



A comparison of popular approaches to optimize landscape resistance surfaces

William E. Peterman · Kristopher J. Winiarski · Chloe E. Moore · Carolina da Silva Carvalho · Anthony L. Gilbert · Stephen F. Spear

Received: 5 November 2018 / Accepted: 8 July 2019 / Published online: 16 July 2019
© Springer Nature B.V. 2019

Abstract

Context Landscape resistance surfaces are often used to address questions related to movement, dispersal, or population connectivity. However, modeling landscape resistance is complicated by the selection of the most appropriate analytical approach and the assignment of resistance values to landscape features.

Objectives We compare three common approaches used in landscape genetics to assign resistance values

to landscape features and assess the ability of each approach to correctly identify the data generating resistance surfaces from competing resistance surfaces, as well as the accuracy of each method in recreating the true resistance surface.

Methods Using simulated genetic data and landscape resistance surfaces, three optimization approaches were evaluated: constrained optimization using reciprocal causal modeling (RCM-CO), constrained optimization using linear mixed effects (MLPE-CO) models, and true optimization using ResistanceGA, which combines MLPE models with a genetic algorithm.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10980-019-00870-3>) contains supplementary material, which is available to authorized users.

W. E. Peterman (✉)
School of Environment and Natural Resources, The Ohio State University, 2021 Coffey Road, Columbus, OH 43210, USA
e-mail: Peterman.73@osu.edu

K. J. Winiarski
Department of Environmental Conservation, University of Massachusetts, Amherst, MA, USA

K. J. Winiarski
Northeast Climate Adaptation Science Center, University of Massachusetts, Amherst, MA, USA

C. E. Moore
Department of Biological Sciences, Virginia Polytechnic Institute and State University, 926 West Campus Drive, Blacksburg, VA 24061, USA

C. S. Carvalho
Instituto Tecnológico Vale, Rua Boaventura da Silva 955, Belém, Pará 66055-090, Brazil

A. L. Gilbert
Department of Biological Sciences, Ohio University, 107 Irvine Hall, Athens, OH 45701, USA

S. F. Spear
The Wilds, Cumberland, OH, USA

Results All methods had low type I error (20% or less) when the simulated surface was continuous, but only MLPE-CO and ResistanceGA had low type I error (10% or less) when the simulated surface was categorical. Error was substantially lower with ResistanceGA than MLPE-CO or RCM-CO for multivariate surfaces. Correlation between true and optimized resistance surfaces was generally high with MLPE-CO and ResistanceGA, but low with RCM-CO.

Conclusions MLPE-based approaches (ResistanceGA and MLPE-CO) were superior to RCM-CO, highlighting their value for landscape genetic analyses. The overall performance, objectivity, and accessibility of ResistanceGA underscore its value as a tool for inferring resistance values from genetic data to better understand how landscapes affect dispersal, movement and population connectivity.

Keywords Gene flow · Genetic differentiation · Landscape genetics · Landscape resistance · ResistanceGA · Resistance optimization

Introduction

Landscape resistance surfaces are spatially-explicit raster data layers that assign a resistance value to landscape features that denote the degree to which that variable impedes or facilitates movement (Zeller et al. 2012). They are often used to address a range of questions related to individual movement, dispersal, or connectivity, have been integral to both landscape ecology (Gustafson 2018) and landscape genetics (Spear et al. 2010), and are often the foundation for the planning of ecological corridors and protected areas (Beier et al. 2011). With an appropriately parameterized resistance surface, functional connectivity across the landscape can be assessed. However, assigning specific resistance values to landscape features is not a trivial or straightforward process and remains one of the greatest challenges of developing resistance surfaces (Zeller et al. 2012).

As reviewed by Zeller et al. (2012), numerous approaches exist for assigning resistance values to landscape features. Use of expert opinion is pervasive in determining landscape resistance values, although such approaches are often biased or suboptimal (Clevenger et al. 2002; Seoane et al. 2005; Charney

2012). Approaches to empirically link detection, relocation, or habitat use data (e.g., resource selection functions) to landscape resistance values exist, but these may suffer by conflating a species' habitat use with how a species perceives the landscape during movement. Beier et al. (2008) describe these approaches to determining resistance values as subjective translation, whereby researchers subjectively interpret data or make subjective modeling decisions. Movement data (e.g., GPS telemetry) have also been used to model landscape resistance but only incorporate movement and not successful reproduction. Genetic data are arguably the richest, as they incorporate both movement and dispersal across the landscape and the successful breeding and movement of genes (gene flow). Genetic measures that convey pairwise differentiation between populations (e.g., F_{ST} ; Nei 1973) or similarity between individuals or populations (e.g., proportion of shared alleles, D_{ps} ; Bowcock et al. 1994) are proxies for gene flow, and reflect movement of many individuals over many generations along numerous different pathways. As such, genetic data have tremendous potential to summarize and integrate movement patterns of many individuals through time to understand functional connectivity across the landscape.

From its inception, a central focus of landscape genetics research has been to infer resistance values from genetic data as a measure of actual functional connectivity, so as to test or improve models of potential functional connectivity from an organism perspective (Manel et al. 2003; Storfer et al. 2010). This is achieved by statistically relating pairwise genetic measures to pairwise measures that incorporate landscape resistance, such as effective distance (Ferrerias 2001) or cost distance (Adriaensen et al. 2003). These distance measures combine the effects of geographic distance as well as the effects of the landscape matrix. Referred to as isolation-by-resistance (McRae 2006), this representation of the landscape can be exploited to understand the landscape features affecting movement (Spear et al. 2010). However, there is still a strong reliance on expert opinion when assigning landscape resistance values in the analysis of genetic data (Zeller et al. 2012). Efforts have been made to develop analytical frameworks to optimize the resistance values of landscape features so that pairwise effective distances measured across the landscape maximally correlate with pairwise genetic

measures across the same landscape (e.g., Wang et al. 2009; Shirk et al. 2010; Graves et al. 2013; Peterman et al. 2014; Dudaniec et al. 2016). Despite these efforts, many of the optimization approaches based on genetic data still have limitations. Specifically, they are usually suitable for only specific landscape surface types (i.e., discrete or categorical vs. continuous), search a limited set of potential resistance scenarios (i.e., parameter space), or require extensive a priori assumptions about the relationship of landscape features to landscape resistance. Thus, there is a growing demand to compare the performance as well as the accuracy of distinct analytical methods and optimization approaches when modeling landscape resistance. Because of their use in conservation and policy planning, accuracy of resistance surfaces is paramount; incorrect or inaccurate resistance models may do more harm than good (Rudnick et al. 2012).

Optimization of resistance surfaces typically adheres to the following steps: (1) assign resistance values to each landscape feature (i.e., categorical surface, such as a land cover surface) or apply a transformation (i.e., continuous surface, such as percent canopy cover or elevation); (2) calculate pairwise effective distances (e.g., least cost path distance, cumulative cost distance, resistance distance); (3) fit a statistical model using pairwise genetic data as the response and pairwise effective distances as the predictor; (4) evaluate multiple resistance surfaces using some model selection criteria to determine the best-parametrized resistance surface. Methods that extensively search the available parameter space in step 1 to identify a best parameterization could be considered an approach for optimization. A multitude of data types and approaches have historically been used to assign landscape resistance values, but methods to rigorously and objectively determine resistance values are lacking. Mantel and partial Mantel tests have been the most used analysis frameworks for assessing the relationship between pairwise genetic data and effective distance data within the optimization or selection process (Storfer et al. 2010). However, there has been extensive discussion and research concerning the validity and power of Mantel tests to provide robust inference (Raufaste and Rousset 2001; Guillot and Rousset 2013; Legendre et al. 2015; Zeller et al. 2016; Shirk et al. 2017). Concerns about Mantel-based methods effectively stymied progress and confidence in the ability to assess competing

landscape resistance hypotheses and determine landscape drivers of gene flow. In response, numerous regression-based statistical approaches have been used as alternatives to Mantel tests. Shirk et al. (2017) assessed many of these approaches, and found that linear mixed effects models, specifically the maximum likelihood population effects parameterization (MLPE; Clarke et al. 2002), which controls for the non-independence of pairwise genetic data, performed best among the competing methods.

Even with a robust analytical method, assignment of landscape resistance values remains challenging. To date, most approaches to resistance optimization have relied on a constrained grid optimization approach to search a finite and predetermined parameter space (e.g., Shirk et al. 2010; Dudaniec et al. 2016; Bothwell et al. 2017). In these approaches, the available parameter space is defined and then assessed at discrete intervals to assess a predetermined set of alternative resistance surfaces. Constrained optimization approaches can reasonably sample parameter space, depending upon the number of parameters and ranges of values searched, providing a moderately computationally intensive approach to optimizing landscape resistance surfaces. Less often, exhaustive searches (Wang et al. 2009) or true optimization has been conducted with optimization algorithms that completely or comprehensively search parameter space (Graves et al. 2013; Peterman et al. 2014). With true (unconstrained) optimization, there is no predefined set of resistance surfaces to assess. Rather, the optimization algorithm iteratively generates alternative resistance surfaces until no further improvement can be achieved.

Motivated by the challenge to objectively optimize landscape resistance surfaces, Peterman (2018) developed an optimization framework that combines the MLPE model with a true optimization approach by using a genetic algorithm to optimize a resistance surface based on MLPE model performance criteria (e.g., log likelihood). The use of genetic algorithm here has nothing to do with the genetic data being used in the analysis, but rather refers to the way in which optimization is conducted (see “Methods” for detailed explanation). This approach, implemented in the R package ResistanceGA, is capable of optimizing single or multiple surfaces (continuous or categorical), and requires no a priori assumptions about the relationship between a surface’s original values and

final resistance values. As such, ResistanceGA is unique among all other existing methods for its objectivity, and has become a popular method for addressing important fundamental questions in landscape ecology and landscape genetics, such as the correct thematic resolution of a land cover surface (Khimoun et al. 2017), functional connectivity (Beninde et al. 2016), and the critical scales at which landscape features affect gene flow (Winiarski et al. in revision). Assessment of ResistanceGA's ability to accurately optimize resistance surfaces across a range of conditions are encouraging (Winiarski et al. in revision), but to date no efforts have been made to directly compare the performance of ResistanceGA to other popular landscape resistance modeling methods.

In this study, we compare three methods used to optimize landscape resistance surfaces and conduct model selection: reciprocal causal modeling (RCM; Shirk et al. 2010; Cushman et al. 2013) using a constrained optimization (RCM-CO), maximum-likelihood population-effects mixed modeling using a constrained optimization (MLPE-CO) and ResistanceGA (Peterman) which conducts a true optimization of resistance surfaces. Specifically, we assess the ability of each approach to correctly identify the true resistance surface among alternative resistance surfaces, as well as the ability of each method to optimize and select a resistance surface highly correlated with the true resistance surface.

Methods

Simulation

To assess the ability of each method to (i) correctly identify the true resistance surface among alternative resistance surfaces and (ii) optimize a resistance surface highly correlated with the true resistance surface, we built three scenarios: (1) a categorical land cover surface as the surface transformed into the true resistance surface, (2) a continuous surface as the surface transformed into the true resistance surface, and (3) a combination of a categorical land cover and a continuous surfaces both transformed and then combined to create the true resistance surface.

Simulating landscape resistance surfaces and genetic data

Categorical scenario

Raster surfaces (75×75 pixels) were generated using the 'RandomFields' R package using a random exponential correlation structure with variance and scale parameters equal to 2 (Schlather et al. 2015). The categorical surface consisted of five levels, which were determined by binning raster values into five equal quantiles. Assignment of quantiles to categories was done such that consecutive quantiles were not assigned to consecutive rank-order categories (e.g., quantile 1 = category 3, quantile 2 = category 1, etc.). This process does not necessarily create surfaces with random associations between categories, as might occur in a land cover surface, but serves as an adequate proxy for our simulation. For the categorical surface, true resistance values of each category were determined following the approach of Shirk et al. (2010):

$$R_i = (\text{Rank}_i / V_{\max})^x * R_{\max}.$$

Rank_i is the rank order resistance value (1–5) of each raster cell i , V_{\max} is a constant (5) representing the number of categorical levels in the surface, and R_{\max} is the maximum resistance parameter that was held constant at a value of 55. Variable x represents the shape exponent. For the true resistance surface, $x = 2.5$.

Continuous scenario

For the continuous scenario, true resistance values were determined by transforming the continuous raster using a reverse monomolecular transformation with a shape parameter = 10 and maximum value parameter = 85, as implemented in ResistanceGA (Fig. 1; Peterman 2018).

Multivariate scenario

For the multivariate scenario, the categorical and continuous surfaces were transformed as previously described, then summed, and divided by the minimum value in the surface to rescale resistance values to a minimum resistance of 1 (Fig. 1).

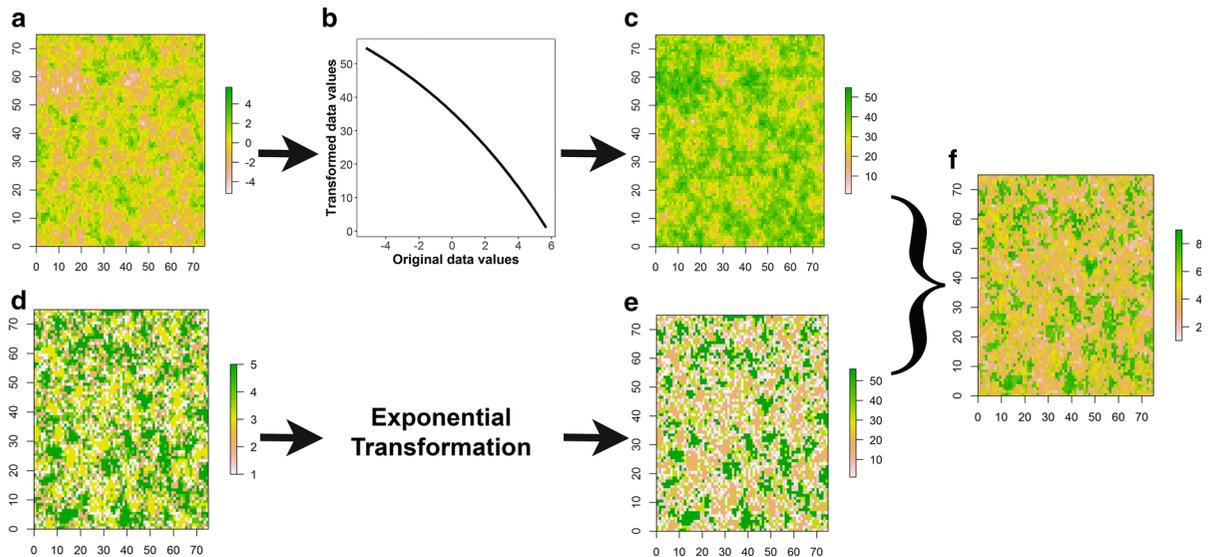


Fig. 1 Visualization of simulated continuous and categorical surfaces (**a**, **d**). These surfaces were simulated using the R package Random Fields using an exponential covariance model with variance and scale parameters equal to 2. The continuous surface is transformed into a landscape resistance surface by applying a reverse monomolecular transformation with a shape parameter of 10 and maximum value parameter of 85 (**b**) to create the true resistance surface for the Continuous scenario (**c**). Applying this transformation (**b**) to the original simulated landscape (**a**) results in a relatively linear conversion of landscape values whereby low values in the original landscape surface have high resistance, and high values in the original landscape have low resistance (**b**). The resistance surface that

results from this transformation has resistance values ranging from 1 to 55 (**c**). The categorical surface (**d**) was transformed by applying an exponential scaling factor of 2.5 (following eq. 1) to create the true resistance surface for the Categorical scenario (**e**), which has resistance values ranging from 1 to 55. The true resistance surface for the Multivariate scenario (**f**) results from summing the Continuous (**c**) and Categorical (**e**) surfaces. The final true resistance surface for the Multivariate scenario is rescaled to have a minimum value of 1 by dividing all resistance values by the minimum of the summed surface. In this example, resistance values of the true Multivariate resistance surface range from 1 to 9

Genetic simulation

To generate genetic data for each scenario, we randomly distributed 100 points (i.e., populations) across the landscape, constraining their locations to be within low resistance areas of the landscape (0.15 percentile or less). We then calculated the effective distance between sample points on the true resistance surface as the random walk commute time between points using the `gdistance` R package (van Etten 2017). This measure is analogous to resistance distance calculated using `CIRCUITSCAPE` (McRae 2006), and represents the effective distance between points, averaged over multiple pathways. Genetic data were then simulated using the population genetic simulation function in `PopGenReport` (Table 1; Adamack and Gruber 2014). Following simulations, we randomly selected 50 of the 100 sample points, and calculated the proportion of shared alleles (D_{ps}) between populations to create pairwise matrix of genetic similarity.

For `ResistanceGA`, it does not matter whether pairwise genetic data is represented as a distance/dissimilarity (e.g., F_{ST}), or a measure of similarity (e.g., D_{ps}).

Simulation structure

Full data simulation code is available in the Online Resource 1. The above steps were repeated ten times for each scenario (categorical, continuous, and multivariate)—a total of 30 genetic data sets, 10 for each of three scenarios. Three raster surfaces were generated at each scenario simulation: one categorical and two continuous. One of the continuous surfaces generated was a random surface, created as described above, but not used in the simulation of genetic data for the continuous or multivariate scenarios. This random surface was included to add an extra level of complexity when evaluating the ability of each optimization approach to correctly identify the true resistance surface.

Table 1 PopGenReport function parameters and settings used to generate genetic data for each of the 30 different simulations assessed

Parameter	Value	Comments
n.pops	100	100 populations simulated, 50 randomly selected for analysis
n.ind	30	Number of individuals per population
sex.ratio	0.5	Males:Females
n.loci	15	Number of loci
n.alleles	15	Number alleles per locus
steps	400	The number of generations simulation was run
n.offspring	2	Number of offspring
mig.rate	0.1	Proportion of population migrating from one population to another
disp.max	0.15	Maximum dispersal is 15% of maximum pairwise resistance distance for each simulation
disp.rate	0.2	The proportion of migrants that will disperse the maximum dispersal distance
mut.rate	0.0005	Mutation rate of microsatellite alleles

The full code for conducting population genetic simulations is available in Online Resource 1

Constrained optimization

Categorical surfaces

Because neither RCM-CO nor MLPE-CO are true optimization approaches, we used a constrained grid search approach that encompassed the true simulation parameters. For both of these approaches, categorical surfaces were optimized by exponentiating the categorical surface by 0.1–1 in increments of 0.1, and 1.5–10 increments of 0.5, using Eq. 1. Selection of the best-supported scenario was completed as described in “[Scenario selection](#)” below. This optimization procedure requires the rank order of resistance values to be specified for categorical surfaces (i.e., a subjective a priori assumption or expert opinion), and the correct rank order was used for analyses due to it being computationally unfeasible to fully explore the complete parameter space for these categorical surfaces (note: ResistanceGA was naïve to the correct rank order of categorical surfaces).

Continuous surfaces

Optimization of continuous surfaces followed a two-step process. First, a coarse grid search was conducted assessing all combinations between transformation, shape, and max value parameters. Four potential transformations (inverse-reverse monomolecular, monomolecular, reverse monomolecular, and inverse

monomolecular), implemented in ResistanceGA (Peterman 2018), were assessed in combination with the shape parameter ranging from 5 to 15 (in increments of 1), and the maximum value parameter ranging from 100 to 500 (in increments of 100). In total, 200 resistance surfaces were assessed in this initial grid search. The best-supported surface was then refined using a second grid search to more precisely define the shape and maximum value parameters. Using the transformation of the best-supported resistance surface from the first grid search, the shape parameter was reassessed at ± 1.25 (in increments of 0.25, minimum value = 0.5) and the maximum resistance value parameter was reassessed at ± 150 (in increments of 50, minimum value = 50). This second grid search resulted in up to 77 scenarios being assessed.

Multivariate surfaces

Multivariate surface (sensu Shirk et al. 2010) optimization with RCM-CO and MLPE-CO was conducted in two ways. Either optimized univariate resistance surfaces were just summed and then divided by the minimum resistance value to create a landscape resistance surface, or univariate resistance surface parameters were adjusted prior to combining surfaces to potentially improve resistance surface fit. To do this, we adjusted the categorical exponent by $\pm 5\%$ and $\pm 15\%$ and adjusted the shape parameter of

optimized continuous surfaces $\pm 10\%$ and $\pm 25\%$. All combinations of these adjusted parameters were assessed when combining resistance surfaces, and the best-supported resistance surface was determined as described below under “Scenario selection”.

Unconstrained optimization

In contrast to the constrained optimization used with RCM-CO and MLPE-CO to optimize a resistance surface, ResistanceGA performs a true, unconstrained optimization by using a genetic algorithm to iteratively adjust resistance surfaces and fit a MLPE model. Genetic algorithms provide a highly flexible and powerful framework for solving optimization problems. In short, genetic algorithms apply the biological principles of natural selection and evolution (Sivanandam and Deepa 2007). A large population of individuals (unique resistance surfaces) is created each generation, each with a set of traits (combination of parameter values) that create unique alternative resistance surfaces. The fitness of each individual (log likelihood of the fitted MLPE model) is then assessed. Genetic algorithms can efficiently solve complex optimization problems through exploration and exploitation (Scrucca 2013). The defined parameter space is explored through random generation of new parameters as a result of random mutation in parameter values as well as exchange through crossover. Exploitation then reduces diversity in the population by selecting the fittest individuals each generation. This approach allows the exploration of the complete parameter space and minimizes the probability of getting “stuck” in a local minima of a complex parameter space. Full details of the optimization process implemented in ResistanceGA can be found in Peterman (2018).

There are two important distinctions to note between how ResistanceGA optimizes resistance surfaces and how we implemented the constrained grid search in the RCM-CO and MLPE-CO. First, when optimizing categorical surfaces ResistanceGA assigns resistance values to each category as opposed to a single exponent parameter. As such, ResistanceGA can identify both the rank order and value of categorical land classes and does not require subjective a priori assumptions about the rank order prior to optimization (different categories can also in theory have equivalent resistance values). Second,

when conducting multivariate optimization, ResistanceGA is simultaneously optimizing all parameters for all surfaces involved, potentially allowing for very different final optimization than would be achieved through the iterative two-step process necessary when using the constrained grid search where surfaces are first optimized in isolation.

Scenario selection

When using RCM-CO, the best-supported resistance surface, was determined using partial Mantel tests as detailed by Cushman et al. (2013) and Bothwell et al. (2017). Briefly, we created a matrix of partial Mantel correlations where N was the number of competing resistance model scenarios. Values in the matrix reflected the partial Mantel correlation between genetic similarity (proportion of shared alleles) and effective distance calculated between sample points on the transformed resistance landscape corresponding to a specified resistance surface model (columns) after controlling for the effect of the alternative resistance models (rows). Relative support for each resistance surface model was calculated as the mean value of each row. The resistance surface model with the largest row mean (i.e., greatest average correlation to genetic similarity after partialling out the effect of other scenarios) was determined to be the most supported resistance surface model. With MLPE-CO and ResistanceGA, the best-supported resistance surface was determined by assessing the log likelihood of the MLPE fit using genetic similarity as the response and resistance distance as the predictor.

Model selection

For each simulated data set, a total of seven surfaces were optimized into resistance surfaces. First, surfaces by themselves (categorical [cat], continuous [cont], continuous_R [contR—random continuous]), then all two- and three-way combinations of these surfaces (cat-cont, cat-contR, cont-contR, cat-cont-contR). Following optimization of potential surfaces into resistance surfaces, a modified bootstrap procedure was used to identify the best-supported optimized landscape resistance surface from the alternative optimized resistance surfaces. This bootstrap procedure was developed to reduce type I error by reducing the probability of selecting a resistance surface that

overfit to the data (Peterman 2018). This procedure subsamples the pairwise effective distance matrices calculated for each optimized resistance surface as well as the genetic similarity matrix (75% of observations selected). The MLPE model is then refit using the subsampled data. Following 1000 bootstrap iterations, the model that was most frequently identified as the top model was determined to be the best-supported resistance surface model. We implemented this bootstrap approach for all of the assessed optimization methods, but only ResistanceGA allowed for an assessment of selection improvement following implementation of the bootstrap procedure. All analyses were conducted using R (R Core Team 2017). Code to conduct RCM-CO and MLPE-CO and model selection are available in Online Resource 2.

Assessment

The performance of each method was evaluated by (i) assessing its ability to correctly identify the true resistance surface among alternative resistance surfaces for each iteration of each scenario using a modified bootstrap analysis and (ii) measuring the correlation between the true resistance surface and the optimized resistance surface and the correlation between the pairwise resistance distances of the true resistance surface and the optimized resistance surface.

Results

Overall, ResistanceGA optimized and correctly identified the data generating scenario in 80% of simulations, resulting in an average correlation between true and optimized resistance surfaces of 0.76 (standard deviation ± 0.21 ; Table 2). Optimization using MLPE-CO and RCM-CO overall had worse performance, correctly identifying the true resistance surface in 74% and 27% of simulations with average resistance surface correlations of 0.57 (± 0.47) and 0.66 (± 0.11), respectively (Table 2). Notably, only ResistanceGA had $\geq 70\%$ success in each of the three scenarios. In looking at specific scenarios, MLPE-CO, RCM-CO and ResistanceGA had low model selection type I error (20% or less) when the simulated surface was continuous, while only MLPE-CO and ResistanceGA had low type I error (10% or less) when the

simulated surface was categorical. Model selection and optimization were generally poorer with multivariate simulations, although type I error was significantly lower with ResistanceGA (30%) than with MLPE-CO (70%) and RCM-CO (100%). Patterns in correlation differed from those patterns found with type I error, with correlation high between the true and selected optimized resistance surface with MLPE-CO (0.98), moderate with ResistanceGA (0.63), and low with RCM-CO (0.0) when the true simulated resistance surface was categorical. Correlation was high with ResistanceGA (0.85), moderate with RCM-CO (0.66), and low with MLPE-CO (0.02) when the true simulated resistance surfaces was continuous. ResistanceGA and MLPE-CO both optimized resistance surfaces that were highly correlated (≥ 0.83) with the true resistance surface in the multivariate simulation scenario, while RCM-CO failed to correctly optimize a multivariate surface.

A bootstrap procedure was implemented prior to final model selection in all scenarios and optimization approaches. Our work flow only permitted a comparison of model selection with and without the bootstrap procedure for ResistanceGA. Model selection success using ResistanceGA for categorical scenarios did not change, selection success increased 10% for continuous scenarios, and selection success increased 40% for multivariate scenarios with the use of the bootstrap procedure. We found that adjustment of univariate optimized values prior to combining surfaces to make a multivariate resistance surface made no difference for RCM (0% success), and improved success from 10 to 30% for MLPE. In cases where the data generating scenario was correctly identified through model selection, the correlation between pairwise effective resistance values calculated on the true and optimized resistance surfaces was typically high (≥ 0.87) for both MLPE-CO and ResistanceGA.

Discussion

Progress toward developing more robust solutions to a foundational problem in the field of landscape genetics, understanding how landscape features affect gene flow, has been limited. In this study, we found that the MLPE based approaches, ResistanceGA and MLPE-CO, performed best in selection of true resistance surface models from a candidate set of alternatives.

Table 2 Summary of the performance of each method: reciprocal causal modeling using a constrained optimization approach (RCM-CO), maximum likelihood population effects

mixed effects model using a constrained optimization approach (MLPE-CO) and ResistanceGA that combines MLPE and a genetic optimization algorithm for unconstrained optimization

Method	Scenario	Correctly identified	Raster correlation: correct	Pairwise resistance correlation: correct	Raster correlation: all surfaces	Pairwise resistance correlation: all surfaces
RCM-CO	Categorical ^a	0/10	NA	NA	0.01 ± 0.01	0.11 ± 0.11
MLPE-CO	Categorical ^a	10/10	0.98 ± 0.02	0.93 ± 0.11	0.89 ± 0.30	0.93 ± 0.11
ResistanceGA	Categorical	9/10 ^b	0.63 ± 0.28	0.87 ± 0.11	0.67 ± 0.29	0.88 ± 0.10
RCM-CO	Continuous	8/10	0.66 ± 0.11	0.11 ± 0.08	0.54 ± 0.28	0.10 ± 0.08
MLPE-CO	Continuous	9/10	0.02 ± 0.01	0.96 ± 0.01	0.02 ± 0.01	0.96 ± 0.01
ResistanceGA	Continuous	8/10 ^b	0.85 ± 0.09	0.97 ± 0.01	0.76 ± 0.28	0.98 ± 0.02
RCM-CO _{add} ^c	Multivariate	0/10	NA	NA	0.18 ± 0.07	0.06 ± 0.05
RCM-CO _{adj} ^c	Multivariate	0/10	NA	NA	0.18 ± 0.07	0.06 ± 0.05
MLPE-CO _{add} ^c	Multivariate	1/10	0.92 ± NA	0.95 ± NA	0.86 ± 0.19	0.94 ± 0.04
MLPE-CO _{adj} ^c	Multivariate	3/10	0.93 ± 0.01	0.89 ± 0.05	0.73 ± 0.30	0.92 ± 0.04
ResistanceGA	Multivariate	7/10 ^b	0.83 ± 0.11	0.95 ± 0.02	0.84 ± 0.10	0.94 ± 0.03

Ability to correctly identify the true resistance surface among alternative resistance surfaces (Type I error rate). Mean and standard deviation in the correlation between the optimized raster surfaces and the true resistance surface, as well as the correlation between pairwise resistance distances calculated on optimized and true resistance surfaces are shown. Correlation statistics were calculated between correctly identified resistance surface (correlation: correct) and the best-supported resistance surface (correlation: all surfaces) separately. NAs in the table indicate that the mean or standard deviation could not be calculated because the method either failed to correctly identify the true resistance surface or only did so once. Scenario indicates the type of surface used to develop the true resistance surface in the genetic simulation. Ten replicates simulations were conducted for each scenario, with correctly identified indicating the number of times the true resistance surface for the scenario was identified as top resistance surface following optimization and a modified bootstrap analysis

^aThe rank-order of resistance values was provided, and the maximum resistance value was fixed for constrained optimization of categorical surfaces

^bFrequency that the true surface was identified by AICc model ranking alone (no bootstrap): Categorical = 9/10; Continuous = 7/10; Multivariate = 4/10

^cTwo approaches were used to create multivariate surfaces for RCM and MLPE. “_add” is the simple addition of univariate optimized surfaces while “_adj” adjusted optimized univariate parameters prior to addition. See text for full details

Additionally, both methods were able to comparably optimize resistance surfaces to create a resistance surface that was highly correlated with the true resistance surface. ResistanceGA stood apart from MLPE-CO in the most challenging, and arguably the most ecologically relevant, multivariate scenarios where ResistanceGA had the lowest type I error rates. Additionally, ResistanceGA and its use of a genetic algorithm for a true optimization has two major advantages over a constrained optimization approach such as MLPE-CO or RCM-CO: accessibility for conducting analyses and objectivity in assigning resistance values.

RCM-CO and MLPE-CO are able to be implemented by the user in software such as R (R Core Team 2017), but requires more manual coding, and the incorporation of even a simple constrained grid search, as in this study, can result in hundreds of scenarios being evaluated. Unfortunately, this level of coding and data analysis management may preclude many from using or adapting existing methods. One of the most challenging aspects of our study was the development of a semi-automated workflow to consistently implement the RCM-CO and MLPE-CO approaches to make an equitable comparison to ResistanceGA (code available in Online Resource

2). ResistanceGA, an existing R package, has been designed to provide a user friendly and comprehensive analysis framework. Further, ResistanceGA requires no subjective a priori assumptions about how landscape features relate to resistance. A priori assumptions most frequently manifest in the form of expert opinion, the shortcomings of which are well documented (Charney 2012; Zeller et al. 2012). Moreover, optimization relying on a priori hypotheses (such as rank order of categorical land classes) may overlook the true landscape resistance process (e.g., Khimoun et al. 2017). For example, categorical surfaces are commonly used to assess patterns of resistance across the landscape, but comprehensive assessment of different rank order resistance of categories quickly becomes computationally unfeasible to assess under a constrained optimization approach.

ResistanceGA fills a prominent void in the landscape genetics toolbox, but it is not without limitations. First, there is potential that resistance surfaces optimized with ResistanceGA may be overfit to the data used in the analysis, rather than providing a generalization of how landscape features affect gene flow. This likely becomes less concerning as sample size and distribution increase (Oyler-McCance et al. 2013), but is a real possibility with small data sets. In an effort to minimize overfitting, a modified bootstrap procedure has been implemented in ResistanceGA to assess the frequency that an optimized resistance surface is the best supported when the MLPE model is refit with a subset of the sample locations. A second limitation is that there is currently no way to assess confidence or uncertainty in the optimized parameter values, as advocated by Dudaniec et al. (2016). A third limitation is the size (number of pixels) of raster surfaces that can feasibly be optimized. During optimization, ResistanceGA may create and calculate effective distance on more than > 2500 possible scenarios. The computation time required to transform resistance surfaces increases with surface size, and the time to calculate pairwise effective distances is a function of both sample and surface size (Peterman 2018). Parallel processing can reduce optimization time, and faster algorithms for calculating pairwise effective distances have tremendous potential to make optimization of larger landscapes more tractable (Anantharaman et al. 2019). If landscape resistance surfaces contain a large number of pixels, optimization with ResistanceGA may not be tractable.

However, our simulations suggest that MLPE-CO may be a computationally intensive, yet feasible, approach, especially if categorical surfaces are not included. A fourth limitation of ResistanceGA is that this approach does not currently account for spatial autocorrelation in MLPE models, the presence of which may increase type I rate. Jaffé et al. (2019) introduced a promising extension of the MLPE model to address this issue by modifying the MLPE model to directly incorporate correlation between pairwise measurements due to comparison of both individuals and spatial locations. Finally, it is important to recognize that no method can resolve issues of poor study design, including the number and spatial distribution of sampled individuals or populations (Oyler-McCance et al. 2013).

Regardless of study design or constraints, we emphasize that care must be taken when interpreting the specific drivers of gene flow and the accuracy of optimized resistance surfaces. We have demonstrated that the true resistance surface can frequently be identified among alternative resistance surfaces (a big step forward), but this does not necessarily mean that the optimized resistance surface is always accurate, such as with our categorical scenarios optimized with ResistanceGA. Here, the lower correlation between the resistance values of the categorical optimized surfaces and the true data generating surface (0.63 ± 0.28) is likely because the correct rank order of categories was identified in only 3 of 9 simulations (although the lowest resistance category was always correctly assigned a resistance of 1). It is important to remember that although MLPE-CO performed better in our study, categorical surfaces were optimized with the correct rank order, and the maximum resistance parameter was held constant. As such, we created the most favorable setting possible for constrained optimization procedures to succeed. Depending on the real-world categorical surface being evaluated, the rank order of relative resistance values may not always be obvious, adding further complexity to the optimization process.

Over the last 15 years, landscape genetics, and increasingly landscape genomics, have been pivotal in advancing our understanding of how landscape features affect gene flow. Such knowledge is vital for effective management of populations in space and time, but critical to many landscape genetic studies and subsequent management decisions is the assignment of resistance values to landscape surfaces.

Despite their importance, accessible, rigorous, and objective approaches to landscape resistance parameterization have been lacking. As demonstrated through our simulations, the R package ResistanceGA fills this gap. We acknowledge that our simulation study was limited in scope relative to what previous simulation studies have assessed (e.g., sample size, sample location, thematic resolution, landscape configuration; Cushman and Landguth 2010; Landguth et al. 2012; Oyler-McCance et al. 2013), but our primary focus was on providing an equitable assessment of methods readily used in current studies. Per its initial intent, ResistanceGA was the easiest and most accessible approach to implement when optimizing resistance surfaces, and its ability to frequently select the true model and accurately optimize a resistance surface that is highly correlated with the true resistance surface suggests that it should become integral to the landscape genetics toolbox to objectively optimize and select resistance surfaces.

Acknowledgements This work resulted from a Distributed Graduate Course in Landscape Genetics. Computational resources for conducting analyses were provided by the Ohio Supercomputer Center (1987) at The Ohio State University. The manuscript was greatly improved by the comments of two reviewers and the coordinating editor.

References

- Adamack AT, Gruber B (2014) PopGenReport: simplifying basic population genetic analyses in R. *Methods Ecol Evol* 5(4):384–387
- Adriaensen F, Chardon JP, De Blust G, Swinnen E, Villalba S, Gulinck H, Matthysen E (2003) The application of ‘least-cost’ modelling as a functional landscape model. *Landscape Urban Plan* 64(4):233–247
- Anantharaman R, Hall K, Shah V, Edelman A (2019) Circuitscape in Julia: High performance connectivity modelling to support conservation decisions. <http://arxiv.org/abs/arXiv:1906.03542>
- Beier P, Majka DR, Spencer WD (2008) Forks in the road: choices in procedures for designing wildland linkages. *Conserv Biol* 22(4):836–851
- Beier P, Spencer W, Baldwin RF, McRae BH (2011) Toward best practices for developing regional connectivity maps. *Conserv Biol* 25(5):879–892
- Beninde J, Feldmeier S, Werner M, Peroverde D, Schulte U, Hochkirch A, Veith M (2016) Cityscape genetics: structural versus functional connectivity of an urban lizard population. *Mol Ecol* 25(20):4984–5000
- Bothwell HM, Cushman SA, Woolbright SA, Hersch-Green EI, Evans LM, Whitham TG, Allan GJ (2017) Conserving threatened riparian ecosystems in the American West: precipitation gradients and river networks drive genetic connectivity and diversity in a foundation riparian tree (*Populus angustifolia*). *Mol Ecol* 26(19):5114–5132
- Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368(6470):455–457
- Charney ND (2012) Evaluating expert opinion and spatial scale in an amphibian model. *Ecol Model* 242:37–45
- Clarke RT, Rothery P, Raybould AF (2002) Confidence limits for regression relationships between distance matrices: estimating gene flow with distance. *JABES* 7(3):361–372
- Clevenger AP, Wierzchowski J, Chruszcz B, Gunson K (2002) GIS-generated, expert-based models for identifying wildlife habitat linkages and planning mitigation passages. *Conserv Biol* 16(2):503–514
- Cushman SA, Landguth EL (2010) Scale dependent inference in landscape genetics. *Landscape Ecol* 25:967–979
- Cushman S, Wasserman T, Landguth E, Shirk A (2013) Re-evaluating causal modeling with Mantel tests in landscape genetics. *Diversity* 5(1):51
- Dudaniec RY, Worthington Wilmer J, Hanson JO, Warren M, Bell S, Rhodes JR (2016) Dealing with uncertainty in landscape genetic resistance models: a case of three co-occurring marsupials. *Mol Ecol* 25(2):470–486
- Ferreras P (2001) Landscape structure and asymmetrical inter-patch connectivity in a metapopulation of the endangered Iberian lynx. *Biol Conserv* 100(1):125–136
- Graves TA, Beier P, Royle JA (2013) Current approaches using genetic distances produce poor estimates of landscape resistance to interindividual dispersal. *Mol Ecol* 22(15):3888–3903
- Guillot G, Rousset F (2013) Dismantling the Mantel tests. *Methods Ecol Evol* 4(4):336–344
- Gustafson EJ (2018) How has the state-of-the-art for quantification of landscape pattern advanced in the twenty-first century? *Landscape Ecol*. <https://doi.org/10.1007/s10980-018-0709-x>
- Jaffé R, Veiga JC, Pope NS, Lanes ÉCM, Carvalho CS, Alves R, Andrade SCS, Arias MC, Bonatti V, Carvalho AT, de Castro MS, Contrera FAL, Francoy TM, Freitas BM, Giannini TC, Hrcir M, Martins CF, Oliveira G, Saraiva AM, Souza BA, Imperatriz-Fonseca VL (2019) Landscape genomics to the rescue of a tropical bee threatened by habitat loss and climate change. *Evol Appl*. <https://doi.org/10.1111/eva.12794>
- Khimoun A, Peterman W, Eraud C, Faivre B, Navarro N, Garnier S (2017) Landscape genetic analyses reveal fine-scale effects of forest fragmentation in an insular tropical bird. *Mol Ecol* 26(19):4906–4919
- Landguth EL, Fedy BC, Oyler-McCance SJ, Garey AL, Emel SL, Mumma M, Wagner HH, Fortin M-J, Cushman SA (2012) Effects of sample size, number of markers, and allelic richness on the detection of spatial genetic pattern. *Mol Ecol Resour* 12(2):276–284
- Legendre P, Fortin M-J, Borcard D (2015) Should the Mantel test be used in spatial analysis? *Methods Ecol Evol* 6(11):1239–1247
- Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol Evol* 18(4):189–197

- McRae BH (2006) Isolation by resistance. *Evolution* 60(8):1551–1561
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* 70(12):3321–3323
- Ohio Supercomputer Center (1987) Ohio Supercomputer Center. Columbus, OH: Ohio Supercomputer Center. <http://osc.edu/ark:/19495/f5s1ph73>
- Oyler-McCance SJ, Fedy BC, Landguth EL (2013) Sample design effects in landscape genetics. *Conserv Genet* 14(2):275–285
- Peterman WE (2018) ResistanceGA: an R package for the optimization of resistance surfaces using genetic algorithms. *Methods Ecol Evol* 9(6):1638–1647
- Peterman WE, Connette GM, Semlitsch RD, Eggert LS (2014) Ecological resistance surfaces predict fine-scale genetic differentiation in a terrestrial woodland salamander. *Mol Ecol* 23(10):2402–2413
- R Core Team (2017) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>. Accessed 16 Dec 2017
- Raufaste N, Rousset F (2001) Are partial mantel tests adequate? *Evolution* 55(8):1703–1705
- Rudnick D, Ryan SJ, Beier P, Cushman SA, Dieffenbach F, Epps C, Gerber LR, Hartter JN, Jenness JS, Kintsch J (2012) The role of landscape connectivity in planning and implementing conservatiand restoration priorities. *Issues in Ecology*. Report No. 16. Ecological Society of America, Washington, DC
- Schlather M, Malinowski A, Menck J, Oesting M, Strokorb K (2015) Analysis, simulation and prediction of multivariate random fields with package randomfields. *J Stat Softw* 63(8):1–25
- Scrucca L (2013) GA: a package for genetic algorithms in R. *J Stat Softw* 53(4):1–37
- Seoane J, Bustamante J, Díaz-Delgado R (2005) Effect of expert opinion on the predictive ability of environmental models of bird distribution. *Conserv Biol* 19(2):512–522
- Shirk AJ, Landguth EL, Cushman SA (2017) A comparison of regression methods for model selection in individual-based landscape genetic analysis. *Mol Ecol Resour* 18(1):55–67
- Shirk AJ, Wallin DO, Cushman SA, Rice CG, Warheit KI (2010) Inferring landscape effects on gene flow: a new model selection framework. *Mol Ecol* 19(17):3603–3619
- Sivanandam S, Deepa S (2007) Introduction to genetic algorithms. Springer, Berlin
- Spear SF, Balkenhol N, Fortin MJ, McRae BH, Scribner K (2010) Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Mol Ecol* 19(17):3576–3591
- Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? *Mol Ecol* 19(17):3496–3514
- van Etten J (2017) R package gdistance: distances and routes on geographical grids. *J Stat Softw* 76(13):1–21
- Wang JJ, Savage WK, Shaffer HB (2009) Landscape genetics and least-cost path analysis reveal unexpected dispersal routes in the California tiger salamander (*Ambystoma californiense*). *Mol Ecol* 18:1365–1374
- Winiarski KJ, Whiteley AR, Peterman WE, McGarigal K (Forthcoming) Multi-scale resistant kernel surfaces derived from inferred gene flow: An application with vernal breeding amphibians. *Mol Ecol Resour*
- Zeller KA, Creech TG, Millette KL, Crowhurst RS, Long RA, Wagner HH, Balkenhol N, Landguth EL (2016) Using simulations to evaluate Mantel-based methods for assessing landscape resistance to gene flow. *Ecol Evol* 6(12):4115–4128
- Zeller K, McGarigal K, Whiteley A (2012) Estimating landscape resistance to movement: a review. *Landscape Ecol* 27(6):777–797

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.