the two southern most populations are feedgrounds in Wyoming where elk are fed hay in winter to keep them away from cattle and private ranches. Yellowstone National Park (YNP) is shown in gray. Numbers on the lines are pairwise mtDNA F_{ST} values for the connected populations connected by the lines. Population abbreviations are as follows: PV = Paradise Valley (Montana), MV = Madison Valley (Montana), NR = Northern Range (Yellowstone National Park), MF = Madison-Firehole (Yellowstone National Park), SR = Shoshone River (Wyoming), MC = Muddy Creek (Wyoming feedground), BC = Bench Corral (Wyoming feedground), and SC = Sand Creek (Idaho).

2.3 Methods

Blood, tissue, or fecal pellets were collected from elk in eight populations in the GYE (Figure 2-1). Samples ($n \approx 20$) were collected during multiple years from four populations (Paradise Valley, Madison Valley, Northern Range and Muddy Creek) to test for temporal stability of allele frequencies. Temporal F_{ST} values were found to be zero. All samples were collected within a 4-year time frame (one generation) to reduce potential intergenerational effects (e.g., on spatial F_{ST} estimates). Blood or tissue was collected from captured (Northern Range, Paradise Valley, and Madison-Firehole) or hunter-killed (Madison Valley, Shoshone River, Muddy Creek) animals. Fecal pellets were collected within 1–2 hours after defecation in Sand Creek and Bench Corral and from 5 of 62 individuals from Muddy Creek (Figure 2-1). To prevent repeated sampling of the same individual, feces collection was only from individuals observed defecating, small groups (5–10) of individuals that were at least 0.5–1 km apart, or from individuals with distinctive natural marking or unique ear tags, or radio collars.

Isolated DNA was taken from tissue and blood using the Qiagen QIAamp isolation kit (Chatsworth, California) and from feces using the QIAamp blood kit as described in Maudet et al. (2004). We conducted polymerase chain reaction (PCR) amplification on a 596 base pair fragment of the mtDNA control region using primers 275-294F (5'-CTCGTAGTACATAAAATCAA-3') and 990-968R (5'-ATAAGGGGGAAAAATAAGAA-3') and reaction conditions given in Polziehn and Strobeck (1998). The PCR and sequencing were conducted by the University of Washington High-Throughput Genomics Center (UW-HTGC), Department of Genome Sciences, Seattle, WA, USA (http://www.htseq.org/). Each 100mL PCR was performed on a 9600 Perki-Elmer Cetus Thermocycler using the following conditions: a 3-min denaturing step at 94°C; 30 cycles at 94°C for 15 s, 56°C for 30 s, and 72°C for 30 s.

Forward and reverse strand sequencing was conducted on all samples to ensure data quality. Each sequencing reaction was performed using approximately 8 uL of PCR product, as described in the Perkin-Elmer Dye Terminator Cycle Sequencing Ready Reaction kit. Cycle sequencing reaction parameters on a 9600 Perkin Elmer Cetus Thermocycler were as follows: denaturation at 96°C for 15 s, annealing at 50°C to for 1 s, and extension at 60°C for 4 min. Sequencing reactions were separated by electrophoresis on an ABI Prism 377 Perkin Elmer automated sequencer. The sequin file for the 380 mtDNA sequences can be found in GenBank (accession numbers: JX125702 - JX126108).

Eight microsatellite loci were genotyped for three populations: Muddy Creek, Northern Range, and Paradise Valley (Figure 2-1). The same individuals were used for both microsatellites and mtDNA sequencing from the Muddy Creek and Northern Range populations. The Paradise Valley samples for mtDNA were from hunter-killed elk, whereas microsatellite DNA samples were taken from live captured elk in the same geographical location. The captured elk provided better quality DNA from fresh blood, which typically yields more reliable microsatellite genotypes than hunter-kill samples and were obtained after the mtDNA sequences. The microsatellite DNA loci used were as follows: *BM5004*, *BM888*, *BM1009*, *BM4208*, *FCB193*, *OarkP6*, *RM006*, *BM415* (Buchanan and Crawford 1992; Kossarek et al. 1993; Bishop et al. 1994; Paterson and Crawford 2000).

All microsatellite PCRs consisted of an initial denaturation at 96°C for 15 s, annealing at 50°C to 56°C (Table 2-1) for 1 s, and extension at 60°C for 4 min. The *BM5004*, *BM888*, and *BM1009* loci were amplified together in one PCR. The *BM4208* and *FCB193* loci were amplified together in another (separate) PCR, and *OarkP6* and *RM006* in a third PCR. The *BM415* locus was amplified alone at 50°C (Table 2-1). Whole genomic DNA was extracted from elk tissue and blood samples using the QIAGEN Dneasy Tissue Kit (Qiagen, Valencia, CA, USA) according to manufacturer's instructions. The reaction volume (10 μl) contained 1.0μL DNA, 1x reaction buffer (*Applied Biosystems*), 2.0 mM $MgCl_2$, 200μM of each dNTP, 1μM reverse primer, 1μM dye-labeled forward primer, 1.5 mg/ml BSA, and 1U *Taq* polymerase (*Applied Biosystems*). The resultant products (PCR profiles) were visualized on a LI-COR DNA analyzer (LI-COR Biotechnology). A GENEPOP formatted file of microsatellite genotypes is available upon request from the corresponding author.

Sequences of mtDNA were corrected and aligned using DNAstar 5.0 software package (DNASTAR Inc., Madison, WI, USA). Sequences were double-checked visually for quality and correctness, including every polymorphic site. For quality control, 5% of all samples were randomly re-extracted and re-sequenced to monitor for potential errors; none were found. Amplification and sequencing success was relatively high even for fecal samples, where ~85% of samples yielded useable sequences, considering the relatively long mtDNA fragment amplified (596 base pairs).

Multiplex	Locus	Allele lengths	Annealing temp	Loci reference
Mix A	BM5004	130–140	56	Bishop et al. (1994)
	BM888	180–194		Bishop et al. (1994)
	BM1009[a]	268–284		Bishop et al. (1994)
Mix B	BM4208	145–157	56	Bishop et al. (1994)
	FCB193	118–146		Buchanan and Crawford (1992)
Mix C	OarkP6	161–163	54	Paterson and Crawford (2000)
	RM006	123–139		Kossarek et al. (1993)
Mix E	BM415	154–164	50	Bishop et al. (1994)

[a] amplified separately for fecal pellet (lower-quality) DNA samples

Table 2-1 Microsatellite loci and polymerase chain reaction (PCR) conditions. Listed are the sets of loci co-amplified together in each multiplex PCR, the observed range of allele lengths (in nucleotides) for each locus, PCR annealing temperatures, and the source reference for each locus (including prime sequences).

Phylogenetic relations among mtDNA haplotypes were constructed by median-joining network (Bandelt *et al.* 1999) with the program Network 4.5.1 (http://www.fluxus-engineering.com/). Reticulations were resolved through maximum parsimony criteria. Diversity indices (haplotype diversity and nucleotide diversity for each population), population pairwise F_{ST} values and a global F_{ST} value were computed using Arlequin 3.5 (Excoffier & Lischer 2010) and confirmed by GenAlEx 6.5 (Peakall & Smouse 2012). All F_{ST} values considered only the differences in haplotype frequencies using 10,000 permutations to test for statistical significance (i.e., to test if Fst > 0.0). Mantel tests (Mantel 1967) were used to evaluate correlations between genetic distances (population pairwise F_{ST} values) and pairwise geographic distances (Euclidean distance) between sample populations. Isolation-by-distance was tested in maternal genetic structure using the R package 'vegan' using 10,000 permutations to test for significance (Oksanen *et al.* 2013). Isolation-by-distance was also tested using the relationship of $F_{ST}/(1- F_{ST})$ with the natural log-transformed Euclidean distance that has found in certain cases to be more appropriate than assuming a linear relationship (Rousset 1997).

For microsatellite data, Arlequin 3.5 was used to calculate pairwise and total F_{ST} value averaged over all loci (using AMOVA tests and 10,000 permutations for significance and 10,000 replicates for confidence intervals bootstrapped over loci). We calculated per population based measures of allelic richness, observed and expected heterozygosity, deviation from Hardy-Weinberg proportions (F_{IS}), and linkage disequilibrium. All values for microsatellites were double checked and confirmed in GenAlEx 6.5.

The expected male to female gene flow ratio was calculated using equation 7(c) in Hedrick et al. (2013):

$$\frac{m_m}{m_f} = \frac{(1 - F_{st})F_{st(f)} - 2F_{st}(1 - F_{st(f)})}{2F_{st}(1 - F_{st(f)})} \quad (1)$$

In equation 1, m_m is male gene flow and m_f is female gene flow, F_{ST} is measured overall genetic differentiation for a population (considering both male and female gene flow) and $F_{ST(f)}$ is the measured genetic differentiation for females in a population using maternally inherited, mtDNA markers. Using equation 1, the ratio of male to female gene flow using global F_{ST} values was calculated for elk populations in the GYE. These estimates of gene flow assume an island model of gene flow, with assumptions that may be violated in some populations. Thus, only relative levels of gene flow can be roughly approximated (Whitlock & McCauley 1999). Equation 1 can help identify cases of sex-biased gene flow in species related to important gene flow process such as sex-biased dispersal or philopatry.

Hedrick (2005) and Jost (2008) noted the tendency of F_{ST} to be lower than expected for populations with high gene diversity (or high heterozygosity). Therefore, as an alternative analysis of male versus female gene flow, the standardized genetic differentiation measure G''_{ST} was computed to remove potential bias due to relatively high haplotype (gene) diversity for mtDNA (Meirmans & Hedrick 2011). The SMOGD program was used to calculate the value of global maternal gene flow ($G''_{ST(f)}$; Crawford 2010). Confidence intervals are often calculated from bootstrapping over loci in programs such as GenAlEx and Arlequin. As mtDNA is treated as a single locus, bootstrapping for

mtDNA (single locus) is not available in most genetic programs (Fstat, Arlequin, Genepop, GenAlex, etc.). Confidence intervals for mtDNA, instead, were calculated by bootstrapping over individuals. This method of bootstrapping is available in the SMOGD program (which does not provide an F_{ST} calculation). For microsatellites, the value for global G''_{ST} was calculated in GenAlEx 6.5 (confidence intervals were arrived at by bootstrapping over loci). Both programs SMOGD and GenAlEx 6.5 provided nearly identical estimates of both global values of G''_{ST}. This analysis assumed that G''_{ST} can replace F_{ST} in equation 1 (since G''_{ST} is an analog of the original F_{ST}); however, Hedrick et al. (2013) recommend that G''_{ST} should be investigated theoretically (as they did for F_{ST}) to ensure it is appropriate to use G''_{ST} in equation 1.

2.4 Results

Samples of blood, tissue, or fecal pellets were collected from 380 elk (223 females, 19 males, 138 unknown) in eight populations (Table 2-2). For mtDNA, haplotype diversity (gene diversity) ranged between 0.75–0.91 with an average of 0.85, and 13.0 haplotypes per population. The total number of variable mtDNA nucleotide sites was 27, which defined 30 haplotypes (Table 2-2, Figure 2-2). Most substitutions were transitions except for one transversion, which is typical for mammals including ungulates (e.g., Luikart et al. 2001). The average number of differences between haplotypes was 3.0 nucleotide sites and the mean number of variable sites within populations was 14.6.

Pop.	UTM (m East)	UTM (m North)	n	Haplotypes (No. of)	Haplotype diversity (SE)	Nucleotide diversity (SE)
PV	526045	5029408	61	16	0.889 (0.022)	0.0059 (0.0034)
MV	453908	4995027	80	17	0.848 (0.031)	0.0055 (0.0032)
NR	536585	4979966	44	15	0.831 (0.049)	0.0049 (0.0029)
MF	495338	4945231	42	12	0.858 (0.035)	0.0049 (0.0029)
SR	608769	4925150	59	15	0.835 (0.036)	0.0056 (0.0032)
MC	634277	4721083	62	12	0.746 (0.049)	0.0035 (0.0022)
BC	569692	4730309	13	7	0.871 (0.067)	0.0045 (0.0029)
SC	439437	4845368	19	10	0.906 (0.040)	0.0059 (0.0035)

Table 2-2 Eight elk (*Cervus canadensis*) populations in the Greater Yellowstone Ecosystem with spatial coordinates, number of mtDNA haplotypes, and two genetic diversity estimates. All coordinates are in the Universal Transverse Mercator (UTM) NAD83 zone 12 projection. Population abbreviations are as follows: PV = Paradise Valley (Montana), MV = Madison Valley (Montana), NR = Northern Range (Yellowstone National Park), MF = Madison-Firehole (Yellowstone National Park), SR = Shoshone River (Wyoming), MC = Muddy Creek (Wyoming feedground), BC = Bench Corral (Wyoming feedground), and SC = Sand Creek (Idaho). Standard error values are in parenthesis beside some values.

For mtDNA, the global $F_{ST(f)}$ was 0.161 (P = 0.001). Pairwise F_{ST} between populations ranged between 0.103–0.213 (P <0.001 for all pairwise comparisons, Table 2-3). No significant correlation was found between geographic distance and mtDNA genetic population pairwise distances using a Mantel test (r = 0.274, P = 0.168). Similarly, a second Mantel test using the relationship of F_{ST} /(1- F_{ST}) against the natural log-transformed Euclidean distance was not significant (r = 0.202, P = 0.228). Thus, there was no evidence of geographic isolation-by-distance among populations.

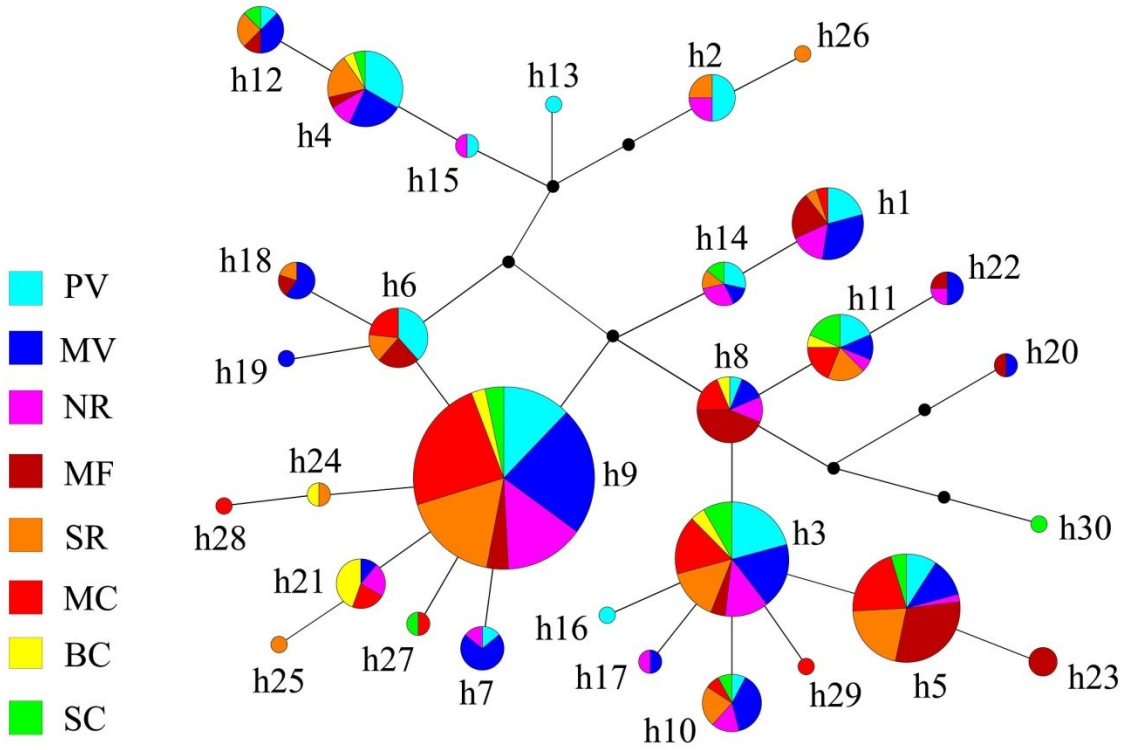

Figure 2-2 Phylogenetic network of mtDNA haplotypes (h1–h30) for eight elk (*Cervus canadensis*) populations in the Greater Yellowstone Ecosystem. Size of circle is proportional to the frequency of the haplotype. Population abbreviations are as follows: PV = Paradise Valley (Montana), MV = Madison Valley (Montana), NR = Northern Range (Yellowstone National Park); MF = Madison-Firehole (Yellowstone National Park), SR = Shoshone River (Wyoming); MC = Muddy Creek (Wyoming feedground), BC = Bench Corral (Wyoming feedground), and SC = Sand Creek (Idaho).

	PV	MV	NR	MF	SR	MC	BC	SC
PV	-							
MV	0.132	-						
NR	0.139	0.160	-					
MF	0.126	0.147	0.155	-				
SR	0.138	0.158	0.167	0.154	-			
MC	0.183	0.201	0.213	0.200	0.210	-		
BC	0.118	0.142	0.151	0.136	0.149	0.203	-	
SC	0.103	0.126	0.134	0.119	0.133	0.184	0.110	-

Table 2-3 Pairwise genetic differentiation (FST) estimates for the eight elk (*Cervus canadensis*) populations in the Greater Yellowstone Ecosystem for mitochondrial (mt)DNA. All F_{ST} values were significantly greater than zero. Population abbreviations are as follows: PV = Paradise Valley (Montana), MV = Madison Valley (Montana), NR = Northern Range (Yellowstone National Park), MF = Madison-Firehole (Yellowstone National Park), SR = Shoshone River (Wyoming), MC = Muddy Creek (Wyoming feedground), BC = Bench Corral (Wyoming feedground), and SC = Sand Creek (Idaho).

For microsatellite loci, mean heterozygosity ranged between 0.56–0.62 for the study populations with an average allelic richness of 3.88 (Table 2-4). Populations were all in Hardy-Weinberg proportions, with no significant gametic disequilibrium. The global overall F_{ST} from microsatellites was 0.002 (95% CI: 0.000–0.011; $P = 0.332$) when averaged across loci. Average differentiation for population pairwise F_{ST} values was slightly higher (average $F_{ST} = 0.005$) than the global F_{ST} value for microsatellites ($F_{ST} = 0.002$; Table 2-5). The estimated ratio of male to female gene flow was $m_m/m_f =$ 46 using our global F_{ST} values derived from mtDNA ($F_{ST(f)} = 0.161$) and from microsatellites ($F_{ST} = 0.002$) in equation 1.

The high gene diversity observed for mtDNA within populations could potentially be problematic when using F_{ST} as a measure of genetic differentiation between populations. In addition to F_{ST}, G''_{ST} (an analog of F_{ST} that corrects for potential bias caused by very high gene diversity within populations) was computed to provide a thorough evaluation of the ratio of male to female gene flow (Hedrick 2005; Meirmans & Hedrick 2011). Global maternal gene flow ($G''_{ST(f)}$) for mtDNA was equal to 0.277 (95% CI: 0.153–0.430). For microsatellites the global gene flow ($G''st$) value was 0.005 (95% CI: 0.000–0.030; $P = 0.331$). Again, the ratio of male to female gene flow was found to be $m_m/m_f = 37$, which was lower than $m_m/m_f = 46$ produced from using $F_{ST.}$ To calculate a lower bound on the ratio of male to female gene flow, G''_{ST} was used to calculate confidence intervals, taking the maximum for global overall gene flow (0.03) and the minimum interval for global female gene flow (0.153), which gave a value of 2.

Pop.	n	Allelic richness (SE)	Observed Heterozygosity (SE)	Expected Heterozygosity (SE)	F_{IS} (SE)
NR	20	3.75 (0.491)	0.617 (0.027)	0.569 (0.038)	-0.161 (0.052)
PV	20	4.13 (0.693)	0.556 (0.056)	0.596 (0.033)	0.066 (0.086)
MC	19	3.75 (0.313)	0.578 (0.064)	0.567 (0.042)	-0.053 (0.074)

Table 2-4 Characteristics of microsatellite variation from three elk (*Cervus canadensis*) populations using eight loci. Values include the number of samples per population, allelic richness, observed and expected heterozygosity, and deviation from Hardy-Weinberg proportions (F_{IS}). Population abbreviations are as follows: PV = Paradise Valley (Montana), MC = Muddy Creek (Wyoming feedground), and NR = Northern Range (Yellowstone National Park). Standard error values are in parenthesis beside some values.

	PV	MC	NR
PV	--		
MC	0.011 (0.120)	--	
NR	0.000 (0.907)	0.004 (0.297)	--

Table 2-5 Pairwise genetic differentiation estimates (FST) for three elk (*Cervus canadensis*) populations in the Greater Yellowstone Ecosystem using eight microsatellite loci. All p-values are in parentheses and are nonsignificant. Population abbreviations are as follows: PV = Paradise Valley (Montana), MC = Muddy Creek (Wyoming feedground), and NR = Northern Range (Yellowstone National Park).

2.5 Discussion

Elk in the GYE had a comparatively high ratio of male to female gene flow compared to findings for other large mammals in the literature, including some species in the same genus (Table 6). The global $F_{ST(f)}$ value derived from mtDNA was 81 times larger than the global overall F_{ST} value derived for microsatellites, which yielded an estimated rate of gene flow that was 46 times higher for males than females for elk in the GYE. The results from G''_{ST} (an analog of F_{ST}) suggested male-biased gene flow with non-overlapping confidence intervals for maternal and overall G''_{ST}. In comparison, Scottish Highland red deer (*Cervus canadensis*) have gene flow for males that is 13 times higher than for females (Table 6; Pérez-Espona et al. 2010; Pérez-Espona et al. 2009). Also, Yellowstone bison (*Bison bison)*, present in the same geographical region, have a ratio of male to female gene flow of 5. These comparisons highlight the rather high male-biased gene flow in GYE elk compared to other species known or suspected to experience male-

biased dispersal and female philopatry (the behavior of remaining in or near one's birthplace). In this context, dispersal is defined as the movement of individuals from their place of birth to a spatially discrete population, with permanent or long-term settlement (Lowe & Allendorf 2010). Hicks et al. (2007) also reports low F_{ST} values for elk microsatellite data (F_{ST} = 0.004; P = 0.281) gathered from tissue samples from the northern and southern portions of the GYE. Their study uses similar microsatellite markers (*BM5004*, *BM888*, *BM4208*, *FCB193*, *BM415*) and samples from a similar geographic area (~260 km between the Hicks et al. (2007) study populations as compared to ~ 300 km between our Paradise Valley and Muddy Creek populations).

Gene flow estimates from equation 1 assume equal effective population sizes (or equal variance in reproductive success) for males and females. High variance in male reproductive success can reduce effective population size for microsatellite loci and result in an increase in observed values of F_{ST} for microsatellites (due to more genetic drift compared to mtDNA; Hedrick et al. 2013). For example, if only 20% of males reproduce (e.g., due to a few males dominating reproduction), then overall effective size is reduced by approximately 50% (Allendorf *et al.* 2013), and the overall F_{ST} for microsatellites doubles from the value expected when effective population sizes are equal. In other words, if there is high variance in male reproductive success for elk, which likely exists given male dominance and harems, then the expected difference between F_{ST} values for mtDNA versus microsatellites would be smaller (all else being equal). This low male effective size suggests a greater magnitude of male-biased gene flow is needed to explain the difference in F_{ST} computed from mtDNA versus microsatellites.

Female gene flow (or genetic structure) was not significantly correlated (r = 0.274, P = 0.168) with straight (Euclidian) geographic distance between populations, when tested for genetic isolation-by-distance. This lack of correlation between geographic and genetic distance was best illustrated by populations that have high pairwise F_{ST} values, but are located close together geographically. For example, Muddy Creek and Bench Corral were separated by a small geographic distance (~65 km), but have one of the large pairwise genetic distances (F_{ST} = 0.203; Figure 2-1). The lack of isolation by geographic distance raised the question that factors (e.g., behavioral patterns of migration) other than straight (Euclidian) distance, and possibly landscape features are important in explaining female gene flow.

In summary, maternal gene flow among elk populations in the GYE was low compared to male gene flow that results in high sex-biased gene flow compared to other large mammals. This low female gene flow over distances of 50 to 325 kilometers was an intriguing result for such a mobile species. Future studies should apply a landscape genetics approach to test for effects of landscape on female gene flow because simple geographic (Euclidean) distance did not explain maternal genetic differentiation. The growing availability of the genetic data to compare relative male to female gene flow provides many exciting opportunities to develop and explore the magnitude, causes, and implications of sex-biased gene flow. When genetic data is not available, however, it is clear there should be methods to address gene flow for several species and explore a variety of scenarios. This is especially true if there is sex-biased dispersal present within species like the results found for elk. The next section will explore this alternative, more theoretical approach to modeling connectivity and gene flow.

Populations (Reference)	F_{ST}	$F_{ST(f)}$	m_m/m_f
(GYE) elk	0.002	0.161	45.9
Yellowstone bison (Halbert *et al.* 2012)	0.032	0.292	5.25
Highland red deer (Perez-Espona et al. 2010)	0.020	0.358	12.7
Canadian white-tailed deer (Cullingham et al. 2011)	0.006	0.015	0.261
Texas collared peccary (2 populations) (Cooper *et al.* 2010)	0.003	0.314	75.1
	0.003	0.861	1,030

Table 2-6 Estimated male to female gene flow ratio (mm/mf) for studies involving large mammal species calculated using equation 7(c) in Hedrick et al. (2013). Values given are the overall observed F_{ST} (considering both male and female gene flow), gene flow related to mitochondrial (mt)DNA ($F_{ST(f)}$), and the reference publication reporting F_{ST} values used for each set of calculations. Reference populations are for Greater Yellowstone Ecosystem (GYE) elk (*Cervus canadensis*), Yellowstone bison (*Bison bison),* Scottish Highland red deer (*Cervus canadensis*), West Canadian white-tailed deer (*Odocoileus virginianus*), and collared peccary (*Pecari tajacu*).

Chapter 3 The influence of dispersal ability, population size and human footprint on habitat fragmentation and loss for ecotype-associated species in the Western United States

3.1 Chapter Summary

The previous chapter explored an empirical, single species system to investigate sex-biased dispersal using traditional population genetic methods. In order to extend the analysis to several species, a more theoretical approach is necessary. This chapter focuses on such an approach to study several species across the Western United States. Quantifying the effects of landscape change on population connectivity is compounded by uncertainties about population size and distribution and a limited understanding of dispersal ability for most species. In addition, the effects of anthropogenic landscape change and sensitivity to regional climatic conditions interact to strongly affect habitat fragmentation and loss. To develop better conservation strategies and to understand the interplay between all of these factors, we simulated habitat fragmentation and loss across the Western United States for several hypothetical species expressing a range of habitat requirements and dispersal abilities. Existing landscape data simplified to cover 4 broadly inclusive biome classifications (Mixed Conifer, Grassland/Shrubland, Desert and Subalpine) that cover a majority of the region. A least-cost resistant kernel model was implemented to evaluate the relative importance of anthropogenic ("pristine", low and high human footprint scenarios) and biological factors (dispersal ability and population size) to create a combined 48 resistance scenarios for a broad range of hypothetical species throughout the Western United States. Resistance kernel modeling uses a dispersal kernel to predict dispersal probability distributions. The strength of the resistance kernel model is to assess contiguous maps of dispersal and regions for hypothetical dispersal scenarios. Habitat extent was found to be equally sensitive to dispersal ability and population size of the focal species, while habitat fragmentation is more sensitive to dispersal ability. Grassland and forest associated species are most at risk from habitat loss and fragmentation driven by human related land-use. Hypothetical simulation studies such as this can be of great value to scientists and managers in helping develop conservation theory, and evaluate spatially-explicit management scenarios. Results from this research are available for download in a web-based interactive mapping prototype useful for guiding conservation and management.

3.2 Introduction

Much of the difficulty in predicting the effects of landscape change on population connectivity (Cushman, Shirk, *et al.* 2013) is due to uncertainty about species population sizes and distributions, how different landscape features affect movement, and limited understanding of species dispersal abilities (e.g., Cushman et al. 2006). Several past

hypothetical based studies evaluated the interactive effects of population size and dispersal ability on the extent and fragmentation of connected habitat (Cushman *et al.* 2010, 2011). For example, Cushman et al. (2010) studied habitat connectivity across 200 combinations of dispersal ability and population size for vernal pool-breeding amphibians in Western Massachusetts. They found that dispersal ability had larger influences on population connectivity than did population size, while the effects of roads and human land-uses had greatest relative impact at middle population sizes and middle-to-high dispersal abilities. Similarly, (Cushman *et al.* 2011) evaluated the sensitivity of habitat connectivity for three focal species and two biome types (grassland and forestland) for the Great Plains region of the United States using an array of dispersal abilities and varying scenarios of landscape resistance.

These past studies did not distinguish between the effects of population size, dispersal ability and landscape resistance on the extent versus the fragmentation of connected habitat. It is important to distinguish between a loss in extent and fragmentation of habitat (Fahrig 2003), given that the interpretation and management response to a given change will differ substantially depending on whether it is loss or fragmentation driven (McGarigal & Cushman 2002; Fahrig 2003). Considering habitat fragmentation in terms of the change in contiguous habitat may also be important because of greatly varying life-history traits (e.g., limited gap crossing ability), and similarly, contiguous habitat has direct effect on area sensitive species (Freemark & Merriam 1986; Robbins *et al.* 1989; Cushman *et al.* 2011). Contiguous habitat may also impact species diversity as smaller patch sizes can contain fewer species than larger patches, with small patches containing a subset of species found in larger patches (Vallan 2000; Debinski & Holt 2000; Fahrig 2003). Large, contiguous patches are likely preferable to small, disconnected patches for these reasons.

This project was designed to predict and map the extent and fragmentation of connected populations of four species groups of differing habitat association across a wide range of dispersal abilities and population sizes for the full extent of the conterminous Western United States. The relative impact that roads and human land-uses had on the extent and fragmentation of these species populations was tested using three specific hypotheses:

1. Dispersal ability will have a larger influence than population size on the extent and fragmentation of connected habitat across the Western United States for a wide range of species.
2. Human footprint (roads and land-use) will have greatest effect on decreasing extent and increasing fragmentation of habitat for grassland associated species, followed in order by desert, mixed conifer forest and subalpine associated species.
3. Human footprint will have a larger impact on fragmentation of habitat than habitat loss, due to the fragmenting effect of roads and the dendritic pattern of human land-uses along transportation networks.

3.3 Methods

3.3.1 Study System

The study system covered the conterminous Western United States from 100° longitude westward (Figure 3-1). This region includes most of the federally protected land and forested areas in the conterminous United States. Also, many important wildlife preserving National parks are found in this region including the Yellowstone, Glacier, Yosemite, Redwood, and Mount Rainer National parks.

3.3.2 Vegetation Class Data

The original vegetation data were from the Mapped Atmosphere-Plant-Soil System (MAPSS; http://www.fs.fed.us/pnw/mdr/mapss/) vegetation model (Neilson 1993, 1995; Neilson & Marks 1994). MAPSS is a vegetation distribution model developed to simulate potential biosphere impacts and biosphere-atmosphere feedbacks from climate change. Output from MAPSS has been used extensively in the Intergovernmental Panel on Climate Change's (IPCC) regional and global assessments of climate change. From the MAPSS model, vegetation was reclassified from the original 10 km resolution climate data model including 62 vegetation classes (Table 3-1) into four broadly inclusive biomes; Grassland/Shrubland (GS), Mixed Conifer (MC), Desert (DE) and Sub-alpine (SA). The reclassified data were then interpolated to 1 km resolution using bilinear interpolation.

3.3.3 Road and Land Cover Data

Road and land coverages for the Western United States were taken from the Census 2000 TIGER line files (http://www.census.gov/geo/www/tiger/tiger2k/tgr2000.html) and the 2001 National Land Cover Database at http://www.mrlc.gov/nlcd.php (UA Census 2000; Homer et al. 2007). The purpose of the road and land cover layers was not to assess habitat suitability, (which was already addressed by the biome layer), but as a measure of the human footprint on resistance to dispersal. Different land-use types (e.g., agriculture and urban centers) carry respectively higher resistances (Table 3-1). Water was also considered to be higher resistance, since the focus was on terrestrial species.

Road cover was taken from a vector map of all major and minor roadways from the 2000 Census. The data were converted to a 30 m resolution raster map. Road features were reclassified (Table 3-1) and smoothed to 1 km resolution using a 50 x 50 moving window in the focal statistics toolkit in ArcGIS (Esri 2011). Land cover was bilinearly interpolated from 30 m to 1 km resolution before being combined with the roads and biomes layers to have all layers at the same grid resolution. All dataset creation and interpolation was done in the ArcGIS 10 (Esri 2011) and/or using the library 'raster' (Hijmans & van Etten 2013) in the statistical software package R (R Core Team 2013).

3.3.4 Resistant kernel Connectivity Modeling

For each scenario of biome association and human footprint impact, predictions for structural or dispersal connectivity were based on a least-cost resistant kernel approach implemented in UNICOR v2.0 (Compton *et al.* 2007; Cushman *et al.* 2010; Landguth *et al.* 2012). Unlike most corridor prediction efforts, the resistant kernel approach is spatially synoptic and provides prediction and mapping of expected dispersal rates for

every pixel in the study area extent, rather than only for a few selected "linkage zones" (Compton *et al.* 2007). Also, in resistant kernel modeling scale dependency of dispersal ability can be directly included to assess how species of different vagilities may be affected by landscape fragmentation. Resistant kernel modeling is also computationally efficient, enabling simulation and mapping across the entire Western United States for multiple species (Cushman *et al.* 2010).

All resistance scenarios below provide values for all locations in the study area, in the form of the cost of crossing that pixel. Cost-distances then refer to the cumulative cost of traveling from a source point to any other location. These cost distances are used as weights in the dispersal function, such that the expected density of dispersing individuals in a pixel is down-weighted by the cumulative cost from the source, following the least-cost route (Compton *et al.* 2007). The initial expected density was set to one for each source cell. The predicted density in each surrounding cell is predicted density relative to the maximum at a source cell. The model calculates the expected relative density of each species in each pixel around the source, given the dispersal ability of the species, the nature of the dispersal function, and the resistance of the landscape (Compton *et al.* 2007; Cushman *et al.* 2010).

The UNICOR v2.0 (Landguth *et al.* 2012) program used a scaled resistant kernel value so kernel volume was constant (which equates to a constant population size at all occupied source locations). Thus, volume was kept constant across different dispersal abilities and assured that more mobiles species were not misconstrued to have larger population sizes. Our approach was somewhat similar to Compton et al. (2007) where a normal probability density function was used as a basis of the dispersal model. Here, we assumed a linearly scaled dispersal function dependent on cost distance.

3.3.5 Modeling Scenarios

Three modeling scenarios were assessed; the null or "pristine" scenario, a low human footprint (LHF), and a high human footprint (HHF). The "pristine" scenario considered only the resistances of the biome resistance maps for each specific biome (GS, MC, DE and SA) without the influence of roads or land cover (Table 3-1; Figure 3-1,Figure 3-2, Figure 3-3, Figure 3-4). Each biome was assumed to be a natural habitat for some species and each pixel in that biome to have an associated value of one compared to other, less desirable habitats for those same species (i.e., resistance values of one refer to the easiest traversable areas on each landscape). When a species entered a non-native habitat it was assigned additional resistance penalty per pixel and based on the native habitat of the species (Table 3-1). Extreme differences in biomes were assumed to represent large resistance differences. For example, DE species have a high resistance in SA and MC biomes and vice versa. Values were based on expert opinion; which was appropriate in this case given the lack of species-specific resistance relationship data across the extent of our study and our goal to represent a broad range of plausible species responses (Table 3-1; Sawyer et al. 2011).

The LHF and HHF scenarios included roads and land cover as separate resistance maps, in addition to the biome resistance map (Table 3-1). Points were seeded at every 10 km and only where the resistance map had a value of one (i.e., indicative of where the habitat-associated species occurred). For the LHF and HHF, this was consistent with an animal choosing its preferred biome type and natural habitat without roads. For each of

the biomes in the null model starting locations (seeded points) consisted of 22,667 for GS; 6,895 for MC; 4,085 for DE; and 627 for SA. four dispersal distances (12.5km, 25km, 37.5km and 50km) were considered for each of the scenarios and biomes for a total of 48 different resistance scenarios.

3.3.6 FRAGSTATS Analysis of Modeling Scenarios

Output of the kernel surfaces maintained constant volume, meaning constant population size, that allowed for comparison across population sizes by multiplying by the appropriate constant (Cushman *et al.* 2010). Different population sizes were produced of 625, 1250, 2500, 5000, and 10000 individuals per 100 square kilometers. A binary map of presence/absence in the study area was used for further analysis in the FRAGSTATS program (McGarigal *et al.* 2013). To do this cells were considered occupied only if cell values were greater than 0.00001 (corresponding to a minimum of one individual per 10 square km).

The FRAGSTATS (McGarigal *et al.* 2013) program was used to calculate four class-level metrics to compare the scenarios in their impact on the extent and fragmentation of connected habitat: 1) the extent of connected habitat as a proportion of the total study area (PLAND), 2) number of patches of connected habitat (NP), 3) correlation length of connected habitat (CL, denoted as GYRATE_AM in FRAGSTATS), 4) the size of the largest patch as a proportion of the total study area (LPI). The PLAND metric is a quick and useful measure for the comparative measure of the amount of loss of habitant extent between different scenarios. The NP quantifies fragmentation of connected habitat in each scenario. The CL metric is the area-weighted mean radius of gyration, where the mean radius of gyration is the mean distance between each cell in a patch and the patch centroid (McGarigal *et al.* 2013). The CL gives the expected distance of travel while staying in that particular patch type, from a random starting point and moving in a random direction. Correlation length has been shown to be a strong predictor of the effects of habitat fragmentation on population connectivity (Keitt *et al.* 1997; Cushman *et al.* 2010; Cushman, Shirk, *et al.* 2013). The LPI metric provides a direct measure of the extent of the largest connected patch of habitat for each scenario.

Resistances by Cover or Biome Type	Scenario	
	LHF	HHF
Land Cover		
Natural: Perennial Ice/Snow; Barren Land; Deciduous Forest; Evergreen Forest; Mixed Forest; Scrub/Shrub; Herbaceous; Emergent Herbaceous Wetlands; Woody Wetlands	1	1
Agricultural: Hay/Pasture; Cultivated Crops; Open water	5	15
Residential: Developed; Open Space; Developed; Low Intensity	5	15
Urban: Developed; Medium Intensity; Developed; High Intensity	10	20
Road Cover		
Primary Highway With Limited Access: Interstate highways and some toll highways with interchanges.	50	80
Primary Road Without Limited Access: US highways, some state highways and county highways that connect cities and larger towns.	30	50
Secondary and Connecting Road: State highways, county highways that connect smaller towns, subdivisions, and neighborhoods.	10	20
Local, Neighborhood, and Rural Road: Local traffic, single lane.	2	5
Vehicular Trail: Roads usable only by four-wheel drive vehicles, one-lane dirt trails in rural areas.	2	5
Road with Special Characteristics: Portions of a road, intersections of a road, or the ends of a road that are parts of the vehicular highway system and have separately identifiable characteristics.	2	5
Road as Other Thoroughfare: Foot and hiking trails located on park and forest land.	2	5

Biome Resistances	GS	MC	DE	SA
Grass/Shrub				
Shrub Savanna (Deciduous Broadleaf, Mixed Warm, Mixed Cool, Evergreen Micro)	1	15	10	20
Shrubland Temperate (Conifer, Xeromorphic Conifer); Grass (Short, Mid, Tall); Grass Dry (Short, Mixed Short); Grass Prairie (Short, Tall); Grass Northern (Short, Tall, Mixed Mid, Mixed Tall); Grass Southern (Mixed Mid); Open Shrubland (No Grass)	1	15	5	20
Shrub Savanna Tropical	1	20	10	25
Chaparral	1	10	5	10
Mixed Conifer				
Forest Evergreen Needle (Tiaga, Maritime, Continental)	15	1	20	5
Tree Savanna (Mixed Cool, Mixed Warm, Evergreen Needle Maritime, Evergreen Needle Continental)	10	1	15	10
Forest Mixed Warm	15	1	20	10
Desert:				
Grassland Semi Desert; Desert Boreal	5	20	1	25
Desert (Subtropical, Tropical, Extreme)	10	20	1	25
Subalpine:				
Taiga/Tundra	20	5	25	1
Tundra	25	15	30	1
Other:				

Continuation of table from above…

Tree Savanna PJ Continental; Tree Xeric Continental Savanna PJ Maritime; Tree Savanna PJ	5	10	5	10
Shrub Savanna SubTropical Mixed	15	20	5	10
Shrubland SubTropical (Xeromorphic, Mediterranean)	20	25	5	10
Desert Temperate	20	25	5	5
Tree Savanna (Deciduous Broadleaf, Mixed Warm)	5	10	10	15
Forest (Deciduous Broadleaf, Mixed Warm, Mixed Cool)	5	10	15	20
Forest Hardwood Cool	5	10	15	20
Forest (Evergreen Broadleaf Tropical, Seasonal Tropical, Savanna Dry Tropical)	10	15	20	20
Ice	20	5	25	30

Table 3-1 Resistances by Cover or Biome Type. Cover and biome types with assigned resistance values used in the kernel resistance simulation. Numbers under scenario refer to the resistance values assigned to land cover or road cover class for the LHF = low human footprint or HHF = high human foot print scenario. Biome abbreviations are GS = grassland/shrub, MC = mixed conifer, DE = desert, SA = subalpine. The resistance values differ for each biome type and are organized by land types assumed to be native to each biome type, and a category of other land types which are not native to any biome.

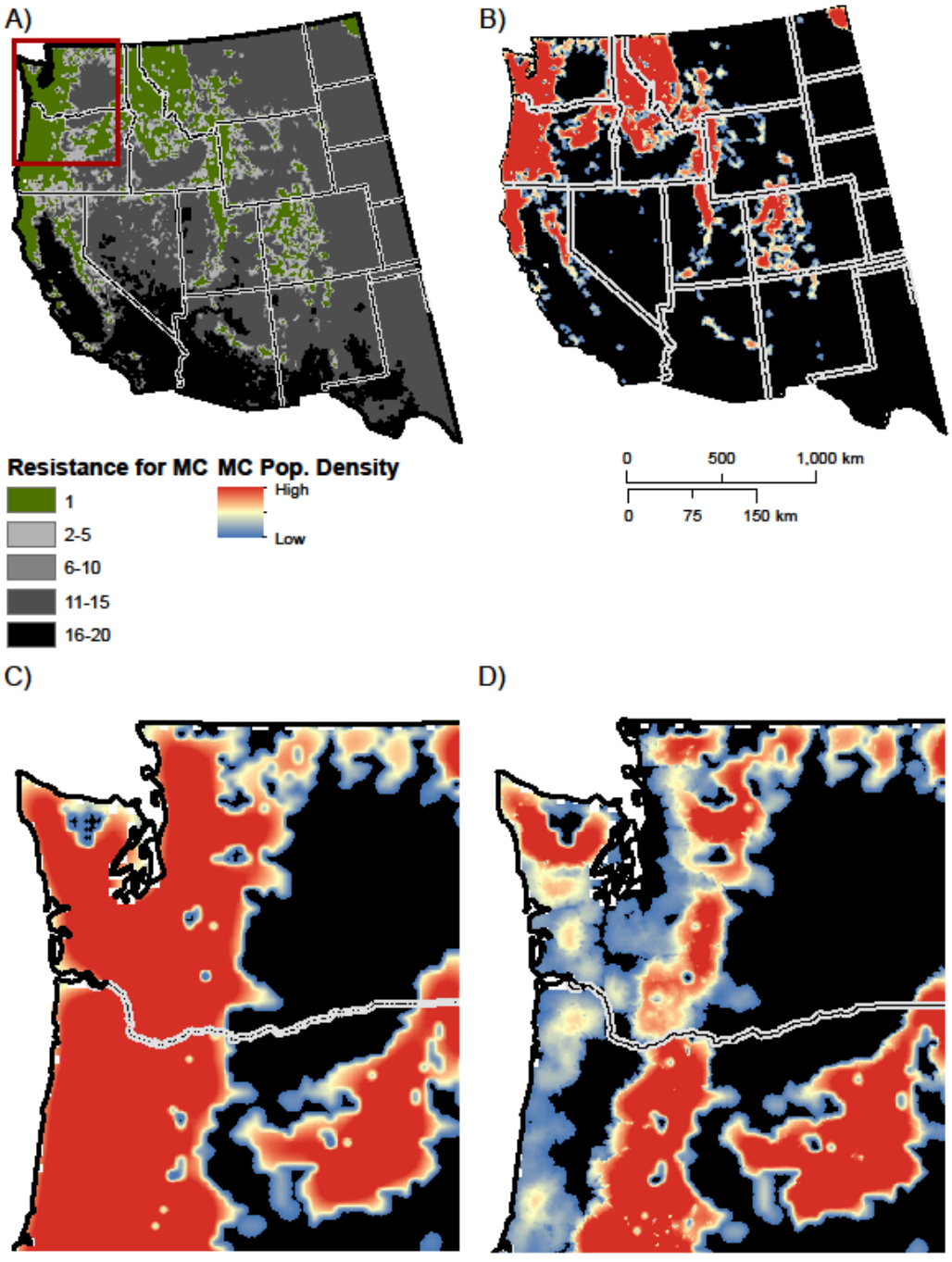

Figure 3-1 A map of the Western United States highlighting the Mixed Conifer (MC) biome. Panel A) is the base resistance map for the Western United States for MC; all resistances in green are equal to 1. Panel B) is the resistant kernel map for the entire study area using the null model showing the variation in predicted population density in blue-red. Panel C) shows an extent of the West Coast centered over Washington State, using the null or "pristine" resistance kernel model. Panel D) is the population density in the high human footprint. Population density is represented on a red-blue scale, with high population density in red. Areas of black are high underlying resistance.

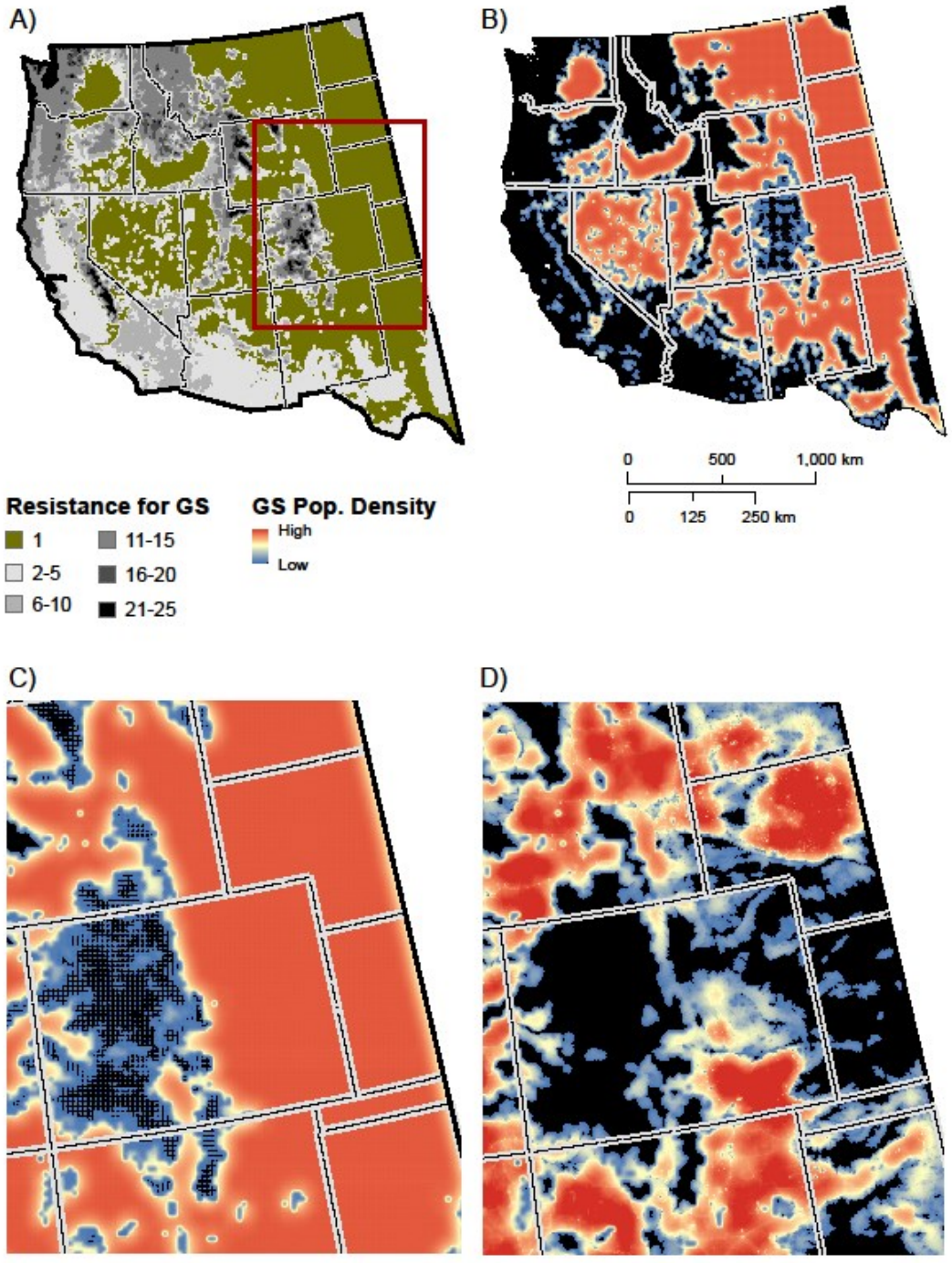

Figure 3-2 A map of the Western United States highlighting the Grassland/Shrub (GS) biome. Panel A) is the base resistance map for the Western United States for GS; all resistances in olive are equal to 1. Panel B) is the resistant kernel map for the entire study area using the null model showing the variation in predicted population density in blue-red. Panel C) shows an extent of the Midwest, using the null or "pristine" resistance kernel model. Panel D) is the population density in the high human footprint. Population density is represented on a red-blue scale, with high population density in red. Areas of black are high underlying resistance.

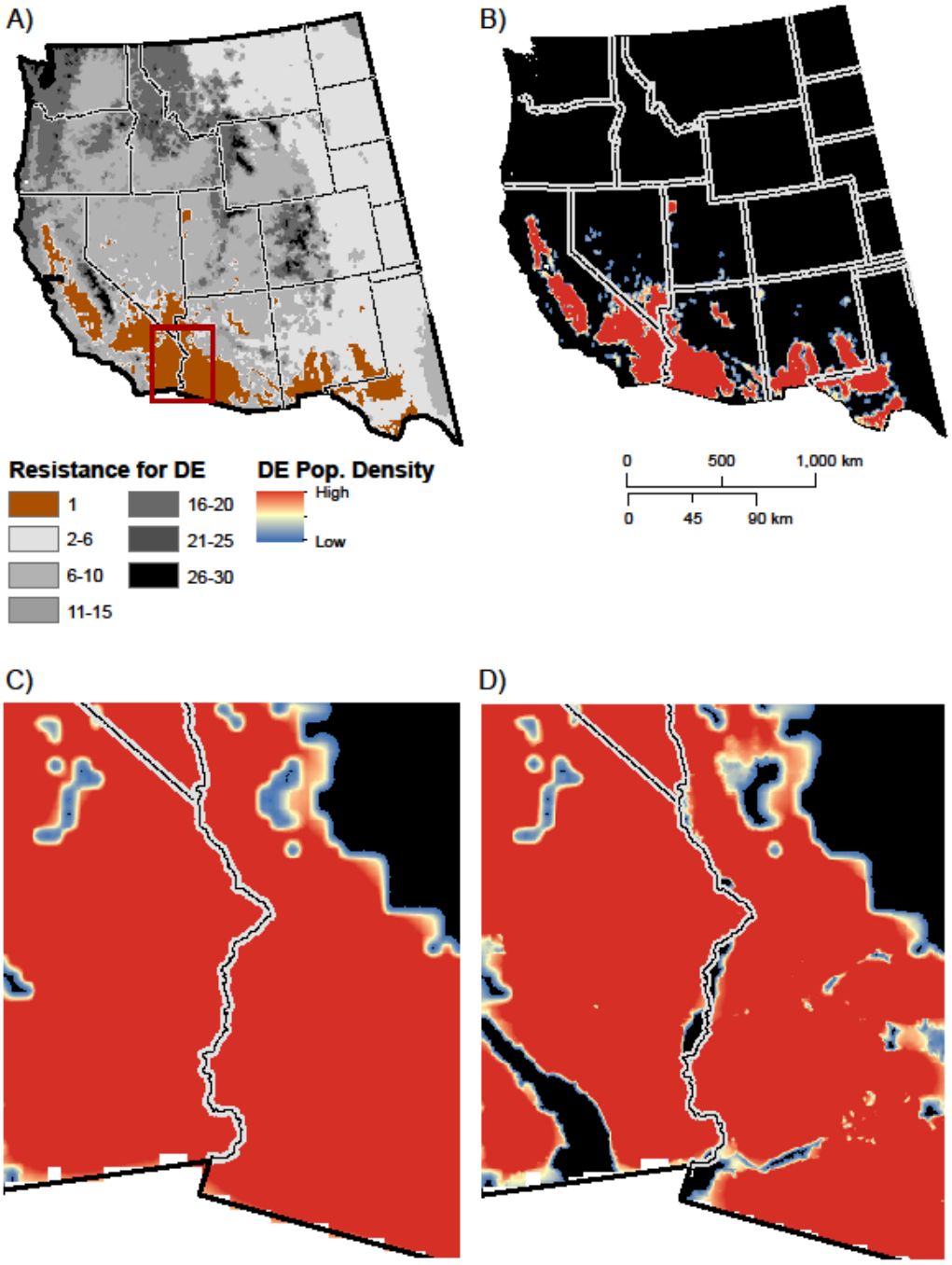

Figure 3-3 A map of the Western United States highlighting the Desert (DE) biome. Panel A) is the base resistance map for the Western United States for DE; all resistances in olive are equal to 1. Panel B) is the resistant kernel map for the entire study area using the null model showing the variation in predicted population density in blue-red. Panel C) shows an extent of Southern California and New Mexico, using the null or "pristine" resistance kernel model. Panel D) is the population density in the high human footprint. Population density is represented on a red-blue scale, with high population density in red. Areas of black are high underlying resistance.

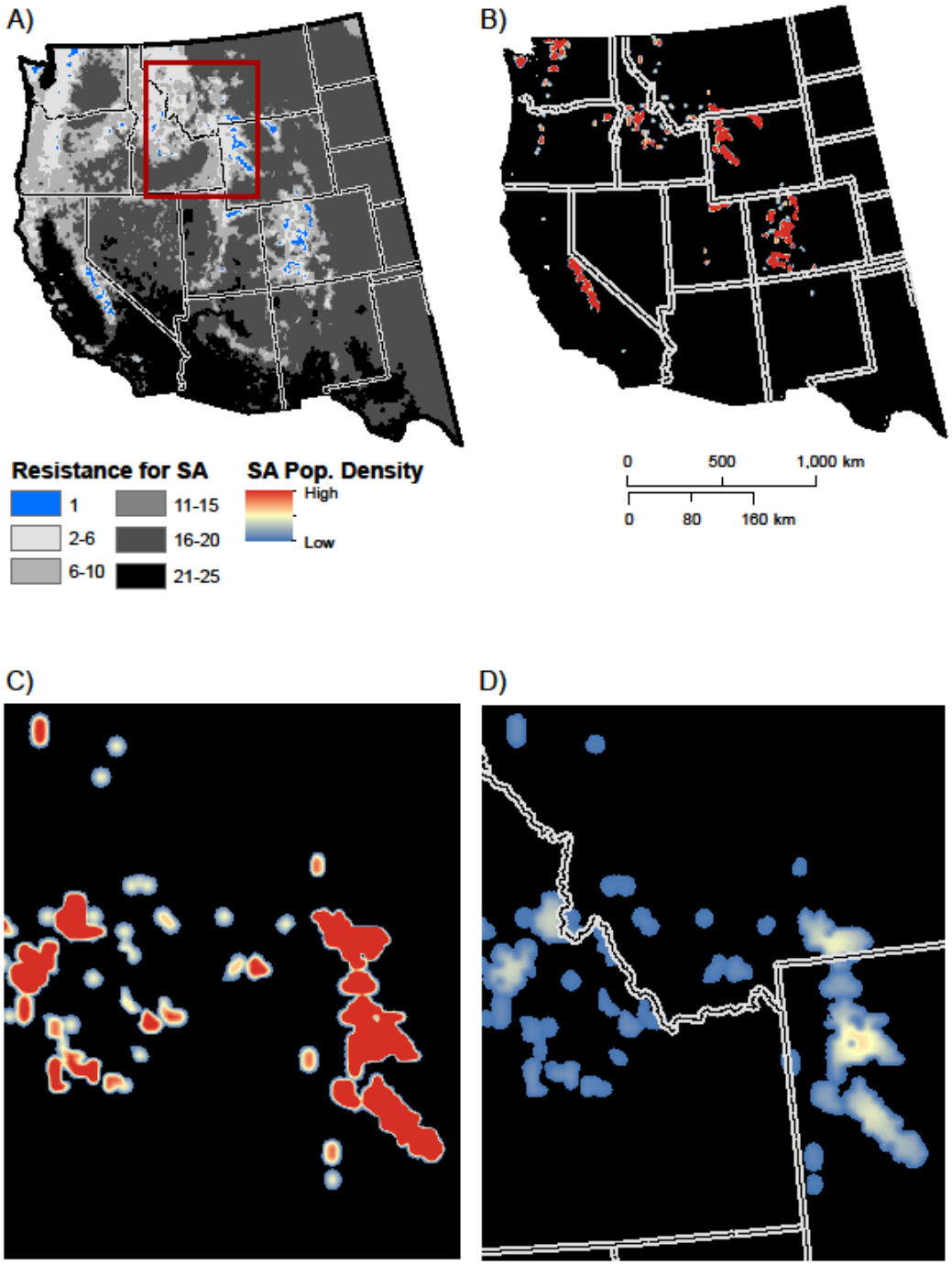

Figure 3-4 A map of the Western United States highlighting the Sub-alpine (SA) biome. Panel A) is the base resistance map for the Western United States for SA; all resistances in olive are equal to 1. Panel B) is the resistant kernel map for the entire study area using the null model showing the variation in predicted population density in blue-red. Panel C) shows an extent of the Northern Rockies, using the null or "pristine" resistance kernel model. Panel D) is the population density in the high human footprint. Population density is represented on a red-blue scale, with high population density in red. Areas of black are high underlying resistance.

3.4 Results

3.4.1 Percentage of the total land area (PLAND)

There was a non-linear trend of decreasing extent of connected habitat (PLAND) as population size and dispersal ability decreased that accelerated as both life history traits decreased (Figure 3-5; Figure 3-6; Figure 3-7; Figure 3-8, panel a; Table 3-2). PLAND tended to decrease more slowly with decreases in the other life history trait, however, when at low population size or dispersal ability. In other words, the two factors interact. This reverse in trend suggested the combination of life history traits had a compounding effect, but only up to a certain threshold where PLAND was much less responsive to changes in life history traits. The greatest change in PLAND occurred for the SA biome where the maximum value of 3.4% dropped to 1.3% at the minimum population size and high dispersal ability, a 62% decrease. For the GS, MC and DE scenarios, the decrease in the maximum value of PLAND from high to low values of dispersal ability and population size was 18%, 28 % and 22 % respectively.

3.4.2 Number of patches (NP)

Across all biome types and population sizes, dispersal ability had a much greater influence on the number isolated patches of habitat internally connected by dispersal than did population size. The increase in NP was non-linear along dispersal gradients and tended to accelerate at low (12.5 - 25 km) dispersal ability (e.g., at low population sizes NP experienced increased with decreased dispersal ability in ratios of 80:1 (GS),11:1 (MC), 9:1 (DE) and 6:1 (SA), Table 3-2; Figure 3-5; Figure 3-6; Figure 3-7; Figure 3-8, panel b). Along constant levels of dispersal ability, NP tended to peak at medium population sizes (2500) for mid-range dispersal abilities (25-37.5 km) and then greatly decreased thereafter. The resulting decrease from medium to low population sizes in NP was often quite large with as much as a 94% loss (Table 3-2) and in many cases was > 50%. The only exception to this sharp decline in NP was at the lowest dispersal ability, where NP tended to remain unchanged (GS, DE, and SA) or slightly increased (MC).

3.4.3 Correlation length (CL)

Dispersal ability had more effect on CL than did population size (Figure 3-5; Figure 3-6; Figure 3-7; Figure 3-8, panel c, Table 3-2) as the relative changes between the extremes in life history traits (high dispersal ability and population size to either low population size or low dispersal, while keeping the other variable constant) showed a larger difference along the dispersal ability gradient (Table 3-2). For all four biome types there were distinct threshold effects where correlation length of connected habitat dropped dramatically in response to changes in population size or dispersal ability, with the tendency to slow or stop for further decreases (with the exception of the SA biome that had several large decreases). Values dropped for CL at these various thresholds in a range between 7% (DE) and 31% (MC).

Percentage of total Landscape (PLAND)

		Low Pop. (625)			Med. Pop. (2,500)			High Pop. (10,000)		
	Disp. Dist.	Null	LHF	HHF	Null	LHF	HHF	Null	LHF	HHF
GS	12.5k	53.2	39.0	33.6	54.7	41.6	36.3	55.1	42.2	99.5
	25k	54.1	40.4	34.9	58.4	47.7	42.2	60.1	50.1	44.5
	37.5k	54.9	40.8	35.0	59.9	49.9	44.2	62.8	54.2	48.3
	50k	55.3	41.1	35.0	60.4	51.1	45.2	64.5	56.6	50.5
MC	12.5k	17.3	14.6	12.2	18.2	15.7	13.5	18.5	16.0	13.9
	25k	18.1	15.4	14.6	20.5	18.5	16.7	21.4	19.5	17.8
	37.5k	18.5	15.6	17.3	21.6	19.5	17.8	22.9	21.2	19.8
	50k	18.5	15.5	12.6	22.2	20.2	18.4	24.0	22.3	21.0
DE	12.5k	10.4	8.3	7.6	10.7	8.8	8.2	10.8	9.0	8.3
	25k	10.6	8.6	7.9	11.7	10.0	9.5	12.1	10.5	10.0
	37.5k	10.8	8.7	8.0	12.2	10.5	9.9	12.9	11.4	10.8
	50k	10.9	8.6	7.9	12.4	10.8	10.1	13.4	12.0	11.4
SA	12.5k	1.4	1.4	1.3	1.6	1.5	1.4	1.7	1.6	1.5
	25k	1.4	1.4	1.2	2.2	2.1	2.0	2.5	2.4	2.3
	37.5k	1.4	1.3	1.2	2.4	2.3	2.2	3.0	2.9	2.8
	50k	1.3	1.2	1.1	2.4	2.3	2.2	3.4	3.3	3.1

Number of Patches (NP)

		Low Pop. (625)			Med. Pop. (2,500)			High Pop. (10,000)		
	Disp. Dist.	Null	LHF	HHF	Null	LHF	HHF	Null	LHF	HHF
GS	12.5k	2724	962	844	2724	861	715	2473	770	3
	25k	107	197	234	1682	290	325	1571	234	271
	37.5k	68	99	112	1344	201	253	1178	134	178
	50k	34	49	50	127	88	100	875	88	129
MC	12.5k	478	359	401	476	340	344	368	299	319
	25k	89	106	135	185	144	155	169	131	137
	37.5k	68	66	96	165	131	128	124	102	103
	50k	43	41	58	63	63	69	80	72	74
DE	12.5k	140	158	173	140	146	155	120	131	144
	25k	41	64	61	99	98	102	74	87	91
	37.5k	32	41	38	67	81	84	54	69	74
	50k	16	20	18	32	42	40	50	59	67
SA	12.5k	119	119	119	119	118	115	118	117	114
	25k	56	55	54	95	93	91	89	90	88
	37.5k	38	36	33	89	88	88	69	68	71
	50k	21	20	20	45	46	47	57	56	55

Table continued below…

Correlation Length (x 10^5 m; CL)										
		Low Pop. (625)			Med. Pop. (2,500)			High Pop. (10,000)		
	Disp. Dist.	Null	LHF	HHF	Null	LHF	HHF	Null	LHF	HHF
GS	12.5k	4.98	4.16	3.73	4.93	4.13	3.76	4.93	4.14	7.44
	25k	5.02	4.30	3.93	4.95	5.89	5.84	6.09	5.90	5.86
	37.5k	5.05	4.39	4.24	6.15	5.97	5.94	6.07	6.03	5.88
	50k	5.09	4.90	4.34	6.21	6.06	6.02	6.15	6.20	5.96
MC	12.5k	2.18	2.06	1.40	2.16	2.05	1.42	2.16	2.05	1.42
	25k	2.23	2.12	1.50	2.19	2.19	1.80	2.22	2.19	2.06
	37.5k	2.25	2.18	1.58	3.19	2.20	2.10	3.19	2.24	2.09
	50k	2.31	2.22	1.85	3.22	2.24	2.13	3.21	3.19	2.23
DE	12.5k	1.65	1.36	1.25	1.64	1.36	1.32	1.64	1.42	1.32
	25k	1.66	1.45	1.44	1.64	1.64	1.60	1.63	1.65	1.60
	37.5k	1.68	1.67	1.66	1.65	1.67	1.67	1.75	1.65	1.67
	50k	1.70	1.70	1.70	1.77	1.70	1.69	1.75	1.68	1.69
SA	12.5k	.227	.226	.215	.226	.226	.224	.226	.226	.224
	25k	.245	.245	.243	.288	.288	.290	.294	.286	.288
	37.5k	.266	.266	.262	.303	.296	.298	.357	.358	.344
	50k	.336	.335	.328	.383	.374	.373	.426	.426	.429
Largest Patch Index (% of sum of patches; LPI)										
		Low Pop. (625)			Med. Pop. (2,500)			High Pop. (10,000)		
	Disp. Dist.	Null	LHF	HHF	Null	LHF	HHF	Null	LHF	HHF
GS	12.5k	39.6	24.7	19.2	40.1	26.0	21.0	40.4	26.5	100.0
	25k	40.4	26.4	21.1	42.5	43.5	38.8	55.3	45.6	40.7
	37.5k	41.1	27.2	22.4	55.7	45.9	41.1	57.5	49.9	44.2
	50k	41.6	29.9	23.0	56.4	47.6	42.5	59.6	53.1	46.6
MC	12.5k	7.78	6.69	5.42	8.07	7.10	5.92	8.20	7.21	6.07
	25k	8.21	7.22	6.05	9.05	8.33	7.76	9.54	8.74	8.20
	37.5k	8.51	7.50	6.26	15.5	8.82	8.30	16.2	9.45	8.89
	50k	8.71	7.65	6.40	15.9	9.20	8.61	16.9	15.8	9.40
DE	12.5k	5.34	4.46	3.76	5.43	4.66	4.27	5.46	4.71	4.32
	25k	5.45	4.68	4.32	5.69	5.25	4.82	5.80	5.44	5.00
	37.5k	5.54	4.79	4.47	5.86	5.51	5.24	6.03	5.74	5.48
	50k	5.61	4.84	4.54	5.97	5.65	5.38	6.15	5.92	5.64
SA	12.5k	0.13	0.13	0.12	0.14	0.14	0.13	0.15	0.14	0.14
	25k	0.15	0.14	0.13	0.27	0.26	0.26	0.30	0.29	0.28
	37.5k	0.15	0.15	0.14	0.30	0.29	0.29	0.34	0.34	0.33
	50k	0.23	0.22	0.20	0.33	0.32	0.31	0.52	0.51	0.50

Table 3-2 Effects of dispersal ability and population size on four graph metrics:
1) percentage of total land area (PLAND), 2) number of patches (NP), 3) CL (GYRATE_AM), 4) largest patch index (LPI). The table includes the three different modeling simulations the null or "pristine" landscape, the low human footprint (LHF, low road and human development resistances), and the high human footprint (HHF, high road and human development resistances). Dispersal distances are separated based on biome type (GS = Grassland/Shrub, MC = Mixed Conifer, DE = Desert, SA = Subalpine).

3.4.4 Largest patch index (LPI)

The LPI metric decreased non-linearly with decreased dispersal ability and population size for all biome types. The LPI and CL metrics appeared to correlate strongly for the GS and MC biome types (Figure 3-5Figure 3-6, Figure 3-7, panel d, Table 3-2). For both GS and MC, strong threshold patterns were present at identical dispersal ability and population sizes for both LPI and CL. This was to be expected because both GS and MC had the largest starting patch sizes that were likely heavily weighted in the CL metric. In contrast, DE showed smooth (close to linear) changes in LPI in response to changes in life history traits that did not indicate any strong threshold effects. The pattern of change in LPI mirrored that of PLAND in the DE heat maps (Figure 3-7, panel d). The LPI ranged from 5.3% to 6.2% for DE, a maximum of a 14.5% decrease. Among all biomes, SA experienced the most extreme change in LPI (from 0.5% to 0.1%, an 80% decrease, and had strong threshold effects where LPI dropped as much as 45-50%).

3.4.5 Grassland/Shrub (GS)

The extent of the GS biome covered over half the Western United States (PLAND was between 53-65%) with 40-60% of this being a contiguous patch (LPI) across all life-history trait combinations in the "pristine" scenario. Thus, it was no surprise GS was also the most connected of all biome types in the "pristine" scenario (the largest CL among all biome types, Table 3-2). Habitat loss was also the largest of any biome type for both human footprint scenarios and varied between a 6-14% loss for LHF and 14-20% loss for HHF, a percentage difference of 12-27% and 22-37% relative to the "pristine" scenario. When dispersal ability or population size was low, the losses in LPI were exacerbated (between 35-50%); reaching a low in the HHF scenario of 19.2%. There were dramatic effects of human footprint on the number of disjunct patches of internally connected habitat, where NP decreased by over 70% in most cases. It was only for small population sizes that this trend did not hold and there were large gains in NP (coupled with large decreases in PLAND, LPI and CL) as habitat fragmentation increased in medium and large sized patches.

3.4.6 Mixed Conifer (MC)

In the "pristine" scenario, the percentage of the landscape in connected habitat (PLAND) ranged from 17.3-24 % (Figure 3-6, panel a; Table 3-2), a difference of 28%. The amount of habitat loss due to the human footprint was slightly higher (a maximum difference of ~ 32% in PLAND) than the difference in extent due to life history traits (~28% difference in PLAND). The greatest habitat loss due to human footprint occurred when population size or dispersal ability were low (e.g., > 25% reduction; Figure 3-6, panel a). In contrast, habitat fragmentation occurred at all levels and was strongest at moderate to high population sizes (2500-10000) and high dispersal abilities (37.5 km-50 km; Figure 3-6, panels c and d). In this zone, LPI and CL tended to decrease rapidly (44-46% in LPI and 30-34% in CL) in the HHF scenario, while PLAND decreased less (13-17%) and coincided with moderate losses to small gains in NP (-10-22%).

3.4.7 Desert (DE)

The DE biome was best characterized as composed of several medium sized patches (initial LPI = 6.15%) covering 10.4-13.4% of the landscape in the "pristine" scenario. Habitat loss due to the LHF and HHF scenarios (~21% and ~27%, respectively) was similar to the difference due to life history traits alone (~ 22%). Habitat fragmentation caused by human footprint was consistently observed across all life history traits with increases in NP and little change in CL. Habitat loss was most noticeable when populations were small and/or dispersal abilities limited with the largest proportional decreases in PLAND, CL, and in LPI (Figure 3-7, panels a, b and c).

3.4.8 Sub-alpine (SA)

For the SA biome type, habitat extent (PLAND) was much smaller compared to the other three biome types and ranged between 1.3-3.4% for the "pristine" scenario (a difference of ~62% due to changes in dispersal ability and population size). The relative impact of human footprint was also much less with decreases in PLAND between 0-0.1% and 0.1-0.3%, for the LHF and HHF scenarios (a maximum difference of 8% and 17%, respectively). Most metrics were nearly linear in change (or remained constant) in response to human footprint, instead life history traits dominated changes in habitat occupancy and fragmentation. The most habitat loss due to human footprint happened at the lowest population size. Habitat fragmentation due to human footprint was nearly negligible.

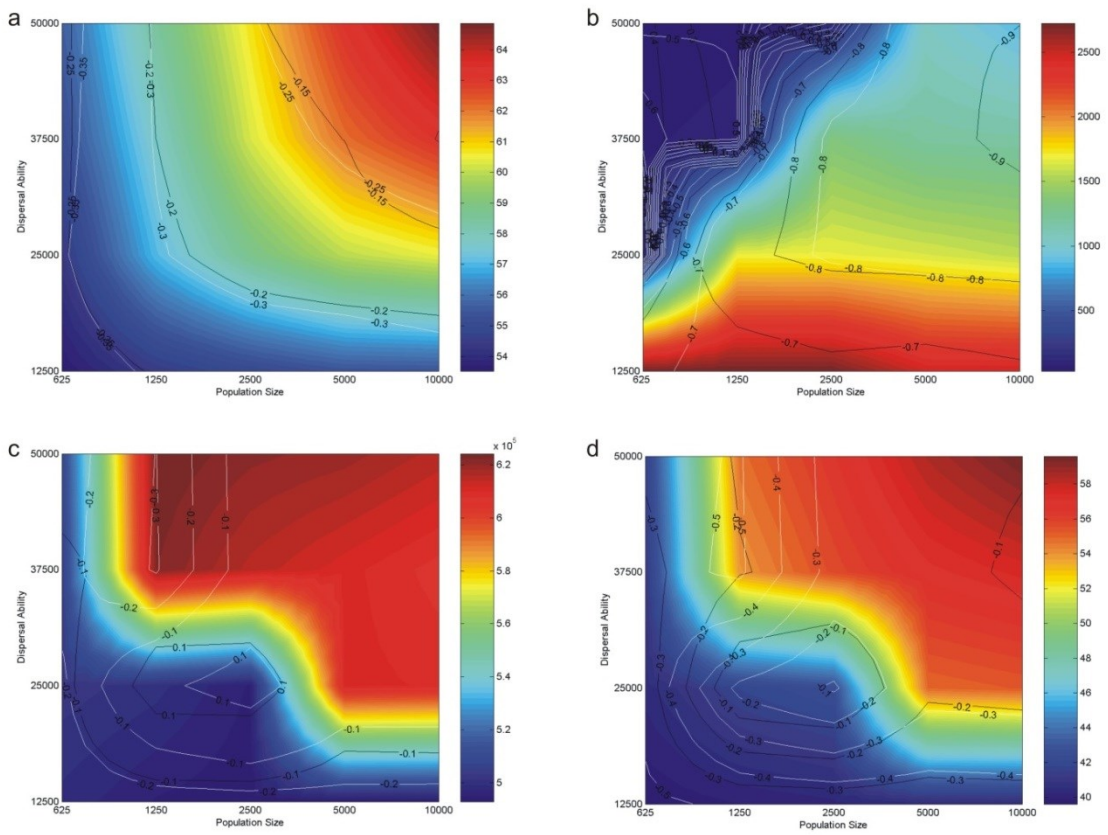

Figure 3-5 Four-panel heat maps of the relationship between fragmentation metrics, dispersal ability and population size for Grassland/Shrub (GS) biome type. Graphs are from resultant output from the FRAGSTATS graph metrics program. By panel, a) is the percentage of total land (PLAND) covered by each biome, b) the number of patches (NP) for each, c) the CL (GYRATE_AM), and d) the percentage of land area of the largest patch (LPI) according to biome type. Contours are: white = proportional reduction in the metric from the null to the high human footprint scenario; black = proportional reduction in the metric from the null to the low human footprint scenario.

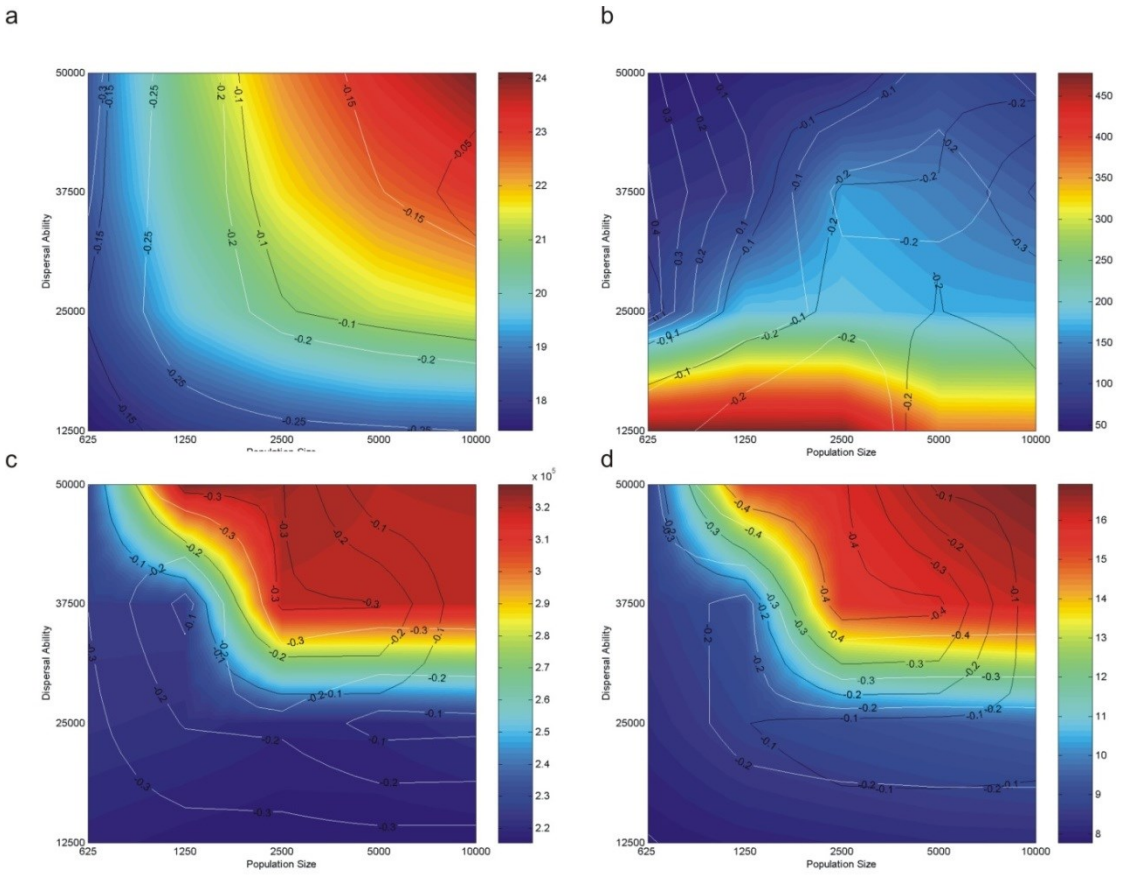

Figure 3-6 Four-panel heat maps of the relationship between fragmentation metrics, dispersal ability and population size for Mixed Conifer (MC) biome type. Graphs are from resultant output from the FRAGSTATS graph metrics program. By panel, a) is the percentage of total land (PLAND) covered by each biome, b) the number of patches (NP) for each, c) the CL (GYRATE_AM), and d) the percentage of land area of the largest patch (LPI) according to biome type. Contours are: white = proportional reduction in the metric from the null to the high human footprint scenario; black = proportional reduction in the metric from the null to the low human footprint scenario.

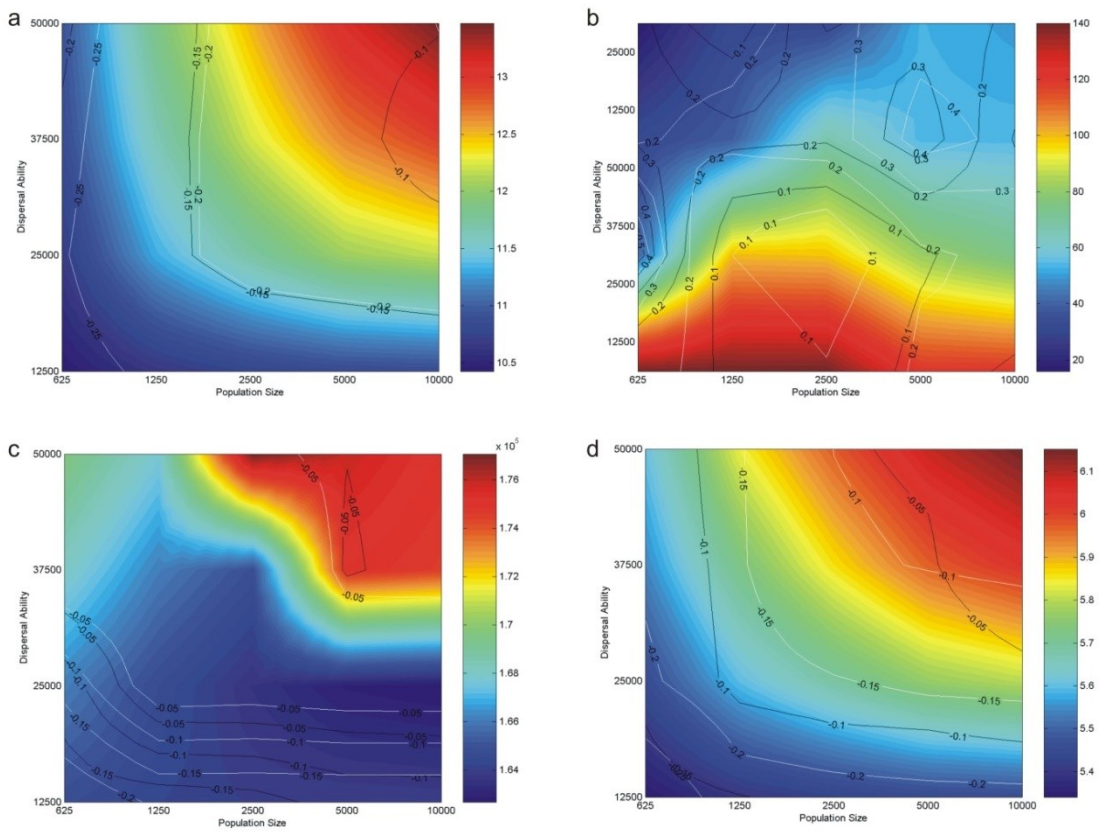

Figure 3-7 Four-panel heat maps of the relationship between fragmentation metrics, dispersal ability and population size for Desert (DE) biome type. Graphs are from resultant output from the FRAGSTATS graph metrics program. By panel, a) is the percentage of total land (PLAND) covered by each biome, b) the number of patches (NP) for each, c) the CL (GYRATE_AM), and d) the percentage of land area of the largest patch (LPI) according to biome type. Contours are: white = proportional reduction in the metric from the null to the high human footprint scenario; black = proportional reduction in the metric from the null to the low human footprint scenario.

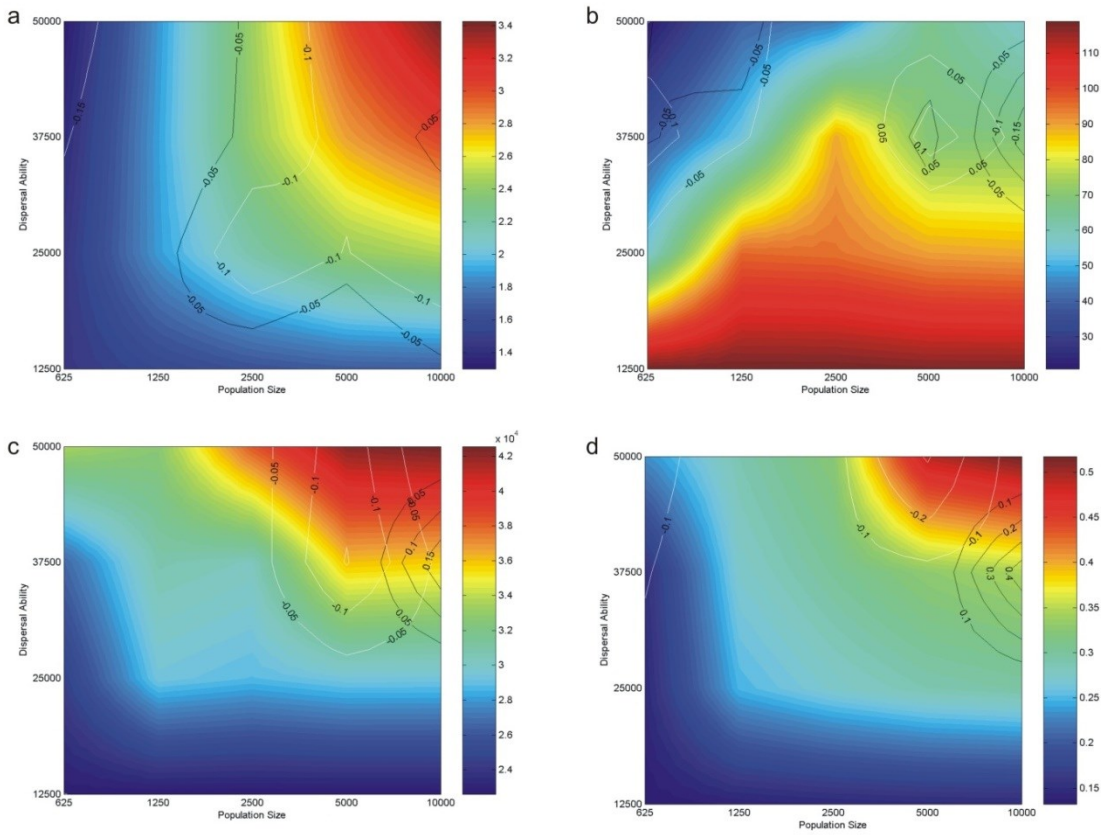

Figure 3-8 Four-panel heat maps of the relationship between fragmentation metrics, dispersal ability and population size for Sub-alpine (SA) biome type. Graphs are from resultant output from the FRAGSTATS graph metrics program. By panel, a) is the percentage of total land (PLAND) covered by each biome, b) the number of patches (NP) for each, c) the CL (GYRATE_AM), and d) the percentage of land area of the largest patch (LPI) according to biome type. Contours are: white = proportional reduction in the metric from the null to the high human footprint scenario; black = proportional reduction in the metric from the null to the low human footprint scenario.

3.5 Discussion

3.5.1 Hypothesis 1: Dispersal ability will play a greater role than population size in its influence on habitat fragmentation and extent.

Our results indicated that the extent of connected habitat decreased rapidly with the decline of either population size or dispersal ability, and did not show a clear difference in the relative importance of either life history trait. There was a strong threshold effect for both traits that caused the extent of connected habitat to decline abruptly when either life history trait was low. Here and in Cushman et al. (2010) when either dispersal or population size passed below this lower threshold, even large increases in the other life-history trait caused very little additional impact on the extent of connected habitat. This made both dispersal ability and population size equally important in protecting against habitat loss, and having high dispersal ability or large population size, did not make up for a major loss in the other. Similar to Cushman et al. (2010), and consistent with the first hypothesis, the correlation length of connected habitat, as well as the extent of the largest connected patch and especially the number of disjunct patches, were all more strongly affected by dispersal ability than population size. These results suggested that population size and dispersal ability were of equal importance in determining the amount of habitat loss, but that habitat fragmentation was likely to increase more rapidly as dispersal ability decreased than as population size increased. Decreases in population size only consistently equated to decreases in extent of connected habitat (PLAND), but did not necessarily lead to increasing fragmentation (a decrease in CL or increase in NP). While decreases in dispersal ability always lead to increased fragmentation, population size suggested a much more dynamic relationship with habitat fragmentation.

3.5.2 Hypothesis 2: In order, GS, DE, MC and SA will show the greatest to least impact of the human footprint on habitat fragmentation and loss in the Western United States.

Cushman et al. (2011) found human footprint to have a greater impact on the amount of habitat loss and fragmentation on species associated with a grassland biome versus a forestland biome across the Great Plains in the United States. Consistent with Cushman et al. (2011) and the second hypothesis, species associated with the GS biome had the largest decrease in extent and largest increase in fragmentation of habitat by roads and human land-uses. The GS biome type had the largest amount of loss in contiguous habitat (PLAND) due to human footprint, with decreases as great as 20%. Both MC and GS, suffered large percentage losses in the largest patch of contiguous habitat (LPI and CL) due to human footprint relative to other biome types, while MC had the largest percentage decrease. The DE biome was less impacted by habitat fragmentation though the amount of habitat loss was similar in both the DE and MC biome. Our second hypothesis was not fully supported, because DE did not reflect a greater change in fragmentation and loss than MC. Specifically, the relative change in three of four landscape metrics (PLAND, LPI, CL) from the "pristine" to the human footprint scenarios was greater for the MC associated species than for the DE associated species. One of the major differences between the biome types was the marked loss in contiguous habitat for MC even for high population and dispersal sizes. Because MC showed a highly fragmented distribution to begin with, the MC biome type might be most at risk

for fragmentation and more so than the DE biome. It was expected that human footprint would have a larger impact on DE associated species given the high overlap of human development with the DE biome, especially in areas of Southern California and Arizona. Major crop related cultivation takes place in these areas, while there is less human habitation in the MC biome, which is also often more protected (e.g., National forest land and parks). Results indicated that roads and human land-use likely have had a larger impact on the extent and connectivity of native wildlife and plant populations in the MC biome than in the DE biome. Finally, consistent with hypothesis two, species associated with SA habitats are predicted to experience far less relative impact of roads and human land-uses on the extent of connected habitat. This was not surprising; given that most of the SA biome is federally managed, with a large portion protected by National Park, National Forest and Wilderness designation.

3.5.3 Hypothesis 3: Human footprint will have a larger impact on fragmentation of habitat than on habitat loss, due to the fragmenting effect of roads and the dendritic pattern of human land-uses along transportation networks.

Cushman et al. (2010) found roads to cause greater habitat fragmentation than the relative effect of habitat loss due to changes in land-uses (based solely on changes in CL). Here, land-uses and roads were not separated and their impact was considered together. Habitat loss always occurred in response to human footprint (PLAND always decreases, and most times LPI also). The relationship between human footprint, biome type and habitat fragmentation, however, was not as simple. In some cases the hypothesis of greater fragmentation than habitat loss held, but did not have full support from all biome types.

The SA biome type showed the least support for hypothesis 3 with very little change due to habitat fragmentation (NP mostly decreased with little change in CL). Perhaps the most supportive, at first glance, was the DE biome where the number of disjunct patches (NP) almost always increased from the "pristine" scenario. The DE biome, however, started out with smaller relative average patch size to begin (low LPI) and with smaller changes in NP, fragmentation was only moderate as all patch sizes tended to shrink. The MC biome time is likely the most susceptible to habitat fragmentation because it experienced the largest percentage change, and the most consistent high rates of change in CL, the metric most indicative of fragmentation. Additionally, there were sharp declines in the area of largest contiguous habitat (LPI) coinciding with moderate decreases in the total amount of habitat (PLAND) and moderate decreases to small increases in the number of disjunct patches (NP). Alone, NP is not a pure indicator of habitat fragmentation, if one considers the largest contiguous patches to be most important to habitat connectivity and therefore most susceptible to habitat fragmentation, as well as loss.

Fragmentation was often most impactful at low population sizes in all but the SA biome. At these thresholds, there were large decreases in the largest contiguous habitat relative to total habitat extent (high relative loss in LPI to PLAND, with large decreases in CL, and increases in NP). This trend was also often true for low dispersal ability, without an increase in the number of disjunct patches (NP). At these thresholds, fragmentation was likely most impactful and the difference in a positive or negative gain in NP was due to the net gain from large patches fragmenting into smaller patch sizes

over small patches disappearing completely. Overall, the hypothesis was not well supported at high population sizes or larger dispersal abilities, but was often supported when either of these life history traits was low, and was partly dependent on biome type.

3.5.4 Conclusion

The extent of connected habitat was approximately equally sensitive to the dispersal ability and the size of the population of the focal species. In contrast, the fragmentation of habitat was more sensitive to dispersal ability than to population size. Habitat loss and fragmentation due to human footprint were also equal in effect when both dispersal ability and population sizes were large. The effects of fragmentation due to human land and road use also increased when dispersal ability or population size was low. Human footprint had the largest effects on the population connectivity of grassland (GS) associated species, as would be expected given the very high impact on these ecosystems relative to other biomes in the Western United States. The MC biome appeared to be the second most affected biome, followed by DE, with the population connectivity of species associated with the SA biome much less affected by human footprint.

The 48 cumulative resistant kernel maps of expected population connectivity produced here could be of great value to scientists and managers. These maps considered several important aspects of biome association, dispersal ability, population size, and human footprint effects and would benefit those wishing to evaluate spatially-explicit management and conservation scenarios. For example, the resistant kernel predictions could be intersected with highways and other anthropogenic landscape features to identify potential barriers to dispersal (e.g., Cushman et al. 2009) and evaluate their relative importance on connectivity (e.g., Cushman et al. 2010). The cumulative resistant kernel maps could also be used to identify core areas and fracture zones for a range of species with different dispersal ability and population size (e.g., Cushman et al. 2011), identify species at risk (Cushman & Landguth 2012a), evaluate the effectiveness of protected lands in connecting habitat for different taxa (Cushman *et al.* 2012), or the effectiveness of one taxa in providing umbrella protection for connected habitat of others (Cushman & Landguth 2012b). Our results are available for download, as well as illustrated in a web-based, interactive mapping prototype (http://ptolemy.dbs.umt.edu/westwide/) that should be useful for evaluating population connectivity to guide conservation and management efforts for such umbrella species.

This chapter focused solely on producing a method for addressing gene flow for several species and two major life history traits (dispersal and population size) when there is not genetic data available. This is done using a more theoretical approach, however, it is important to have improved methods when empirical genetic data exists to parameterize and guide resistance map creation. The next chapter presents a software tool for this purpose of optimizing resistance map creation using measures of genetic distance.

Chapter 4 GARM: A machine learning algorithm for creating resistance maps in landscape genetics

4.1 Chapter Summary

In many systems, genetic data is becoming widely available for resistance map creation and parameterization. It is important to have an optimized tool for these systems because the search for optimal weightings of resistance maps is a computationally intensive endeavor. A Genetic Algorithm for Resistance Map creation (GARM) attempts to converge upon the optimal solution to relate landscape features to genetic structure with the additional attribute of finding the weighted resistance of each landscape feature. GARM is a new tool to aid in understanding genetic connectivity by developing a more rigorous approach to corridor modeling. The aim of GARM is to facilitate a less-biased, or expert opinion driven process for the creation of resistance maps, as well as to provide a promising exploration and optimization tool for landscape genetic studies.

4.2 Introduction

The field of landscape genetics attempts to discover the environmental drivers of genetic variation typically by using distance-based correlative methods (e.g., Manel *et al.*, 2003; Cushman *et al.,* 2006). Common methodology used for explaining the observed pattern of gene flow in a species of interest relies heavily upon weighted, individual landscape features (e.g., elevation, habitat type, or road coverage) that are combined into resistance surfaces (e.g., Spear *et al.*, 2010). A resistance surface then becomes a hypothesis for movement (i.e., gene flow), allowing for identification of areas that impede or enhance connectivity. There is often no general consensus, however, on the appropriate selection and weighting scheme of landscape features.

Resistance map creation is often biased by expert opinion due to a lack of alternative methods for validation (for discussion see Zeller *et al.*, 2012). Furthermore, many different resistance maps of various combinations of the same landscape features and weights are needed to test for competing hypotheses (Cushman *et al.*, 2006; Shirk *et al.*, 2012). This becomes a time consuming process done by hand that reduces the number of environmental layers considered and weight classes of each. For example, consider landscape resistance hypotheses for a system that is comprised of two landscape features with three weight classes for elevation (low, medium, high) and two weight classes for forest cover (forest, non-forest), and each weight class was allowed to vary resistance between 1-10, then there are 10^5 possible combinations for this simple system. A typical landscape genetics study may only consider, to date, at most ~100 such landscape configurations or only 0.1% of the total combinations. For example, Cushman *et al.*, (2006) used individual-based landscape genetics analysis to predict landscape resistance for American black bear and found that population connectivity is facilitated by middle elevation forest and resisted by non-forest areas and roads using 110 such landscape

configurations. For this reason, many landscape genetic systems are not thoroughly tested for parameter sensitivity (Sawyer *et al.* 2011).

Due to temporal and spatial complexities inherent in modeling genetic connectivity, recent literature has suggested resistance map creation will benefit from using machine learning approaches for optimization (Spear *et al.* 2010). Needed are automated, robust methods to explore the large parameter space of the relationship between complex landscape feature and genetic structure, such as offered with machine learning approaches. To help answer this need, we introduce GARM, a machine learning approach designed to take any number of landscape (e.g., elevation, habitat type, or road systems), environmental (e.g., temperature or precipitation), or behavioral (e.g., predator/prey presence or resource selection functions) layers used in resistance map creation.

4.3 GARM v1.0 Program Architecture

The GARM (v 1.0) program utilizes the UNIversal CORridor network simulator (UNICOR; Landguth *et al.*, 2012) to generate shortest-path models. The UNICOR simulator uses a modified version of the Dijkstra's algorithm and parallel-processing to efficiently find multiple paths between sets of pair-wise point combinations which represent locations of individuals on the input landscape (Dijkstra 1959). The GARM program leverages the efficiency of the UNICOR program to search for an optimal match between a set of derived landscape resistance weights and observed genetic distances.

The user begins by describing their environmental variables as ASCII input files, which are easily created using a Geographic Information System (GIS) or equivalent program (e.g., 'raster' package in the R open source software R; Hijmans and Van Etten, 2013; R Core Team, 2013). Environmental layers are user-defined and the user can define attributes for each layer, such as the number of weight classes, ranges for each weight class, and a range of possible weights for each class, among other options. From the previously mentioned example system, a user may have a simple system with two environmental layers; elevation and forest cover. Elevation is composed of three classes; low, medium and high, while forest cover has only two; forested and non-forested areas. The user then suggests a relative weight range (e.g., 1-100) for all classes in each environmental layer. The user also supplies the starting locations for source populations or individuals and observed genetic-distance data (e.g., proportion of shared alleles, (e.g. D_{ps}; Bowcock *et al.*, 1994) or local pairwise (e.g. F_{ST}; Nei, 1973) they have for each source point.

The algorithm begins with the random creation of a set of resistance maps (or population of solutions; a typical starting number would be 100 resistance maps). For each map using the least-cost path connectivity model within UNICOR, a cost-distance matrix is calculated (the accumulated least-cost paths between all locations). These cost-distance matrices are then correlated to the observed genetic-distance matrix provided by empirical data (e.g., population pairwise F_{ST} values for populations on a landscape). Correlative tests are performed using the C-coded zt: Software tool for the simple and partial Mantel Tests (Bonnet & de Peer 2002). The measure of goodness-of-fit for each resistance map (in genetic algorithmic terms, the fitness calculation) is a user option and

can be either of the resulting correlation coefficient or *p*-value from a simple or partial Mantel test.

Once a starting set of maps have been produced and goodness of fit assigned, the algorithm enters the main algorithmic loop (Figure 4-1). There are three major components to each iteration (generation) of the main loop: 1) the combination of two solutions, in a user-specified manner, to create new solutions (crossover), 2) the random change of a single weight, for a single class in a single resistance map configuration (mutation), 3) new solutions are chosen from the full set of previous solutions with newly created solutions (fitness-proportional selection using parents and children). The GARM tool ceases to run either upon reaching a user-specified number of iterations (generations) or when the correlation coefficient (fitness) is above a specific threshold.

There are five types of crossover available (names in italic represent configuration file keywords), including: the *traditional* method where each parent contributes equally to the child's weights, *random* assignment of weights from each parent, *random alleles* where each parent contributes an entire layer set of weights, a *random split* similar to the traditional method, but at a random spot instead of equally half from each parent, and lastly, the *average* of the weights from the parent. Using a timer on the child creation stage, one can impose more stochasticity on the model by introducing randomly created models (immigration) to alleviate endless loops in child creation if the unique models are difficult to generate from the current available population of models (Cobb 1993). Other useful features include a "hot start" for run restarts, visited solutions tracking and output in an R (i.e., comma separated value) friendly format.

The GARM tool was verified by creating resistance maps with random weights and ten randomly placed starting points using the hypothetical system of elevation and forest cover on a small (70 x 110) test grid. In total, ten sets of randomly created starting weights and source points were then used in UNICOR to produce hypothetical cost-distance matrices. Next the output cost-distance matrices from UNICOR were run in the GARM tool for the ten simple hypothetical resistance maps. Tests were run until reaching perfect correlation between the simulated input genetic-distance matrix and the cost-distance matrix calculated by GARM ($r = 1.0$) for all verification tests. On average, it took 205 generations for this simple example system.

The GARM tool is compatible with Python 2.7 and is provided with installation instructions for most platforms, along with the needed inputs for the example system described in this note. The main input file accepts parameters organized as name-value-pairs in a stanza-oriented, text file format. The inputs are parsed using the RipMgr package (Glassy, unpublished library), a flexible symbol table manager for science models that includes special parsing capabilities. The GARM tool is freeware and can be downloaded at the Computational Ecology Lab (CEL) website (cel.dbs.umt.edu/garm) with information for users, including a user manual.

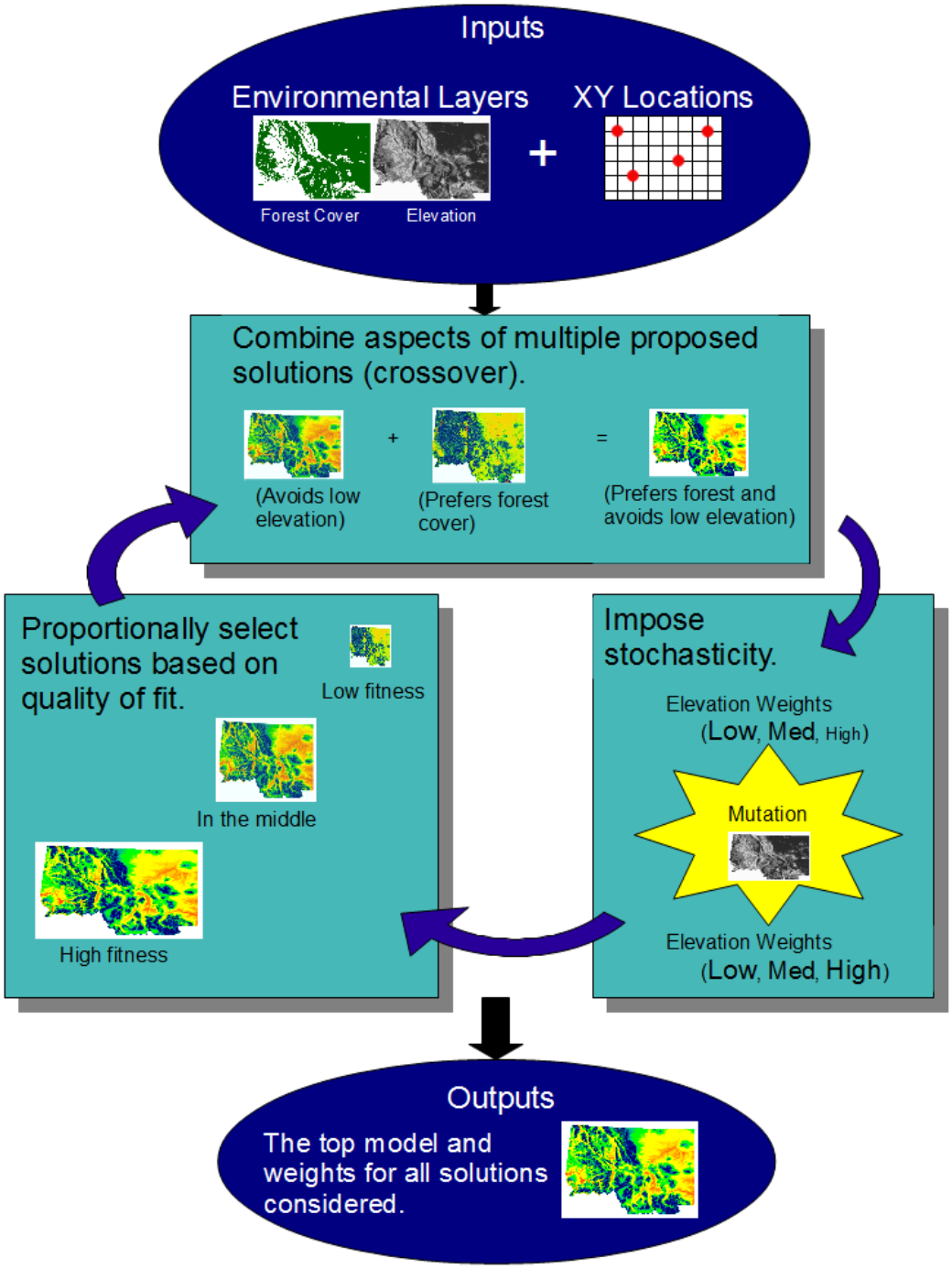

Figure 4-1 Main algorithm workflow of GARM. 1) Combination of multiple proposed solutions, 2) imposing stochasticity on the newly created solutions, and 3) selecting a new solution set based on the quality of fit for the combination of the beginning set of solutions with newly created solutions.

4.4 Conclusion

Both the GARM tool and UNICOR simulator are written in the Python programming language taking advantage of the parallel-processing in Python, and the optimization of the Numpy package (numpy.scipy.org), to greatly expedite least-cost path calculation, as well as flexibility to be easily adapted to any landscape connectivity study. GARM is modular in its development with the structural connectivity modeling software of UNICOR being able to be replaced with any number of other more synoptic modeling approaches (e.g. CIRCUITSCAPE; http://www.circuitscape.org/; McRae, 2006).

The GARM tool helps to remove the bias of expert opinion on resistance map creation. It is important to note that users must still exercise caution in interpreting results from any automated software process and model sensitivity analysis is a necessary step to avoid spurious conclusions. In this regard, GARM will allow users to more proficiently run parameter sensitivity tests with a large number of environmental inputs. Automation will allow users to search much larger solution domains than many previous landscape genetics focused studies. Particularly important is when a user is unsure of which layers are influencing observed gene flow on the landscape. Overall, GARM will aid in building more consistent and explorative landscape genetic studies with flexibility to employ future improvements in connectivity modeling when they become available. In the next chapter, the GARM tool is applied to the same elk system (with the addition of several more populations) presented in chapter 2.

Chapter 5 New landscape genetics approaches for assessing uncertainty in genetic connectivity: Examples using elk from the Greater Yellowstone Ecosystem

5.1 Chapter Summary

The research here involves utilizing GARM (chatper 4) to develop a new computational framework quantifying uncertainty in landscape genetic connectivity modeling. The GARM software program was important in illustrating and comparing several approaches to identifying putative corridors of maternal gene flow using mitochondrial (mt) DNA from 19 elk (*Cervus canadensis*) across the Greater Yellowstone Ecosystem. We compared support of >300,000 alternative connectivity scenarios (resistance surfaces) using 14 combinations of statistical tests, genetic distance metrics, and connectivity modeling approaches (least-cost path and circuit theory). Estimators of explained variance were high for some methods ($r^2 > 0.8$ using Redundancy Analysis) suggesting that ecological distance can explain over 80% of variation in genetic distances between population pairs. The same major corridors were detected with regularity across all statistical tests within a given metric of genetic distance. For example, forested areas were consistently identified as conducive to gene flow. However, sensitivity analysis (by leaving one population out) showed uncertainty in some corridors identified from different combinations of statistical tests, and genetic distance metrics. Our results suggested that future landscape genetic studies will benefit from the following improvements: (1) model optimization, (2) model uncertainty assessment, and (3) performing sensitivity analysis. Quantifying uncertainty will improve confidence in genetic connectivity modeling, thus aiding in corridor (and barrier) identification in conservation and ecology, and also help landscape genetics to develop into a more rigorous scientific discipline.

5.2 Introduction

Population connectivity has been a major focus of conservation planning in response to the present and future threats of habitat fragmentation, invasive species, and environmental change (Noss 1987; Hanski 1998; Crooks & Sanjayan 2006; Cushman *et al.* 2009; Allendorf *et al.* 2013). All of these threats can cause increased rates of inbreeding, genetic drift, and loss of adaptive alleles in local populations (Saccheri *et al.* 1998; Keller & Waller 2002). Fragmentation can also reduce rates of recolonization and demographic and genetic rescue (Hames *et al.* 2008), while it also influences spread of invasive species, diseases, and can negatively interact with climate change (Crooks & Sanjayan 2006; Frankham *et al.* 2010).

Landscape genetics can help identify routes of connectivity (corridors) by quantifying the influence of landscape features on gene flow and dispersal (Manel *et al.* 2003; Storfer *et al.* 2007). It is crucial for managers to have confidence in predictions of

genetic connectivity as these areas are often are targeted for implementation of corridors or barriers to maintain or limit connectivity (Beier *et al.* 1998; Rees *et al.* 2008). Most previous corridor model studies have suffered from a lack of uncertainty assessment. Uncertainty in corridor models can be addressed by using a variety of computational methods (statistical tests, connectivity models or genetic distance measures) and through sensitivity analysis. Many studies have also relied heavily on "expert" opinion when determining the impact on genetic connectivity of specific landscape features. Also, they tend to compare few alternative models, rather than taking a more rigorous, optimization approach (Beier *et al.* 1998; Sawyer *et al.* 2011; Manel & Holderegger 2013).

In landscape genetics, landscape permeability to gene flow is often represented by using a resistance surface where each point is the hypothetical weight (or cost) for an animal to traverse through that point (Spear *et al.* 2010; Zeller *et al.* 2012). The fundamental unit of measure of a resistance surface at a single point in space is the sum of the underlying landscape layer resistance weights (e.g., for elevation, land cover type, etc.). For example, one might hypothesize high forest cover and low elevations facilitate gene flow for a species and therefore assign low resistance weights to those two layer values. From the resistance surface and a set of starting populations (or individuals) one can then calculate the effective geographical distance between each population pairwise combination.

Effective geographical distance is the cumulative cost of all pixels traversed between two populations on a resistance surface (Adriaensen *et al.* 2003; Epps *et al.* 2007; Sawyer *et al.* 2011). Connectivity models like least-cost path are commonly used to calculate effective geographical distances on a resistance surface (using an algorithm such as Dijkstra's shortest path; Dijkstra 1959). Goodness of fit of a resistance surface to empirical genetic data is determined by the correlation between matrices of pairwise effective geographical and genetic distances. The most common statistical test for finding this correlation are Mantel and partial Mantel tests (Mantel 1967; Smouse *et al.* 1986; Sawyer *et al.* 2011). Producing the resistance surfaces necessary to test all alternative corridor models (generated as a result of different sets of resistance values) is a computationally intensive task, but is crucial for finding the optimal (strongest correlation) resistance surface (i.e. optimization).

The first two studies to use a limited form of optimization for resistance surface creation were Wang *et al.* (2009) and Shirk *et al.* (2010). Wang *et al.* (2009) produced 24,843 resistance surfaces and correlated least-cost paths population pairwise distances from each surface with genetic distances (i.e., effective geographical distances) for California tiger salamanders. Shirk *et al.* (2010) introduced a new framework for resistance surface fitting using univariate optimization (i.e., a single landscape feature) followed by iterative multivariate adjustment in study of mountain goats. The only study to attempt a more robust form of optimization was Graves *et al.* (2013), that used optimization with simulated genetic data to evaluate the usefulness of Mantel tests for fitting resistance surfaces. To date, no study we are aware of, has applied optimization to assess, in detail, the uncertainty in modeling genetic connectivity using empirical data.

Optimization is crucial to assess uncertainty (or consistency) among corridor prediction approaches by finding the optimal fitting resistance surface among all method combinations. The GARM software program is an automated, optimization tool valuable for the computationally intensive task of fitting resistance surfaces to genetic structure

(GARM; cel.dbs.umt.edu/garm; Hand, Raiford, Landguth, *et al.* 2013). The GARM tool is also flexible, allowing inclusion of new and future improvements in statistical tests and additional connectivity model approaches. Lastly, it facilitates comparisons of several landscape variables (e.g. forest, elevation, etc.) and resistance weight ranges, allowing for wide searches of landscape variable combinations in a single run.

Uncertainty was assessed by investigating the sensitivity of genetic connectivity modeling to the choice of the statistical test, connectivity model and the genetic distance metric. The results when using Mantel and partial Mantel tests were compared. Also included in the tests were transform-based Redundancy Analysis (tb-RDA) and an additional measure of matrix congruence, the R_v coefficient (a multivariate generalization of the squared Pearson correlation coefficient), as an alternative to Mantel tests (Robert & Escoufier 1976; Legendre & Legendre 2012).

There is currently no consensus on the superiority of one connectivity model over another, so we included least-cost path and circuit theory (McRae 2006; McRae & Beier 2007; Zeller *et al.* 2012). As suggested in Bird *et al.* (2011), our analysis included two different measures of genetic distance, a fixation index (F_{ST}; Excoffier *et al.* 1992), and a standardized genetic differentiation measure, G'_{ST}, an analog of F_{ST}, detailed in Hedrick (2005). For tb-RDA we used haplotype frequencies as an additional, non-distance matrix, non-summary statistic measure of gene flow. Finally, we explored the effects of removing a single population (leave-one-out sensitivity analysis) on model fit.

The overall goal of this study was to build a rigorous framework for identifying landscape features and corridors facilitating gene flow. To meet this goal, we addressed three specific questions using data on maternal gene flow (mtDNA) among populations of elk (*Cervus canadensis*) in the Greater Yellowstone Ecosystem. The specific questions were as follows: (1) Is there strong agreement between different combinations of the methods used, for example when using a different genetic distance metric or connectivity model?, (2) How do results change when removing a single population of genetic information such as performing sensitivity analysis?, and, (3) Is there greater variation in the top-fitting models when using different methods or when removing a single population of genetic information (sensitivity analysis)? The GARM tool allowed the creation and testing of more than 300,000 resistance surfaces (hypotheses of animal gene flow) by considering 14 combinations of methods including two connectivity models, four statistical tests and three genetic distance metrics. From these resistance surfaces we qualitatively compared top-fitting models as identified per each of the 14 methods to test for agreement of corridor predictions among methods. Our results were useful to help draw general conclusions that will benefit future landscape genetics work.

5.3 Materials and Methods

5.3.1 Study species and area

The Greater Yellowstone Ecosystem (GYE) is a world-renowned ecosystem and important ecological system in the United States (Figure 5-1). Spanning three states (Montana, Idaho and Wyoming) the GYE has at its center the world's oldest national park, Yellowstone. The GYE is well-suited to study female elk dispersal due to relatively pristine habitat, large elk population sizes (elk are the most numerous large mammal in the GYE, with estimated numbers in the 50,000s), and good spatial coverage of samples from several elk populations (641 individuals in 19 populations). Lowe & Allendorf

(2010) define dispersal as the movement of individuals between spatially discrete populations for permanent or long-term residence.

Identifying elk genetic structure and movement corridors is important for management and conservation of elk populations and critical habitats. This information can also improve our understanding of disease movement pathways and inform management decisions related to transmission risk to people, livestock and other wildlife. Brucellosis (caused by the bacteria *Brucella abortus*) is of major concern in the GYE, primarily due to the possible transmission from elk to cattle which has negatively impacted all three states (Wyoming, Idaho, and Montana) after recent cattle outbreaks (Beja-Pereira *et al.* 2009). Brucellosis is a disease that causes individuals to prematurely abort their offspring and is transmittable between wildlife and livestock when individuals feed near infected fetuses, placentas, or birthing fluids (Cheville *et al.* 1998). Brucellosis seroprevalence has been reported to be as high as between 20-40% on many of the 23 feedgrounds in Wyoming and has recently increased to 8-20% in populations not associated with feedgrounds (Cross, Cole, *et al.* 2010; Cross, Heisey, *et al.* 2010). Maternal elk genetic structure and dispersal patterns are of great interest in the GYE because elk have been identified as the most likely route of transmission to livestock and there is concern that the area affected by brucellosis in elk may be increasing (Cross, Cole, *et al.* 2010).

5.3.2 Elk sampling

Blood, tissue or fecal pellets were collected from 19 elk populations in the GYE (Table 5-1; Figure 5-1). Here populations were defined as large groups or collections of individuals from (1) areas where elk congregate, such as winter ranges with hundreds to thousands of elk, or (2) by distinct local geographical areas where there were a sufficient number of samples, such as hunting districts where the samples were collected. Blood or tissue was collected from captured individuals (Dell Creek, Fall Creek, Forest Park, Greys River, Jewett, Madison-Firehole, National Elk Refuge, Paradise Valley, Soda Lake, Northern Range; Figure 5-1) or hunter-killed animals (HD311, HD 360, HD 362, Madison Valley, Pioneer Mountains, Shoshone River, Muddy Creek; Figure 5-1)

Fecal pellets were collected from the ground within a few hours after defecation in Canyon, Jenny Lake, Sand Creek, and Bench Corral and from a few individuals (5 of 62 individuals) from Muddy Creek. To prevent sampling more than one fecal pile from the same individual, only very fresh (warm) feces were sampled, from individuals observed defecating, from distant groups (i.e., sets of 5-10 individuals 0.5 to 1 km apart), and from individuals with distinctive natural markings, large ear tags or radio collars. Further sampling details and how samples were sequenced and haplotypes discovered can be found in (Chpter 2; Hand, Chen, *et al.* 2013).

5.3.3 Environmental layers

The original elevation data came from the National Elevation Dataset (NED) at ned.usgs.gov (Gesch *et al.* 2002). Elevation was sorted into five bands of elevation using a quantile classification which generates equally-represented spatial classes. For land type the National Land Cover Database 2001 (NLCD2001) was used, a national database of land cover data found at http://www.mrlc.gov (Homer *et al.* 2007). The original NCLD2001 data, in a 16-class land cover classification system, was kept for this project with some aggregation of lesser represented classes (little to no spatial coverage) to arrive

at four simple, and well represented classes. For example, all levels of development totaled very few pixels over the entire raster; it was decided they were best represented as part of the Pasture/Hay, Cultivated Crops class. All non-forest cover classes (grassland, shrubland, barren land) are included in the grassland/shrub group. Open water and forest remained the same. After reclassification of land type, environmental layers were resampled to 1500 meter resolution using bilinear interpolation for elevation and the majority algorithm for land type in ArcMap 10 (Esri 2011).

5.3.4 The GARM algorithm

The current (in-house) implementation (v1.1.8) of the GARM tool (Chapter 4) was used to search for the optimal fitting resistance surface created from the combination of the environmental layers above. The GARM tool and genetic algorithms, in general, use the concept of natural selection to compete and select among models of varied landscape resistance (Holland 1975; Goldberg 1989). Environmental layer classes (Figure 5-1) were randomly assigned integer weights from 1-100 to create resistance maps that were then allowed to compete using correlation as a measure of each resistance model's fitness (except open water that was set at 100 for all runs). The algorithm was run for at least 200 generations with 100 models in each generation. This is equivalent to considering > 20,000 unique models per run. Stochasticity was imposed at a rate of 4 random weight mutations per generation or per 100 models. Several individuals with randomly assigned weights (immigrants) were introduced each generation to prevent the model from converging prematurely on local maxima (Cobb 1993).

Population	Pop. Abbr.	UTM (m East)	UTM (m North)	Number of Samples	Number of Haplotypes	Haplotype Diversity
Bench Corral*	BC	569692	4730309	13	7	0.872 (0.067)
Canyon Campground	CY	540479	4952566	17	10	0.919 (0.043)
Dell Creek*	DC	550399	4789242	26	12	0.818 (0.073)
Fall Creek*	FC	606925	4746438	26	8	0.793 (0.056)
Forest Park*	FP	524833	4741410	22	12	0.926 (0.031)
Grey's River*	GR	497375	4776953	37	14	0.917 (0.023)
Jenny Lake	JL	522331	4840230	23	12	0.885 (0.050)
Jewett*	JE	548343	4747492	23	13	0.948 (0.024)
Madison Firehole	MF	495338	4945231	42	12	0.858 (0.035)
Hunting District 311	311	466967	5046716	14	8	0.824 (0.097)
Hunting District 360	360	452967	5017564	20	9	0.821 (0.072)
Hunting District 362	362	454247	4977560	58	15	0.901 (0.021)
Muddy Creek*	MC	634277	4721083	89	14	0.706 (0.047)
National Elk Refuge	NER	524075	4820220	28	13	0.886 (0.041)
Paradise Valley	PV	526045	5029408	68	17	0.901 (0.018)
Sand Creek	SC	439437	4845368	19	10	0.906 (0.040)
Shoshone River	SR	608769	4925150	59	16	0.825 (0.039)
Soda Lake*	SL	594270	4756018	13	5	0.782 (0.079)
Northern Range	NR	536585	4979966	44	15	0.830 (0.049)
Totals:				**641**	**47**	

***Denotes a Wyoming feedground.**

Table 5-1 Nineteen elk (*Cervus canadensis*) populations in the Greater Yellowstone Ecosystem with spatial coordinates, and the number of mtDNA haplotypes. All coordinates are in the Universal Transverse Mercator (UTM) NAD83 zone 12 projection.

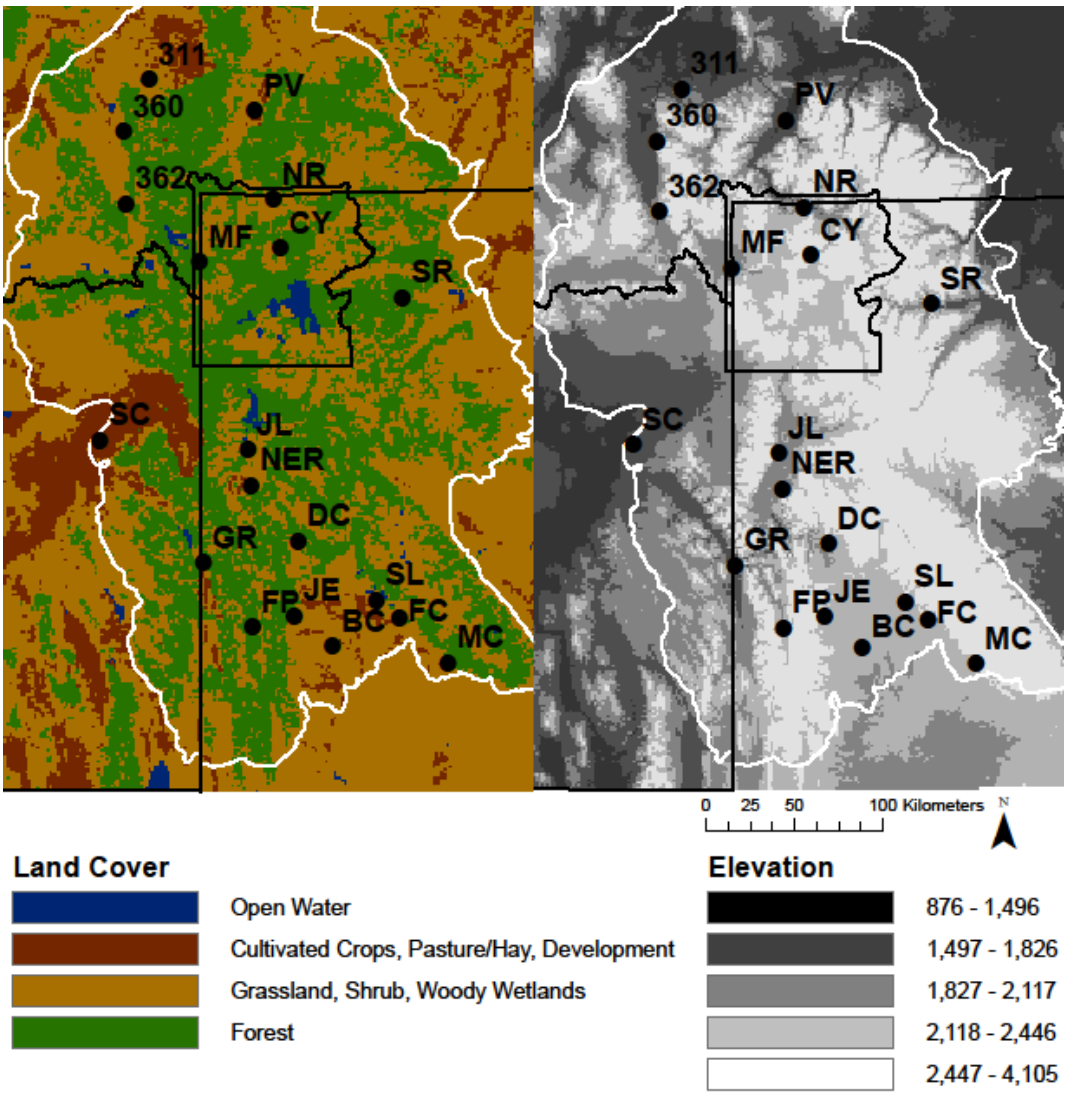

Figure 5-1 Map of 19 elk (*Cervus canadensis*) populations sampled from the Greater Yellowstone Ecosystem. The southernmost populations (GR,FP,DC,JE, BC, SL, FC and MC) are feedgrounds in Wyoming where prevalence of brucellosis is elevated (and where elk are fed hay in winter to keep them from cattle and private ranches). Yellowstone National Park is shown at the center, and the Greater Yellowstone Ecosystem boundary is in white. The map on the left depicts land cover classes used in the landscape genetics study. The map on the right depicts the five elevation classes used.

5.3.5 Genetic distance measures, statistical tests, and connectivity models

Isolation-by-distance was tested for in the pairwise genetic distances (F_{ST} and G'_{ST}) by correlating them to population pairwise Euclidean distances. Isolation-by-distance provided an initial hypothesis of genetic structure for comparison to the computed landscape resistance surfaces (i.e., relative correlations). The F_{ST} genetic distance matrix was calculated in Arlequin 3.5 by considering only the differences in haplotype frequencies using 10,000 permutations to test for statistical significance. The G'_{ST} genetic distance matrix was calculated in SMOGD and checked in the GenAlEx 6.5 (Crawford 2010; Peakall & Smouse 2012). All statistical tests of correlation in GARM were performed using the R packages 'ade4' (Dray & Dufour 2007) and 'vegan' (Oksanen *et al.* 2013).

Mantel tests have long come under scrutiny due to high type I error related to inherent spatial auto-correlation in ecological studies (Guillot & Rousset 2013) and the lack of relative correlative power of distance matrices (Legendre & Fortin 2010; Graves *et al.* 2013). To alleviate type I error stemming from the use of significance-based tests, we used the Mantel r and partial Mantel r, instead of significance values (p values) as correlation values are potentially more useful for model ranking (Fumagalli *et al.* 2011; Cushman, Wasserman, *et al.* 2013).

To quantify resistance surface correlations, the R_v coefficient was considered as an alternative to the Mantel r, due to the recent criticism of Mantel tests. The R_v coefficient also has the desirable mathematical property that several multivariate analysis techniques (principal component analysis, canonical correlation analysis, multivariate regression and discriminant analysis) are equivalent to maximizing the coefficient (Robert & Escoufier 1976; Abdi 2007). To calculate the R_v coefficient principal coordinate analysis (PCoA) was first performed on input effective geographical distance and genetic distance matrices to arrive at tables of eigenvalues projected in Euclidean space (Legendre & Legendre 2012). These tables are then correlated using an R_v test to measure similarity. All R_v coefficient work was done in the 'ade4' R package with significance tested using permutation tests (Heo & Gabriel 1998).

Balkenhol et al. (2009) recommends using Canonical Correspondence Analysis (CCA), but there is currently no software program that calculates the adjusted R^2 for unbiased variance partitioning using CCA (Peres-Neto *et al.* 2006; Legendre & Legendre 2012). The adjusted R^2 for RDA approaches was developed in Peres-Neto *et al.* (2006) to provied an unbiased estimator of the variance explained, therefore, we used RDA which is more flexible than CCA (Legendre & Gallagher 2001). The RDA method is also independent of a genetic distance measure and relies on site-wise haplotype (allele) frequencies. Haplotype frequencies were transformed using a chord transformation in the 'vegan' package, and RDA performed with the PCoA transformed effective geographical distance matrix, termed transformation-based RDA or tb-RDA (Legendre & Legendre 2012).

For modeling connectivity, we applied the two most widely used connectivity models (least-cost path and circuit theory). Least-cost path analysis assumes organisms have complete information of the landscape and will always choose optimal paths (Cushman, McRae, *et al.* 2013). Circuit theory considers organisms to behave like random walkers and that all paths contribute to determining gene flow. The least-cost path models were produced using the UNIversal CORridor simulator (UNICOR;

Landguth *et al.* 2012) to calculate effective geographical distances between populations. This program is the default connectivity model in the GARM tool. The GARM tool was also extended to include a circuit theory connectivity model (CIRCUITSCAPE; McRae *et al.* 2008). Resistance distance matrices from CIRCUITSCAPE function exactly as cost-distance matrices, allowing for the same methods of correlation to be used in GARM to find top models.

5.3.6 Model comparison

The major goal of this study was to investigate a wide range of methods used in identifying corridors by including 14 combinations of statistical tests (Mantel test, partial Mantel test, R_v coefficient and tb-RDA), genetic distance metrics (F_{ST}, G'_{ST} and haplotype frequencies) and connectivity models (least-cost path and circuit theory). The limit of 14 combinations, from potentially 24 total, was dependent on the tb-RDA approach that does not require a genetic distance matrix and uses haplotype frequencies as a metric for genetic differentiation. In analysis, the three matrix dissimilarity measures (Mantel, partial Mantel, and R_v) that used genetic distance metrics were grouped separately from results of tb-RDA.. Paths predicted from least-cost path and circuit theory differ greatly enough qualitatively (visually) that separation of the results was warranted when considering path overlap. For instance, all least-cost path models were included together to study path overlap by looking for agreement of highly traveled paths (> 80% presence in all models). For circuit theory models, cumulative current maps (a measure of effective geographical distance in each cell over a wider area) were summed from the top 50 models and all cells below the top ~10% of cumulative value were removed. Within each connectivity model, results were also separated out between the three genetic distance metrics (F_{ST}, G'_{ST} and haplotype frequencies).

For comparing predicted resistance weights by all combinations, results were organized by the genetic measure used for analysis. The top 50 models were chosen to conduct comparative analysis for several reasons, (1) so that all methods were equally represented in model comparative analysis; (2) the top 50 model cut-off also often represented (visually) clear model thresholds; and (3) the top 50 models often represented (qualitatively) 1-10% of total variation in the model goodness of fit. For the top 50 models in each combination we produced tables of ranges and averages for correlation values, p-values, and variable weights. From these tables we also produced plots of means with the standard errors of weights to visually represent confidence intervals for weight ranges for each landscape variable. Sensitivity analysis results were produced by running the top 50 models for each combination while removing 1 of 19 populations in turn. The resulting correlation values were then averaged and compared to the values from all models. Models that showed a significant drop in the average value from all models when removing a population were considered to not agree well with previous results.

5.4 Results

A total of 641 individuals in 19 elk populations were sampled in the GYE, including 444 females, 18 males, and 179 animals for which sex was unknown. Within population haplotype diversity was high for elk populations in the GYE (Table 5-1), ranging between 0.706-0.948. This high diversity limited the range of F_{ST} for pairwise population

comparisons (0.062-0.265). For example, G'_{ST} pairwise values were found to range between 0-0.716. To roughly compare both genetic distance matrices, a Pearson correlation calculation showed r = -0.06 or almost no correlation existed between the two distance matrices. Because F_{ST} is less appropriate than G'_{ST} for data with high gene diversity within populations, we report only on G'_{ST} and tb-RDA herein.

Initial tests for isolation-by-distance using Mantel tests on the pairwise Euclidean and genetic distance matrices were non-significant using G'_{ST} pairwise distances for the Mantel test (r = -0.011, P=0.53) and the R_v coefficient (R_v = 0.269, P=0.355). The test for isolation-by-distance was also non-significant using raw haplotype frequency (pairwise differentiation) data with tb-RDA (R^2 = 0.002, P = 0.4).

5.4.1 Resistance model comparison

The highest correlation values varied substantially according to the choice of statistical test and connectivity model (Table 5-2). The lowest correlation occurred for least-cost path models using G'_{ST} and a Mantel test for (r = 0.266) while tb-RDA tests had high estimates of the explained variation for the adjusted R^2 (0.712 for least-cost path and 0.831 for circuit theory). For measures of matrix dissimilarity, Mantel consistently produced the lowest correlations (0.266-0.382), while the R_v coefficient produced more highly correlated matrices (0.483-0.750) and the partial Mantel test fell somewhere in the middle (0.450-0.658; Table 5-2).

Test Type	Correlation Coef.	Associated p-value	Forest	Cultivated Crops, Pasture/Hay	Grass/ shrubland	Elevation (876-1496m)	Elevation (1497-1826m)	Elevation (1827-2117m)	Elevation (2118-2446m)	Elevation (2447-4105)
Top 50 Models (G'_{ST}) – Least-cost path (least-cost path)										
Partial Mantel	0.443-0.463 (0.450)	0.001-0.007 (0.002)	1-27 (10.3)	4-70 (33.3)	14-98 (53.8)	4-96 (50.0)	1-20 (7.3)	23-100 (50.5)	50-100 (86.4)	3-31 (18.5)
Mantel	0.230-0.266 (0.242)	0.022-0.078 (0.053)	1-6 (2.2)	1-99 (43.1)	1-100 (74.3)	1-99 (62.2)	1-66 (23.8)	2-91 (22.8)	37-97 (66.0)	1-7 (2.8)
R_v	0.443-0.483 (0.455)	0.050-0.132 (0.082)	1-6 (1.5)	2-99 (53.6)	1-100 (61.4)	1-100 (51.8)	7-97 (31.6)	29-99 (72.4)	2-100 (57.2)	1-7 (2.1)
Top 10 Models (haplotype frequencies) – Least-cost path (least-cost path)*										
tb-RDA	0.621-0.680 (0.650)	0.010-0.031 (0.022)	1-8 (6.1)	3-93 (51.8)	20-99 (78.2)	9-69 (52)	3-91 (70.1)	4-15 (7.2)	35-75 (51.1)	3-40 (15)
Top 50 Models (G'_{ST}) – Circuit theory										
Partial Mantel	0.433-0.450 (0.440)	0.002-0.011 (0.005)	1-73 (23.2)	4-89 (40.6)	15-96 (47.9)	2-67 (26.1)	25-78 (51.3)	51-94 (73.6)	66-100 (90.7)	1-28 (10.6)
Mantel	0.342-0.382 (0.353)	0.005-0.030 (0.013)	1-13 (4.0)	2-88 (41.1)	4-90 (39.7)	1-84 (24.1)	1-64 (24.4)	36-99 (66.5)	63-100 (84.8)	1-12 (4.1)
R_v	0.618-0.630 (0.622)	0.014-0.060 (0.034)	1-19 (5.6)	2-88 (38.0)	1-82 (20.6)	1-89 (29.8)	1-100 (30.0)	41-99 (74.7)	58-100 (84.4)	1-8 (2.7)
Top 50 Models (haplotype frequencies) – Circuit theory										
tb-RDA	0.784-0.831 (0.805)	0.005-0.020 (0.009)	1-10 (4.7)	7-100 (58.0)	36-100 (74.9)	3-95 (48.8)	35-100 (75.6)	1-18 (5.1)	39-98 (72.6)	1-56 (17.4)

Table 5-2 Summary statistics for the top 50 models using the least-cost path connectivity model. Rows are sorted by the genetic distance metric used (G'_{ST} and haplotype frequencies) and by statistical test (Mantel test, partial Mantel test, R_v coefficient and tb-RDA). The ranges include the min and max values, with averages in parenthesis. Values are reported for the correlation coefficient and p-value for the chosen statistical test, and variable weights. Open water was removed from the table as it is set to 100 in all model runs.

5.4.2 Sensitivity analysis

Sensitivity analysis (leave-one out population; Table 5-3) was conducted across all combinations of methods with mixed results. Dissimilarly matrix tests (Mantel, partial Mantel and R_v coefficient) showed some consistency within the same measures of genetic distance across connectivity models. For example, the BC population was identified by sensitivity analysis as problematic for both connectivity modeling approaches when using the three dissimilarity tests for G'_{ST}. The most noticeable effects occurred with the removal of the populations that changed the overall significance of the set of top models. One instance of this effect was when removing BC using G'_{ST} as a genetic distance metric with the least-cost path connectivity model. Previous to removing the BC population, all models were significant (or slightly non-significant using the R_v coefficient). Once BC was removed all models shifted to highly non-significant with the Mantel (P = 0.053 to P = 0.296 on average) and R_v coefficient (P = 0.082 to P = 0.228 on average), while models only remained significant when using the partial Mantel test. This highlights an important point, the partial Mantel test is likely more sensitive to high type I error and could be the reason why significance did not change.

Post-hoc analysis sensitivity analysis was performed using haplotype frequencies (tb-RDA) and G'_{ST} with the circuit theory connectivity model and rerunning the GARM tool. For each test, important populations were identified by comparative drops in correlation from results achieved using all populations (Table 5-3). For haplotype frequency models, the population identified was SR (Shoshone River) and for G'_{ST} models, it was BC (Bench Corral). The analysis was redone using GARM and the results overlapped (visually) with previous runs using all populations. For G'_{ST} models, removing the BC population showed high overlap of 90% compared to using all populations (Figure 5-2). Removing SR for haplotype frequency corridor models showed a much larger decrease where only 46% of models overlapped with the previous test using all populations. This overlap only slightly improved (56%) when using the top 10 models. Comparatively, we found the overlap between haplotype frequencies (tb-RDA) and G'_{ST} to be around 66% (Figure 5-2).

Population Removed	Mantel (r-coef)	Mantel (p-value)	Partial (r-coef)	Partial Mantel (p-value)	R_v coef.	R_v (p-value)
least-cost path with $\mathbf{G'_{ST}}$						
All Models	**0.242**	**0.053**	**0.450**	**0.002**	**0.455**	**0.082**
BC	0.055	0.296	0.284	0.006	0.393	0.228
MF	0.311	0.018	0.519	0.001	0.513	0.027
Circuit theory with $\mathbf{G'_{ST}}$						
All Models	**0.353**	**0.013**	**0.440**	**0.005**	**0.622**	**0.034**
BC	0.193	0.098	0.283	0.037	0.575	0.122

Population Removed	tb-RDA (adj. R^2)	tb-RDA (pvalue)
tb-RDA with least-cost path		
All Models	**0.650**	**0.022**
CY	-0.022	0.514
MF	0.025	0.489
NR	0.034	0.514
tb-RDA with Circuit theory		
All Models	**0.805**	**0.009**
SR	-0.097	0.500

Table 5-3 Sensitivity analysis results using population leave-one-out tests for 19 elk populations in the GYE. The first column lists the population left out of each test. Values in bold along the "**All models**" row depict the average value for each statistical test using all populations for the top models (as chosen by the respective statistical test). Other values represent the average value when tested against the top models and leaving out the population listed in the first column (e.g. the Mantel column reports only for the top models when sorted by the Mantel r-coefficient).

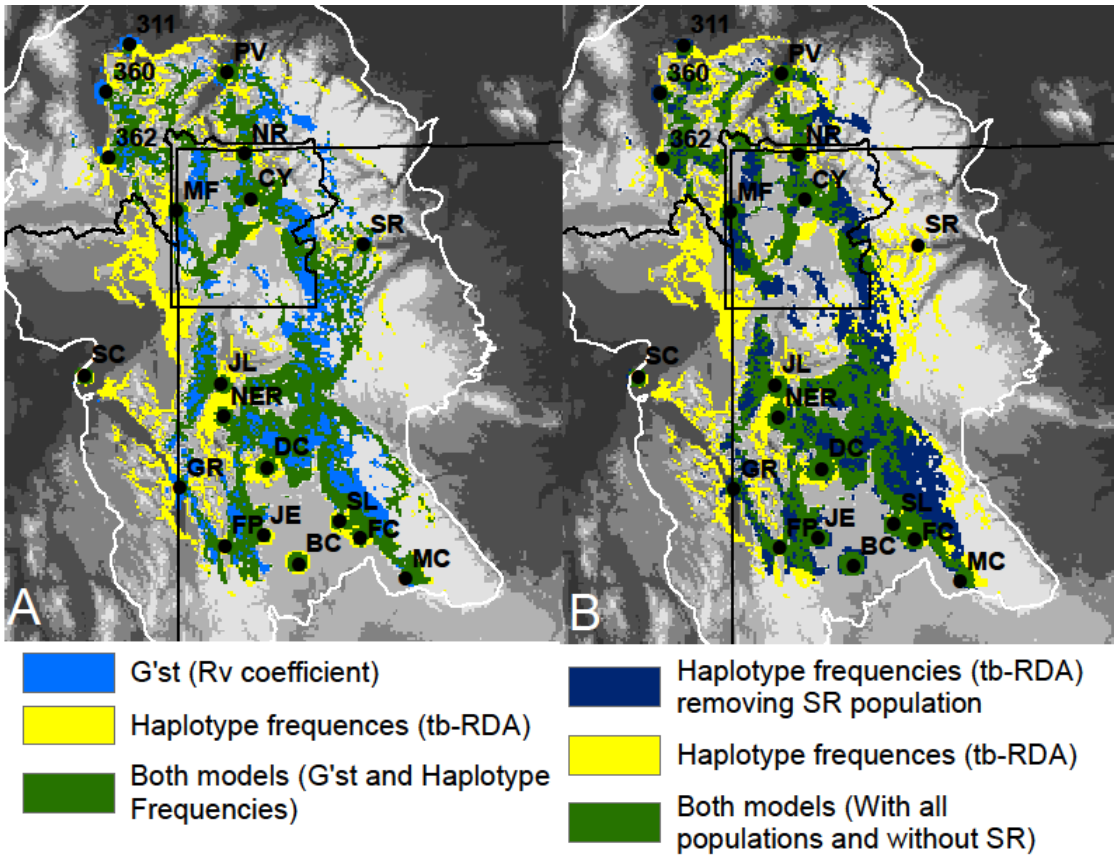

Figure 5-2 Path overlap for circuit theory models. Illustrates differences in results between-genetic distance metrics (G'_{ST} and raw haplotype frequencies with tb-RDA) and sensativity analysis using a single genetic distance metric (raw haplotype frequencies with all populations and removing the SR population). The top models from each method (e.g., G'_{ST} and raw haplotype frequencies) were compared from each test using the top 10% most connected areas. The values leftover after removing the other 90% were then overlapped to produce presence/absence connectivity maps. A) Overlap test between the G'_{ST} (yellow) and raw haplotype frequencies with tb-RDA (blue) and B) overlap when performing sensitivity analysis using haplotype frequencies with all populations (yellow, and the same test represented in panel A), and haplotype frequencies removing population SR (dark blue). Areas of overlap between models are colored as dark green.

5.5 Discussion

5.5.1 Uncertainty assessments identify important features with biological implications

In total we considered >300,000 potential resistance surface models; a number far greater than all previous landscape genetics studies on modeling corridors. For example, Wang *et al.* (2009) used three variables with limited optimization to produce 24,843 models. Only strong relationships between landscape resistance and genetic distance (gene flow) were likely to have a strong or consistent signal among several computational methods, which can help provide certainty in modeling genetic connectivity. This was illustrated by forest being represented as the lowest weighted landscape feature in terms of resistance to gene flow for all 14 method combinations. In contrast, resistance surface fitting results from different methods varied considerably in identifying low-weighted elevation classes. Generally, the amount of variation (confidence interval width) of resistance weights tended to increase as the resistance weights increased.

It was encouraging (and surprising) that our most highly correlated corridor models (using tb-RDA with raw haplotype frequencies) indicated that landscape features explain over 80% of genetic variation using a circuit theory connectivity model. This is a high percentage compared to most landscape genetics studies and suggests one can use landscape features (e.g., forest cover) to model or predict elk gene flow. There was almost no evidence that linear (Euclidian distance) structures genetic variation in that there was zero correlation when testing for isolation-by-distance using tb-RDA (R^2 = 0.002, P = 0.4) and G'_{ST} (r = -0.011, P=0.53). Overall, the lack of isolation-by-distance and strong landscape signals suggest that landscape strongly shapes maternal gene flow in elk of the GYE.

Elevation was found to be important, but the elevation class most conducive to gene flow was partly dependent on the genetic distance metric. Our results from G'_{ST} and haplotype frequencies (tb-RDA) both indicated low resistance weights for the high elevation class (2447-4105m). Very roughly, one might interpret this to indicate that summer movement is most prevalent in the gene flow signal. This result makes biological sense because summer ranges are less distinct than winter ranges, with more mingling among populations. It is also relevant because elk rutting season takes place during early fall months and in elk transitional habitats as they migrate to winter range. Haplotype frequencies, however, also identified mid-range elevation (1827-2117m) to be low resistance to gene flow, which fits elk transitional range (during migration), or the high range preferred elevations for winter movement.

For female elk in the GYE, forest cover was a good predictor of gene flow across all combinations. It is important to note, however, that other biological factors and behaviors are likely masked by this result because this study considered relatively few major and widely distributed landscape features. For instance, forest cover might also be related to hunting or predator pressure, and other dispersal behavioral patterns (e.g. seasonal migration). It is difficult to identify a single or absolute landscape resistance weight when correlating landscape features to gene flow because many landscape features are inherently spatially auto-correlated. For example, forest is likely a surrogate for public land, and therefore, less development. This leads to greater uncertainty in the weight ranges of more heavily weighted landscape features because their impact results from multiple interacting (but hidden) features (Cushman, Wasserman, *et al.* 2013).

Additionally, gene flow is a highly stochastic process dependent not only on dispersal, but on mating, drift and inheritance (Graves et al. 2013).

5.5.2 Sensitivity analysis

In some cases, the amount of uncertainty due to removing a single population was as high as or higher than uncertainty due to the methodological approach (statistical tests, connectivity model and genetic distance metrics). This was illustrated by a larger disagreement in overlap between corridor models produced by a single genetic distance metric (haplotype frequencies) than between genetic distance metrics (haplotype frequencies vs. G'_{ST}). In post-hoc testing, when removing the Shoshone River (SR) population, approximately 56% of areas of high genetic connectivity overlapped in resistance models optimized without the SR population vs. resistance models optimized with all populations. In comparison, the same test performed using G'_{ST} and removing the Bench Corral (BC) population (based on results sensitivity analysis) lead to a 90% agreement in areas of high genetic connectivity (Table 5-3). An additional comparison between differing methodological approaches (haplotype frequencies vs. G'_{ST}) showed a 66% agreement when optimized using all populations.

For all method combinations, it was found removing even one population caused top model correlation values to drop. Interestingly, in some cases correlation of the top models was even found to reverse sign (Table 5-3). A few of these influential populations reoccurred in sensitivity analysis tests, such as the BC and Madison Firehole (MF), but many sensitivity analysis tests identified different populations as being influential (Table 5-3). The BC population strongly influenced methods using G'_{ST} with both connectivity models, while the MF population showed up in both G'_{ST} with least-cost path method and using haplotype frequencies with least-cost paths. When the BC population was dropped, models became much less significant (e.g., considering the least-cost path connectivity model, P =0.053 to P = 0.296 for a Mantel test, and P =0.013 dropped to P = 0.098 using a circuit theory model; Table 5-3). Results were mixed for sensitivity analysis tests for MF. For example, when using G'_{ST} with least-cost path, correlation values improved (Mantel r increased from 0.242 to 0.311), however, when using haplotype frequencies and least-cost paths, models become greatly non-significant (R^2 = 0.025, P = 0.489).

The BC population is an elk winter feedground in the Southern portion of the GYE, and located at upper-middle elevation (2118-2446m) in an area of non-forested habitat. Potentially, BC is more influential in model selection because all other populations are in or near the presence of forest cover (with the exception of Sand Creek; SC). Pairwise G'_{ST} values for the MF population included most of the highest values for all population pairs (ranging from 0.095-0.716, an average of 0.435, or approximately double the average of 0.249 for all pairs). This was expected as the MF population is likely a small, isolated population of elk. In general, sensitivity analysis suggested potential population subsets of connectivity and particular sensitivities in the combinations of methods. Though a powerful tool, sensitivity analysis needs further development as a tool in landscape genetics that allows consistent predictions of areas of genetic connectivity and the associated variable weights (Manel & Holderegger 2013).

5.5.3 Optimizing on correlation vs. model significance

Optimizing on the correlation coefficient, rather than on model significance (i.e. P values), was potentially a more reliable way to predict and compare top resistance

surfaces because of better weight consistency. Variation in the resulting corridor model paths was more likely due to differences in genetic distance metrics or the chosen connectivity model. Ranking by significance, however, would have clouded corridor model choice by greatly increasing the weight ranges identified by top models (results not shown). This is especially true for the partial Mantel test that had a large proportion (several thousands) of equally significant resistance surfaces. The large number of equally significant corridor models reported was likely due to an increase in type I error rates. On the other hand, the R_v coefficient might be a more useful measure for the purpose of identifying significant models. Over all possible combinations, the number of significant models reported from the R_v coefficient was much less. The R_v coefficient might prove to be a much more stringent and useful for model significance testing, but it requires further testing using a full simulation comparison study.

5.5.4 Uncertainty related to the choice of genetic distance

Corridors identified by G'_{ST} and haplotype frequencies (tb-RDA) were similar. The only minor difference was the low weighting of mid-range elevation (1827-2117m) identified by haplotype frequencies (tb-RDA). Uncertainty related to the choice of genetic distance matrix in resistance surface fit (and subsequent corridor mapping) was most pronounced between G'_{ST} and F_{ST}. There was reasonable overlap (~66% of the total area) between models produced using haplotype frequencies (tb-RDA) and G'_{ST} combined with circuit theory (Figure 5-2). This result also did not differ much when comparing the top 10 models (67% overlap) vs. the top 50 models (66% overlap).

There was no correlation between the population pairwise genetic distance matrices for F_{ST} and G'_{ST} ($r = -0.06$). This low correlation was likely due to the large bias produced by high within-population haplotype diversity that constrained the range (maximum value) of F_{ST} vs. the less constrained G'_{ST}. G'_{ST} is a standardized measure allowing for better representation of true genetic differentiation or distance (Hedrick 2005; Jost 2008). This illustrates the importance of comparing results from different genetic distance metrics when assessing uncertainty. F_{ST} should not be used if genetic diversity is high within populations. Results for F_{ST} are mostly unreported and instead serve as important illustration of the potential effects of using different or inappropriate distance metrics.

5.5.5 Uncertainty related to the choice of connectivity model

The uncertainty related to the choice of connectivity model was less than that produced by other factors like the statistical test or the genetic distance metric used. For example, using haplotype frequency, both connectivity models were tightly grouped for landscape variables conducive to gene flow (forest cover, high elevation [2447-4105m] and mid-elevation [1827-2117m]; Figure 5-3). In comparison, mid-elevation (1827-2117m) was the variable of major disagreement between G'_{ST} and haplotype frequency tests. Though it should be noted, it is not certain if this difference can be attributed to the genetic distance metric or the statistical test used because both differed. For methods using haplotype frequencies (tb-RDA) and G'_{ST}, circuit theory connectivity models had higher correlation values between statistical tests, though there was still much greater variation between statistical tests than between connectivity models (Table 5-2).

5.5.6 Modeling inconsistencies

Among all methods, the most inconsistences were in corridor models predicted using the least-cost path connectivity model with the tb-RDA statistical test. Two distinct groups of corridor models were identified using least-cost path based on model spatial patterns and resistance surface weight values. We further separated out these corridor models to better compare the results from other methods by creating two groups of corridor models based on high and low-weighted forest. Low-weighted forest corridor models were more in agreement with all previous results using other combinations of methods, and when using tb-RDA with a circuit theory connectivity model. For this reason, all high-weighted forest cover models were removed from analysis and the main text only reports on low-weighted models. Only the top 10 models were kept because of the large variation over the top 50 low-weighted models (results not shown). The inconsistencies when using the tb-RDA statistical test with the least-cost path connectivity model further stressed the importance of testing multiple and differing methods when the goal is predicting genetic connectivity.

5.5.7 Conclusion

Here, a framework was illustrated to assess landscape genetic connectivity model uncertainty by testing effects of statistical tests, genetic distance metrics, and corridor modeling approaches on resistance surface fit to genetic data. Additionally, we conducted sensitivity analysis by leaving out a single population at a time and rerunning analysis. Results suggest it is plausible to identify landscape features strongly correlated with gene flow consistently across resistance surfaces and independent of the combination of methods used. For example, forest cover was a good predictor of maternal elk gene flow in the GYE for each of 14 combinations of methods (statistics, distance metrics, and connectivity models). The circuit theory model had reasonable agreement of highly connected areas with > 60% of corridors overlapping in highly connected areas among comparisons of haplotype frequencies (tb-RDA) and G'_{ST}. Sensitivity analysis is currently underutilized in landscape genetics studies, but will aid in further identifying where model predictions or corridors change significantly based solely on using a subset of the data.

In summary, future landscape genetic studies should adopt a more exhaustive uncertainty testing framework for identifying corridors and landscape resistance features influencing connectivity. This framework should include comparisons of thousands of resistance surfaces and a variety of methods (statistical tests, genetic distances, and connectivity modeling approaches). For this purpose, GARM is offered as a powerful new tool to optimize the resistance map creation procedure by allowing the automated searching among an intractable number of possible resistance surface maps (to the order of 10^{16} models in this simple study). Use of extensive uncertainty testing will improve confidence in corridor (and barrier) identification for conservation and make landscape genetics a more rigorous scientific discipline. Uncertainty testing will also help prevent waste of limited conservation resources by ensuring protection of the most important areas and landscape features to maintain connectivity and biodiversity.

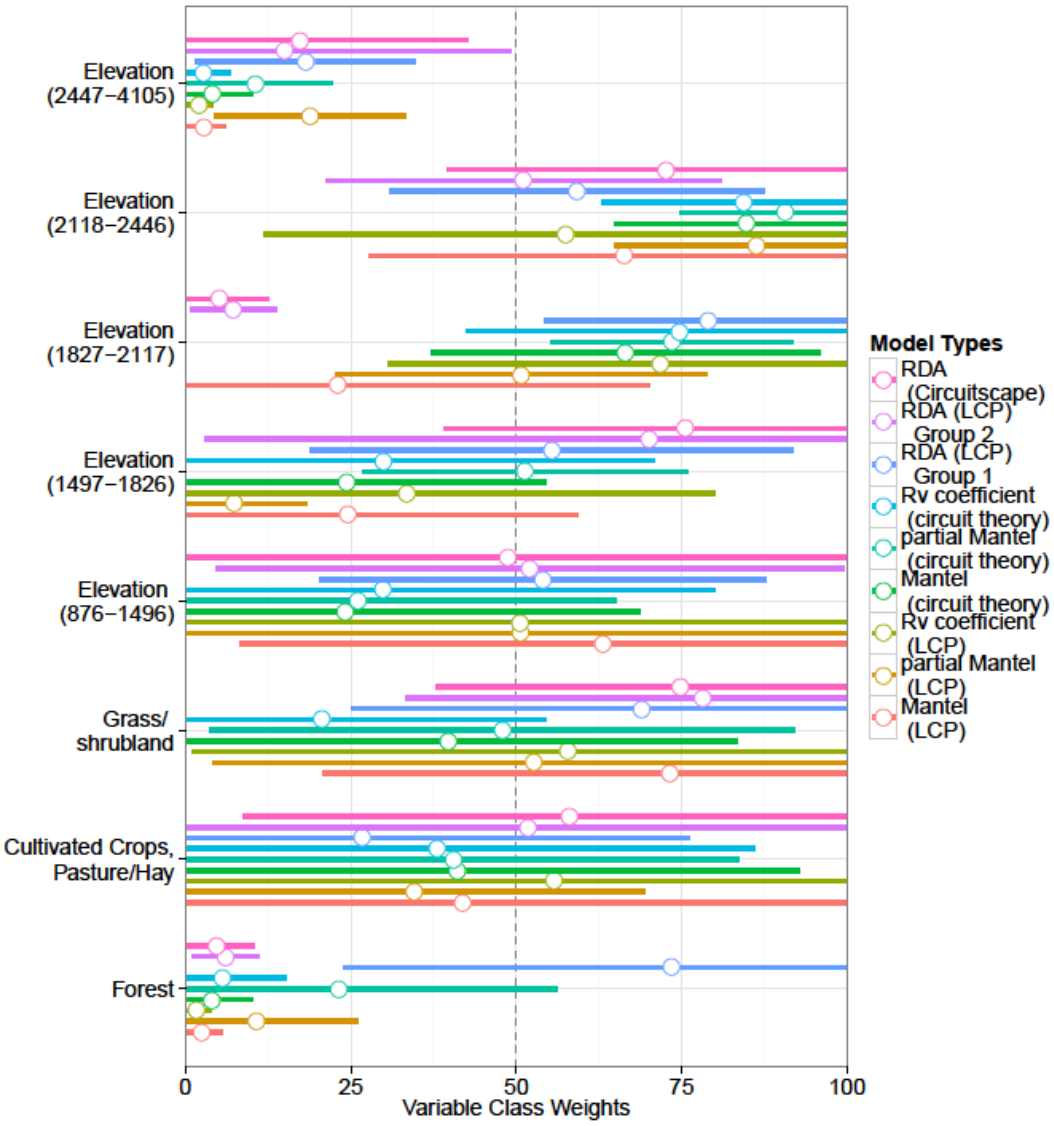

Figure 5-3 Plot of weights with standard error bars for the top models. Points are means; lines represent the 95% confidence intervals. Model types are in the legend on the right with the connectivity model used in parenthesis. Weights have the potential to range from 1-100 (x-axis) with class variable names for each weight listed on the left. Open water is not shown because it was set to 100 in all cases.

Chapter 6 Conclusions

6.1 Summary of Contributions

Chapter 2 compared the relative rates of gene flow between male and female elk in the GYE (Hand, Chen, *et al.* 2013). Identifying sex-biased dispersal is important for systems like the GYE where female elk are solely responsible for the transmission of Brucellosis. There was also no detectable isolation-by-distance as the mediating factor for gene flow in female elk in the GYE. The results in Chapter 2 lead to the conclusion that a full landscape genetics investigation was needed to identify landscape features impacting maternal gene flow in the GYE.

When genetic data isn't available as it was in chapter 2 to conduct gene flow analysis, a more theoretical approach is needed. As an alternative to the parameterization of resistance maps from genetic data, Chapter 3 took a broadly inclusive approach to investigate several species over the extent of the Western United States (Hand, Landguth, *et al.* 2013). Resistance maps were produced based on hypothetical "pristine" and human footprint scenarios. This allowed the results to be widely applicable to several species with varied life-history traits to do with dispersal ability and population size. Chapter 3 also addresses the impact of anthropogenic influences on habitat loss and fragmentation. This chapter, however, does not answer the need for when empirical genetic data is available and should be used in resistance surface parameterization.

Chapter 4 answered the need for a novel software program for optimization of resistance map creation. Optimization is useful for when one wishes to parameterize resistance surfaces based on empirical genetic data (Hand, Raiford, Landguth, *et al.* 2013). Full optimization has been an important omission from many previous landscape genetics studies because it is a computationally intensive process. The GARM tool greatly benefits any landscape genetics study in efficiently searching for top resistance surfaces over intractable search spaces.

Chapter 5 expands upon the initial work in chapter 2 to include several more elk populations towards a full landscape genetics treatment of maternal gene flow in the GYE (Hand, Raiford, Chen, *et al.* 2013). Chapter 5 also employs the GARM tool from chapter 4 to perform a cross-method comparison of four statistical tests, three metrics of genetic distance and two connectivity models. It is one of the first studies in the landscape genetics field to cover such a wide assortment of methods and in such great depth (searching over 300,000 resistance surfaces). From the results it was possible to draw on some general observations about future landscape genetics studies and the importance of considering measures of genetic distance carefully. It also used sensitivity analysis (leave-one-out population), another novel approach that has not been explored fully in any landscape genetics study. The rigorous framework presented in chapter 5 establishes guidelines for future landscape genetics studies by highlighting the great need to explore several different methods and to use some kind of sensitivity analysis approach. The approach outlined was helpful in finding that there is potential (and hope) to consistently identify strong relationships between landscape features and gene flow.

6.2 Future Research

6.2.1 Population vs. Individual methods

The major work of this dissertation has used population based methods to measure genetic distance (e.g. use of F_{ST} in elk populations). One of the great promises of landscape genetics is to include individual-based approaches. There has been a lack of comparison between individual and population based methods in connectivity modeling (e.g. Balkenhol *et al.* (2009) performed population based simulations vs. Graves et al. (2013) that used individual based methods). It is important to explore the differences in modeling individual vs. population connectivity because very different landscape patterns could be influential at various levels of investigation. Also, it is not well known if some methods will work appropriately for both population and individual approaches. This is an important and yet little-explored area in landscape genetics.

6.2.2 Multi-species study

Different species offer very different life-history traits (dispersal ability, population sizes, spatial structure, seasonal migration, etc.) that are all important in predicting gene flow. The work presented herein only focused on a single species, elk. Additional work would be to generate simulated gene flow data from varied life-history parameters to represent several different species to be used in a pilot study. Such a study would compare several methods much like Chapter 5 to investigate if a single method worked better in most cases and under several different combinations of life-history traits. After a simulation study, the results could be applied to empirical data sets.

6.2.3 Landscape Genomics

Manel & Holderegger (2013) point out that to date, landscape genetics and landscape genomics has been separated mostly in the aim and methods used. For the most part, the aim of current landscape genomics approaches has been different from those of landscape genetic studies. Most landscape genetic studies try to relate genetic distance to features on the landscape to better understand how animals *travel* between populations or individuals. Landscape genomics is often more focused on how animals *adapt* to environmental conditions based on the environment at individual or population geographical locations

For example, landscape genetics focuses on connectivity models to use varying landscape gradients to better explain animal movement (Cushman *et al.* 2009). Connectivity models take into account the landscape between points of interest (individuals or populations). On the other hand, landscape genomics approaches use site-based measures (environmental factors like precipitation, levels of humidity, disease prevalence, etc.) at points of interest to relate environment to observed allele frequencies (Joost *et al.* 2007; Coop *et al.* 2010; Hancock *et al.* 2011). Landscape genomics methods have not attempted to use connectivity model approaches and instead focus on more consistent statistical methods (lower type I and II errors) when comparing environment to

genetic structure (raw allele frequencies). There have been several recent simulation papers exploring the power and error rates of several different methods in the landscape genomics literature (De Mita *et al.* 2013; Frichot *et al.* 2013).

Lacking is a general framework that pulls together the best practices of landscape genomics into landscape genetics. There has been much recent development of statistical theory for better prediction of genetic structure due to environmental factors in landscape genomics. Some of the methods used could greatly benefit landscape genetics work and improve current methods. The major bottleneck, currently, is computational time. Connectivity models (least-cost paths, circuit theory, and resistant kernels) are already computationally intensive to run, and so are the statistical methods used in landscape genomics (e.g. latent factor mixed modeling that combines likelihood estimates with latent factors representing the unknown contributions of demographic history among other immeasurable factors).

A general framework would attempt to combine current genomic techniques (e.g. like F_{ST} outlier detection using latent factor mixed modeling) to identify SNPs under selection with connectivity models (circuit theory or least-cost corridors, or both). An important step is to separate out neutral vs. loci under selection. One could then test the robustness of predicting patterns of gene flow using neutral loci. A major benefit of using genomic data vs. tens of microsatellites is the ability to perform advanced sensitivity analysis using the SNPs themselves to account for genetic variation among loci. Several thousand SNPs could be split into several sensitivity analysis sets. A single set is used to train the connectivity model, and the remaining sets are used to test the prediction accuracy of the connectivity model. Sensitivity analysis helps in identifying those connectivity routes or resistance weights that are consistently predicted across sensitivity analysis sets. This connectivity model based on neutral markers can be used as a powerful null hypothesis against those loci under selection can be tested either individually or as sets. Such sets of adaptive loci would be chosen relative to some known function, like a group of disease related genes. The approach of using sets of loci helps alleviate variance due to noise in the genetic signal from using a single locus. Whether routes for genes under selection vs. neutral genes would differ is left to an additional study to explore, but this method would make answering that question possible. In addition, this study could incorporate much more sophisticated methods of contemporary gene flow like assignment tests (Lowe & Allendorf 2010).

These more precise methods, along with extensive sensitivity analysis testing, would give much greater confidence in model predictions. This is a much needed contribution in the field of landscape genetics/genomics. Currently, there is a lack of confidence that landscape genetics models are meaningful (due to papers like Cushman, Wasserman, *et al.* 2013; Graves *et al.* 2013). There is a great need to move past faulty partial Mantel tests and F_{ST} measures to better employ the power of genomic data sets to fine-tune connectivity model predictions.

Bibliography

Abdi H (2007) RV coefficient and congruence coefficient. In: *Encyclopedia of Measurement and Statistics* (ed Salkind N), pp. 849–853. Sage Publication, Thousand Oaks (CA).

Adriaensen F, Chardon JP, De Blust G *et al.* (2003) The application of "least-cost" modelling as a functional landscape model. *Landscape and Urban Planning*, **64**, 233–247.

Allendorf FW, Luikart G, Aitken SN (2013) *Conservation and the genetics of populations.* Wiley-Blackwell.

Balkenhol N, Waits LP, Dezzani RJ (2009) Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography*, **32**, 818–830.

Bandelt HJ, Forster P, Röhl a (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular biology and evolution*, **16**, 37–48.

Beier P, Noss RF, Biology C, Dec N (1998) Do Habitat Corridors Provide Connectivity ? *Conservation biology*, **12**, 1241–1252.

Beja-Pereira A, Bricker B, Chen S *et al.* (2009) DNA genotyping suggests that recent brucellosis outbreaks in the Greater Yellowstone Area originated from elk. *Journal of wildlife diseases*, **45**, 1174–7.

Bird C, Karl S, Smouse P, Toonen R (2011) Detecting and measuring genetic differentiation. In: *Phylogeography and Population Genetics in Crustacea* (ed Schubart C), pp. 31–55. CRC Press 2011.

Bonnet E, de Peer Y Van (2002) zt: A Sofware Tool for Simple and Partial Mantel Tests. *Journal of Statistical Software*, **7**, 1–12.

Bowcock AM, Ruiz-Linares A, Tomfohrde J *et al.* (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, **368**, 455–457.

Boyce MS, Hayden-Wing LD (1979) *North American Elk: Ecology, behavior and management.* University of Wyoming, Laramie, Wyoming.

Boyce MS, Mao JS, Merrill EH *et al.* (2003) Scale and heterogeneity in habitat selection by elk in Yellowstone National Park. *Ecoscience*, **10**, 421–431.

Cheville NF, McCullough DR, Paulson LR (1998) *Brucellosis in the Greater Yellowstone Area*. National Academy Press, Wasington, D.C. USA.

Cobb HG (1993) Genetic algorithms for tracking changing environments. In: *Proceedings of the Fifth International Conference on Genetic Algorithms* , pp. 523–530. Morgan Kaufmann.

Compton BW, McGarigal K, Cushman S a, Gamble LR (2007) A resistant-kernel model of connectivity for amphibians that breed in vernal pools. *Conservation biology : the journal of the Society for Conservation Biology*, **21**, 788–99.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics*, **185**, 1411–23.

Cooper JD, Waser PM, Gopurenko D *et al.* (2010) Measuring sex-biased dispersal in social mammals: Comparisons of nuclear and mitochondrial genes in collared peccaries. *Journal of Mammalogy*, **91**, 1413–1424.

Crawford NG (2010) Smogd: Software for the Measurement of Genetic Diversity. *Molecular ecology resources*, **10**, 556–7.

Crooks K, Sanjayan M (2006a) Connectivity conservation: maintaining connections for nature. In: *Connectivity Conservation* , pp. 1–20.

Crooks KR, Sanjayan A (2006b) *Connectivity Conservation*. Cambridge University Press.

Cross PC, Cole EK, Dobson AP *et al.* (2010) Probable causes of increasing brucellosis in free-ranging elk of the Greater Yellowstone Ecosystem. *Ecol Appl*, **20**, 278–288.

Cross P, Heisey D, Scurlock BM *et al.* (2010) Mapping Brucellosis Increases Relative to Elk Density Using Hierarchical Bayesian Models. *PLoS ONE*, **5**, e10322.

Cullingham CI, Merrill EH, Pybus MJ *et al.* (2011) Broad and fine-scale genetic analysis of white-tailed deer populations: estimating the relative risk of chronic wasting disease spread. *Evolutionary Applications*, **4**, 116–131.

Cushman SA, Compton BW, Mcgarigal K (2010) Habitat fragmentation effects depend on complex interactions between population size and dispersal ability: Modeling influences of roads, agriculture and residential development across a range of life-history characteristics. In: *Spatial complexity, informatics, and wildlife conservation.* (eds Cushman SA, Huettmann F), pp. 369–387. Springer New York, Tokyo.

Cushman SA, Landguth EL (2012a) Ecological Associations , Dispersal Ability , and Landscape Connectivity in the Northern Rocky Mountains. *Research Paper*, 21.

Cushman S a., Landguth EL (2012b) Multi-taxa population connectivity in the Northern Rocky Mountains. *Ecological Modelling*, **231**, 101–112.

Cushman SA, Landguth EL, Flather C (2011) *Climate change and connectivity: assessing landscape and species vulnerability.*

Cushman SA, Landguth EL, Flather CH (2012) Evaluating the sufficiency of protected lands for maintaining wildlife population connectivity in the U.S. northern Rocky Mountains. *Diversity and Distributions*, **18**, 873–884.

Cushman S a., Lewis JS (2010) Movement behavior explains genetic differentiation in American black bears. *Landscape Ecology*, **25**, 1613–1625.

Cushman S, McKelvey KS, Hayden J, Schwartz MK (2006) Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *The American naturalist*, **168**, 486–499.

Cushman S a, McKelvey KS, Schwartz MK (2009) Use of empirically derived source-destination models to map regional conservation corridors. *Conservation biology : the journal of the Society for Conservation Biology*, **23**, 368–76.

Cushman S, McRae B, Adriaensen F *et al.* (2013) Biological corridors and connectivity. In: *Key Topics in Conservation Biology 2* (eds Macdonald DW, Willis KJ), pp. 384–404. John Wiley & Sons, Oxford.

Cushman SA, Shirk AJ, Landguth EL (2013) Landscape genetics and limiting factors. *Conservation Genetics*, **14**, 263–274.

Cushman S, Wasserman T, Landguth E, Shirk A (2013) Re-Evaluating Causal Modeling with Mantel Tests in Landscape Genetics. *Diversity*, **5**, 51–72.

Dale VH, Joyce LA, McNulty S, Neilson RP (2000) The interplay between climate change, forests, and disturbances. *Science of the Total Environment*, **262**, 201–204.

Debinski DM, Holt RD (2000) A Survey and Overview of Habitat Fragmentation Experiments. *Conservation Biology*, **14**, 342–355.

Dennis Jr. JE, Schnabel RB (1983) Numerical methods for unconstrained optimization and nonlinear equations. *Numerical methods for unconstrained optimization and nonlinear equations*, xiii+378 pp.

Dijkstra EW (1959) A note on two problems in connexion with graphs. *Numerische Mathematik*, **1**, 269–271.

Dray S, Dufour AB (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**, 1–20.

Driezen K, Adriaensen F, Rondinini C, Doncaster CP, Matthysen E (2007) Evaluating least-cost model predictions with empirical dispersal data: A case-study using

radiotracking data of hedgehogs (Erinaceus europaeus). *Ecological Modelling*, **209**, 314–322.

Epps CW, Wehausen JD, Bleich VC, Torres SG, Brashares JS (2007) Optimizing dispersal and corridor models using landscape genetics. *Journal of Applied Ecology*, **44**, 714–724.

Esri (2011) ArcGIS Desktop: Release 10. *Redlands CA*.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources*, **10**, 564–567.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application. *Genetics*, **491**, 479–491.

Fahrig L (2003) Effects of Habitat Fragmentation on Biodiversity. *Annual Review of Ecology Evolution and Systematics*, **34**, 487–515.

Frankham R, Ballou JD, Briscoe DA (2010) *Introduction to Conservation Genetics*. Cambridge University Press.

Freemark KE, Merriam HG (1986) Importance of area and habitat heterogeneity to bird assemblages in temperate forest fragments. *Biological Conservation*, **36**, 115–141.

Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models. *Molecular biology and evolution*, **30**, 1687–1699.

Garroway CJ, Bowman J, Carr D, Wilson PJ (2008) Applications of graph theory to landscape genetics. *Evolutionary Applications*, **1**, 620–630.

Garroway CJ, Bowman J, Wilson PJ (2011) Using a genetic network to parameterize a landscape resistance surface for fishers, Martes pennanti. *Molecular ecology*.

Gesch D, Oimoen M, Greenlee S *et al.* (2002) The National Elevation Dataset (D Maune, Ed,). *Photogrammetric Engineering and Remote Sensing*, **68**, 5–11.

Goldberg DE (1989) *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley, Ed,). Addison-Wesley.

Graves T a, Beier P, Royle JA (2013) Current approaches using genetic distances produce poor estimates of landscape resistance to interindividual dispersal. *Molecular ecology*, 1–16.

Guillot G, Rousset F (2013) Dismantling the Mantel tests. *Methods Ecol Evol*, **4**, 336–344.

Halbert ND, Gogan PJP, Hedrick PW, Wahl JM, Derr JN (2012) Genetic population substructure in bison at Yellowstone National Park. *The Journal of heredity*, **103**, 360–70.

Hames RS, Rosenberg K V., Lowe JD, Dhondt A a. (2008) Site reoccupation in fragmented landscapes: testing predictions of metapopulation theory. *Journal of Animal Ecology*, **70**, 182–190.

Hancock AM, Witonsky DB, Alkorta-Aranburu G *et al.* (2011) Adaptations to Climate-Mediated Selective Pressures in Humans. *PLoS Genet*, **7**, e1001375+.

Hand BK, Chen S, Anderson N *et al.* (2013) Limited Maternal Gene Flow amongst Elk in the Greater Yellowstone Ecosystem Revealed by Mitochondrial DNA. *Journal of Fish and Wildlife Management - Accepted*.

Hand BK, Landguth EL, Cushman SA (2013) The influence of dispersal ability, population size and the human footprint on habitat fragmentation and loss in the Western United States. *In prep.*

Hand BK, Raiford D, Chen S *et al.* (2013) A new, more-exhaustive landscape genetics approach: corridor identification for elk across the Greater Yellowstone Ecosystem. *In prep.*

Hand BK, Raiford D, Landguth EL, Glassy J (2013) GARM: A machine learning algorithm for creating resistance maps in landscape genetics. *In prep.*

Hanski I (1998) Metapopulation dynamics. *Nature*, **396**, 41–49.

Hedrick PW (2005) A Standardized Genetic Differentiation Measure. *Evolution*, **59**, 1633–1638.

Hedrick PW, Allendorf FW, Baker CS (2013) Estimation of Male Gene Flow from Measures of Nuclear and Female Genetic Differentiation. *Journal of Heredity*.

Heo M, Gabriel KR (1998) A permutation test of association between configurations by means of the RV coefficient. *Communications in Statistics Simulation and Computation*, **27**, 843–856.

Hicks JF, Rachlow JL, Rhodes OE, Williams CL, Waits LP (2007) Reintroduction and Genetic Structure: Rocky Mountain Elk in Yellowstone and the Western States. *Journal of Mammalogy*, **88**, 129–138.

Hijmans RJ, van Etten J (2013) raster: raster: Geographic data analysis and modeling.

Holland JH (1975) *Adaptation in Natural and Artificial Systems* (JH Holland, Ed,). University of Michigan Press.

Homer C, Dewitz J, Fry J *et al.* (2007) Completion of the 2001 National Land Cover Database for the Conterminous United States. *Photogrammetric Engineering and Remote Sensing*, **73**, 337–341.

Houston DB (1982) *The Northern Yellowstone Elk : Ecology and Management.* Macmillan, New York, NY.

Joost S, Bonin a, Bruford MW *et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular ecology*, **16**, 3955–69.

Jost L (2008) GST and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Keitt TH, Urban DL, Milne BT (1997) Detecting critical scales in fragmented landscapes. *Conservation Ecology*, **1**, 4.

Keller LF, Waller DM (2002) Inbreeding effects in wild populations. , **17**, 19–23.

Laiolo P, Tella JL (2006) Landscape bioacoustics allow detection of the effects of habitat patchiness on population structure. *Ecology*, **87**, 1203–1214.

Landguth EL, Hand BK, Glassy J, Cushman SA, Sawaya MA (2012) UNICOR: a species connectivity and corridor network simulator. *Ecography*, **35**, 9–14.

Legendre P, Fortin M-J (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular ecology resources*, **10**, 831–44.

Legendre P, Legendre L (2012) *Numerical Ecology*. Elsevier Science.

Levins R (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America*, **15**, 237–240.

Lowe WH, Allendorf FW (2010) What can genetics tell us about population connectivity? *Molecular ecology*, **19**, 3038–51.

Luikart G, Gielly L, Excoffier L *et al.* (2001) Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 5927–5932.

Manel S, Holderegger R (2013) Ten years of landscape genetics. *Trends in Ecology & Evolution*, 1–8.

Manel S, Joost S, Epperson BK *et al.* (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology*, **19**, 3760–3772.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189–197.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer research*, **27**, 209–20.

Martin K, Stacey PB, Braun CE (2000) Recruitment , dispersal , and demographic rescue in spatially-structured white-tailed ptarmigan populations. *The Condor*, **102**, 503–516.

Maudet C, Luikart G, Dubray D, Von Hardenberg A, Taberlet P (2004) Low genotyping error rates in wild ungulate faeces sampled in winter. *Molecular Ecology Notes*, **4**, 772–775.

McGarigal K, Cushman SA (2002) Comparative Evaluation of Experimental Approaches To the Study of Habitat Fragmentation Effects. *Ecological Applications*, **12**, 335–345.

McGarigal K, Cushman SA, Ene. E (2013) FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst.

McRae B (2006) Isolation by resistance. *Evolution*, **60**, 1551–1561.

McRae BH, Beier P (2007) Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 19885–90.

McRae BH, Dickson BG, Keitt TH, Shah VB (2008) Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology*, **89**, 2712–2724.

Meirmans PG, Hedrick PW (2011) Assessing population structure: F(ST) and related measures. *Molecular ecology resources*, **11**, 5–18.

De Mita S, Thuillet A-C, Gay L *et al.* (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular ecology*, **22**, 1383–99.

Nei M (1972) Genetic distance between populations. *The American Naturalist*, **106**, 283–292.

Nei M (1973) Analysis of Gene Diversity in Subdivided Populations. *Proceedings of the National Academy of Sciences of the United States of America*, **70**, 3321–3323.

Neilson RP (1993) Vegetation redistribution: A possible biosphere source of CO2 during climatic change. *Water Air Soil Pollution*, **70**, 659–673.

Neilson RP (1995) A model for predicting continental-scale vegetation distribution and water balance. *Ecological Applications*, **5**, 362–385.

Neilson RP, Marks D (1994) A global perspective of regional vegetation and hydrologic sensitivities from climatic change. *Journal of Vegetation Science*, **5**, 715–730.

Nelder JA, Mead R (1965) A Simplex Method for Function Minimization. *The Computer Journal*, **7**, 308–313.

Noss R (1987) Corridors in real landscapes: a reply to Simberloff and Cox. *Conservation biology*, **1**, 159–164.

Oksanen J, Blanchet FG, Kindt R *et al.* (2013) vegan: Community Ecology Package.

Parks SA, McKelvey KS, Schwartz MK (2013) Effects of weighting schemes on the identification of wildlife corridors generated with least-cost methods. *Conservation biology : the journal of the Society for Conservation Biology*, **27**, 145–54.

Peakall R, Smouse P (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, **1**, 6–8.

Peres-Neto PR, Legendre P, Dray S, Borcard D (2006) Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, **87**, 2614–25.

Pérez-Espona S, Pérez-Barbería FJ, Goodall-Copestake WP *et al.* (2009) Genetic diversity and population structure of Scottish Highland red deer (*Cervus canadensis*) populations: a mitochondrial survey. *Heredity*, **102**, 199–210.

Pérez-Espona S, Pérez-Barbería FJ, Jiggins CD, Gordon IJ, Pemberton JM (2010) Variable extent of sex-biased dispersal in a strongly polygynous mammal. *Molecular ecology*, **19**, 3101–13.

Pérez-Espona S, Pérez-Barbería FJ, McLeod JE *et al.* (2008) Landscape features affect gene flow of Scottish Highland red deer (*Cervus canadensis*). *Molecular Ecology*, **17**, 981–996.

Pinto N, Keitt TH (2008) Beyond the least-cost path: evaluating corridor redundancy using a graph-theoretic approach. *Landscape Ecology*, **24**, 253–266.

Polziehn RO, Strobeck C (1998) Phylogeny of wapiti, red deer, sika deer, and other North American cervids as determined from mitochondrial DNA. *Molecular phylogenetics and evolution*, **10**, 249–58.

R Core Team (2013) R: A Language and Environment for Statistical Computing.

Rees EE, Pond B a, Cullingham CI *et al.* (2008) Assessing a landscape barrier using genetic simulation modelling: implications for raccoon rabies management. *Preventive veterinary medicine*, **86**, 107–23.

Riitters K, Wickham J, O'Neill R, Jones B, Smith E (2000) Global-scale patterns of forest fragmentation. *Conservation Ecology*, **4**, 18.

Robbins CS, Dawson DK, Dowell BA (1989) Habitat area requirements of breeding forest birds of the middle Atlantic states. *Wildlife Monographs*, **103**, 1–34.

Robert P, Escoufier Y (1976) A unifying tool for linear multivariate statistical methods: the {RV}-coefficient. *Appl Stat*, **25**, 257–265.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58–62.

Saccheri I, Kuussaari M, Kankare M, Vikman P, Hanski I (1998) Inbreeding and extinction in a butterfly metapopulation. , **45**.

Sawyer SC, Epps CW, Brashares JS (2011) Placing linkages among fragmented habitats: do least-cost models reflect how animals use landscapes? *Journal of Applied Ecology*, **48**, 668–678.

Schnabel RB, Koonatz JE, Weiss BE (1986) A modular system of algorithms for unconstrained minimization. *ACM Transactions on Mathematical Software*, **11**, 419–440.

Schwartz MK, Copeland JP, Anderson NJ *et al.* (2009) Wolverine gene flow across a narrow climatic niche. *Ecology*, **90**, 3222–32.

Shah V, McRae B (2008) Circuitscape: a tool for landscape ecology. *Proceedings of the 7th Python in Science Conference*, 62–65.

Shirk a. J, Cushman S a., Landguth EL (2012) Simulating Pattern-Process Relationships to Validate Landscape Genetic Models. *International Journal of Ecology*, **2012**, 1–8.

Shirk a J, Wallin DO, Cushman S a, Rice CG, Warheit KI (2010) Inferring landscape effects on gene flow: a new model selection framework. *Molecular ecology*, **19**, 3603–19.

Smouse PE, Long JC, Sokal RR (1986a) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, **35**, 627–632.

Smouse PE, Long JC, Sokal RR (1986b) Society of Systematic Biologists Multiple Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence Extensions of the Multiple Regression and Correlation Mantel Test of Matrix Correspondence. *Society*, **35**, 627–632.

Spear SF, Balkenhol N, Fortin M-J, McRae BH, Scribner K (2010) Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Molecular ecology*, 3576–3591.

Spear SF, Peterson CR, Matocq MD, Storfer A (2005) Landscape genetics of the blotched tiger salamander (Ambystoma tigrinum melanostictum). *Molecular Ecology*, **14**, 2553–2564.

Stevens VM, Verkenne C, Vandewoestijne S, Wesselingh R a, Baguette M (2006) Gene flow and functional connectivity in the natterjack toad. *Molecular ecology*, **15**, 2333–44.

Storfer a, Murphy M a, Evans JS *et al.* (2007) Putting the "landscape" in landscape genetics. *Heredity*, **98**, 128–42.

Taylor P, Fahrig L, Henein K, Merriam G (1993) Connectivity is a vital element of landscape structure. *Oikos*, **68**, 571–573.

Thorne ET, Morton JK, Ray WC (1979) Brucellosis, its effect and impact on elk in western Wyoming. In: *North American Elk: Ecology, behavior and management* (eds Boyce MS, Hayden-Wing LD), pp. 212–220. University of Wyoming, Laramie, Wyoming.

Toweill DE, Thomas JW, McCabe RE (2002) *North American elk: Ecology and management.* Smithsonian Institution Press, District of Columbia, Washington.

Vallan D (2000) Influence of forest fragmentation on amphibian diversity in the nature reserve of Ambohitantely, highland Madagascar. *Biological Conservation*, **96**, 31–43.

Wang IJ, Savage WK, Bradley Shaffer H (2009) Landscape genetics and least-cost path analysis reveal unexpected dispersal routes in the California tiger salamander (Ambystoma californiense). *Molecular ecology*, **18**, 1365–74.

Wang Y-H, Yang K-C, Bridgman CL, Lin L-K (2008) Habitat suitability modelling to correlate gene flow with landscape connectivity. *Landscape Ecology*, **23**, 989–1000.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: FST not equal to 1/(4Nm + 1). *Heredity*, **82 (Pt 2)**, 117–25.

Wilson GA, Rannala B (2003) Bayesian Inference of Recent Migration Rates Using Multilocus Genotypes. *Genetics*, **163**, 1177–1191.

Wright S (1943) Isolation by Distance. *Genetics*, **28**, 114–38.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

Zeller K a., McGarigal K, Whiteley AR (2012) Estimating landscape resistance to movement: a review. *Landscape Ecology*, **27**, 777–797.

Glossary of Important Terms

Microsatellite – co-dominant molecular markers of repeating base pairs of DNA useful for assessing overall population gene flow.
Mitochondrial DNA - (mt)DNA is a useful marker for resolving maternal population structure and gene flow because it is a maternally inherited haploid marker (a single chromosome coming only from the mother).
Loci – plural form of the term locus, is the specific location of a gene or DNA sequence or position on a chromosome.
PCR - The polymerase chain reaction (PCR) is a biochemical technology in molecular biology to amplify a single or a few copies of a piece of DNA across several orders of magnitude, generating thousands to millions of copies of a particular DNA sequence.
F_{ST} - A measure of population differentiation due to genetic structure.
G'_{ST} – A standardized analog of F_{ST} to alleviate the problems of high within population heterozgosity.
Resistance Map – A hypothesis of species dispersal based on weighted landscape variables suspected to be important to gene flow.
Optimization – A search for the best fitting model or solution to a problem.
Landscape Genetics - A discipline that analyses the influence of landscape and environmental features on the genetic structure of a population.
Heterozygosity - The state of being heterozygous; having two different alleles of the same gene.
Causal modeling - An approach uses Mantel tests to test between competing hypothesis landscape models (e.g; isolation-by-distance, resistance, isolation-by-barrier).