EDITOR'S CHOICE

BIODIVERSITY RESEARCH

WILEY **Diversity** and **Distributions**

Are all data types and connectivity models created equal? Validating common connectivity approaches with dispersal data

Katherine A. Zeller[1] | Megan K. Jennings[2] | T. Winston Vickers[3] |
Holly B. Ernest[4] | Samuel A. Cushman[5] | Walter M. Boyce[3]

[1]Department of Environmental
Conservation, University of Massachusetts,
Amherst, MA, USA

[2]Biology Department, San Diego State
University, San Diego, CA, USA

[3]Karen C. Drayer Wildlife Health
Center, School of Veterinary
Medicine, University of California, Davis,
CA, USA

[4]Department of Veterinary
Sciences, University of Wyoming, Laramie,
WY, USA

[5]U.S. Forest Service Rocky Mountain
Research Station, Flagstaff, AZ, USA

Correspondence
Katherine A. Zeller, Department of
Environmental Conservation, University of
Massachusetts Amherst, Amherst, MA, USA.
Email: kathyzeller@gmail.com

Editor: Janine Bolliger

Abstract

Aim: There is enormous interest in applying connectivity modelling to resistance surfaces for identifying corridors for conservation action. However, the multiple analytical approaches used to estimate resistance surfaces and predict connectivity across resistance surfaces have not been rigorously compared, and it is unclear what methods provide the best inferences about population connectivity. Using a large empirical data set on puma (*Puma concolor*), we are the first to compare several of the most common approaches for estimating resistance and modelling connectivity and validate them with dispersal data.

Location: Southern California, USA.

Methods: We estimate resistance using presence-only data, GPS telemetry data from puma home ranges and genetic data using a variety of analytical methods. We model connectivity with cost distance and circuit theory algorithms. We then measure the ability of each data type and connectivity algorithm to capture GPS telemetry points of dispersing pumas.

Results: We found that resource selection functions based on GPS telemetry points and paths outperformed species distribution models when applied using cost distance connectivity algorithms. Point and path selection functions were not statistically different in their performance, but point selection functions were more sensitive to the transformation used to convert relative probability of use to resistance. Point and path selection functions and landscape genetics outperformed other methods when applied with cost distance; no methods outperformed one another with circuit theory.

Main conclusions: We conclude that path or point selection functions, or landscape genetic models, should be used to estimate landscape resistance for wildlife. In cases where resource limitations prohibit the collection of GPS collar or genetic data, our results suggest that species distribution models, while weaker, may still be sufficient for resistance estimation. We recommend the use of cost distance-based approaches, such as least-cost corridors and resistant kernels, for estimating connectivity and identifying functional corridors for terrestrial wildlife.

1 | INTRODUCTION

Human populations and associated development continue to expand and fragment natural areas, increasing the isolation of wildlife populations (Haddad et al., 2015). Identifying and conserving wildlife corridors between populations have been shown to increase the movement of individuals and counteract population isolation (Gilbert-Norton, Wilson, Stevens, & Beard, 2010). Movement is critical for fulfilling several biological processes such as mating, dispersal, migration and gene flow (Hilty, Lidicker, & Merenlender, 2012) and allows individuals and populations to spatially redistribute in response to disturbances and climate change (Chen, Hill, Ohlemuller, Roy, & Thomas, 2011). The flow of individuals among populations also aids in maintaining metapopulation viability and alleviates threats faced by small, isolated populations (Haddad et al., 2015; Nicholson et al., 2006).

Modelling wildlife corridors requires two main components, (1) a resistance to movement surface (also known as a cost surface) and (2) a connectivity algorithm. Resistance can be estimated empirically using a wide variety of data types including opportunistic presence-only data, occurrence data, relocation data, GPS collar data and genetic data (Zeller, McGarigal, & Whiteley, 2012). Typically, habitat suitability is estimated from one of these data types. For methods based on habitat selection (e.g., all except genetic), resistance is then represented as an inverse function of habitat suitability so that areas of high habitat suitability confer a low resistance and vice versa. Estimating resistance directly from dispersal data is considered the most reliable approach as dispersal is the main process by which organisms move between populations (e.g., Elliot, Cushman, Macdonald, & Loveridge, 2014). However, dispersal data are extremely difficult to obtain and when it is available, sample sizes are typically low (Fagan & Calabrese, 2006). Therefore, GPS collar data derived from home-range movement in the form of steps, paths or movement-only points, or movement inferred from genetic data, are considered the closest proxies for estimating dispersal movements and may be more appropriate for estimating resistance (Cushman & Lewis, 2010; Zeller et al., 2012).

Once a resistance surface is developed, there are a wide variety of algorithms to estimate connectivity, each with different advantages and limitations. Two of the most commonly used connectivity algorithms are CircuitScape (McRae, 2006) and cost distance-based approaches, such as the least-cost corridor (Adriaensen et al., 2003). CircuitScape (CS), based on electrical circuit theory, runs current across resistance surfaces where each pixel is a node and the resistance values between pixels are the resistors. The CS algorithm results in an estimated current density for each pixel on the landscape, with higher values of current representing higher probabilities that a random walker will pass through a pixel. Cost distance algorithms

(CD) calculate the cumulative cost from source points across a resistance surface. Cost distance surfaces from two source points can be paired, and troughs can be identified where the cumulative cost and distance between the two points are minimized. Most often, a single, least-cost corridor is derived from these surfaces, although the surfaces can also be bounded at some upper connectivity value to derive multiple paths. The advantage of CS is that it does not assume an individual has complete knowledge of the landscape through which it is travelling. CD approaches are often criticized because they do make this assumption, which can be problematic as dispersing individuals almost never have experience of the landscape outside of their natal territories (Theobald, 2006). However, an assumption of CS methods is that each pixel, or grid cell, is a population, which may not be a realistic assumption when modelling connectivity at smaller spatial scales, between discrete populations.

Gaining a better understanding of how each data type and connectivity algorithm performs is crucial for future conservation research (Abrahms et al., 2017; Elliot et al., 2014). For example, if models from opportunistic presence-only data capture the dispersal process as well as models from GPS collar data or extensive genetic sampling, then time-consuming and costly data collection efforts may not be necessary. However, if presence-only data perform poorly relative to more robust data types, then the effort for collaring or genetic studies is better justified (e.g., Mateo-Sánchez et al., 2015a,b). To date, there have been few studies assessing the performance of resistance surfaces and connectivity models with independent dispersal data—likely owing to the difficulty in obtaining data on a sufficient number of dispersers. Fattebert, Robinson, Balme, Slotow, and Hunter (2015), Keeley, Beier, and Gagnon (2016), Keeley, Beier, Keeley, and Fagan (2017), and Newby (2011) compared the ability of resource selection functions to predict dispersal habitat and found that predictive ability was high. Elliot et al. (2014), Gastón et al. (2016), and Jackson, Marnewick, Lindsey, Roskaft, and Robertson (2016) found substantial differences among habitat use during resource use and dispersal and concluded dispersers tolerated a wider array of landscape features than residents, and therefore, analysis of occurrence data or non-dispersal movement data would provide biased predictions. For connectivity algorithms, when CS and CD algorithms have been compared, CS tended to outperform CD in capturing dispersal movements (Jackson et al., 2016; McClure, Hansen, & Inman, 2016). None of these studies assessed different data types for estimating resistance as well as different connectivity algorithms in a single analysis. Additionally, none of these studies explored resistance surfaces derived from opportunistic presence-only data or genetic data.

Here, we use empirical dispersal data to assess the performance of several data types and methods for estimating resistance and modelling corridors. Using data from pumas (*Puma concolor*)
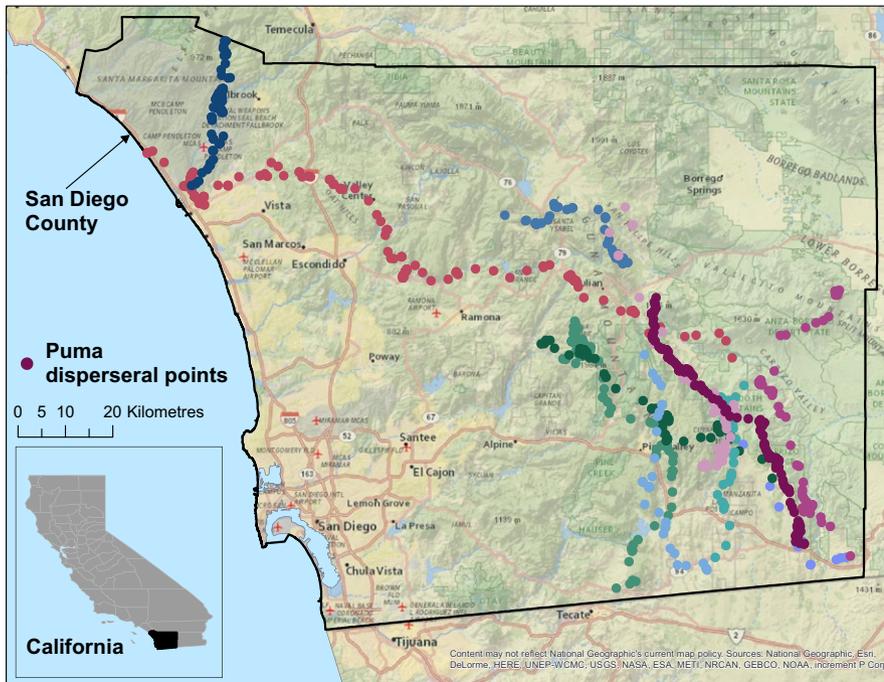
FIGURE 1 San Diego County, California, USA, study area and points from eleven puma dispersal paths from 2005 to 2012. The dispersal points were used for validating the suite of connectivity surfaces. [Colour figure can be viewed at wileyonlinelibrary.com]

in San Diego County, California, USA, we estimated resistance using opportunistic presence-only data analysed with seven different species distribution models (SDMs), GPS collar data analysed with point and path selection functions (PSF; PathSF), and genetic data analysed in a landscape genetic framework. We modelled connectivity between the start and end locations of puma dispersal paths using CS and CD algorithms. We then used independent dispersal data from pumas in the study area to assess the performance of each resistance surface/connectivity algorithm combination. Because previous studies have shown that different data types produce resistance surfaces with similar variables and relationships to resistance (Cushman, Lewis, & Landguth, 2014), we predicted that all our data types would produce qualitatively similar resistance surfaces. We also expected, based on previous research comparing point and movement-based approaches (Abrahms et al., 2017; Zeller et al., 2014), that resistance surfaces derived from movement and genetic models would outperform those based on presence-only data. Lastly, we predicted that the CD connectivity algorithm would outperform the CS algorithm because the assumptions of CS are not met when the scale of analysis is individual movement across the landscape.

2 | METHODS

2.1 | Study area and environmental variables

Our study area, measuring 11,722 km^2, was San Diego County, California, USA (Figure 1). The region experiences a Mediterranean climate with mild winters and hot dry summers. The western part of the study area is comprised of coastal and foothill habitats while the eastern part of the study area is comprised of mountain and desert habitats. Elevations range from sea level to 1,992 m.

We selected human development, topographic and vegetation variables known to influence puma habitat use and movement (Table 1). Our previous work has shown resistance surfaces and corridors for pumas are sensitive to the choice of geospatial layers and their thematic resolutions (Zeller, McGarigal, et al., 2017). Therefore, we represented many of these variables with different thematic resolutions and allowed the models to drive the selection of the thematic representation for a variable. We used 30 m as the spatial grain size for the variables.

To determine the most appropriate scale of selection for each variable in each model, we explored smoothing kernels of various sizes (McGarigal, Wan, Zeller, Timm, & Cushman, 2016). Using the *smoothie* package (Gilleland, 2013) in the R software environment (R Core Team 2016), we applied a Pareto kernel to each pixel in the landscape representing the following scales: 241 m, 681 m, 1,123 m, 1,850 m, 2,312 m, 3,044 m, 3,819 m, 4,461 m. Scales were based on empirically estimating the Pareto curve from the distribution of puma movement distances across eight different time periods from 15 min to 6 hr (details in Zeller et al., 2014).

Univariate models (described in detail in Appendix S1) were run, and AUC values or AICc values were used to determine the most appropriate scale for each variable for the SDMs and resource selection functions, respectively. Pairwise correlations between scaled variables were assessed, and the higher performing variable between pairs with a correlation coefficient |r| > .60 was retained for the multivariate models.

2.2 | Habitat suitability and movement models

We explored four data types in our analysis: (1) opportunistic presence-only data, (2) point data from GPS telemetry collars, (3) path data from GPS telemetry collars and (4) genetic data. The

TABLE 1 Environmental variables used in the puma models. Source or derivation, data year and citation are provided

	Variable	Source/Derivation	Year	Citation
Roads and development	All Roads[a]	Open Street Map	2014	Open Street Map 2014
	Primary roads[a]	Open Street Map; Motorways	2014	Open Street Map 2014
	Secondary roads[a]	Open Street Map; primary road, secondary road and trunk road	2014	Open Street Map 2014
	Tertiary roads[a]	Open Street Map; living street, residential, rest area, road, service, tertiary and unclassified	2014	Open Street Map 2014
	Unpaved roads/trails[a]	Open Street Map; bridleway, cycleway, footway, path and track	2014	Open Street Map 2014
	Percentage Imperviousness[b]	Derived from a hybrid of the National Land Cover Database percentage impervious surface and updated data from the San Diego Association of Governments land use surface	2011/2012	NLCD 2011 (Jin et al. 2013), SANDAG 2012
Topography	Elevation[b]	National Elevation Dataset	2009	USGS 2009
	Percentage Slope[b]	Derived from National Elevation Dataset	–	–
	Terrain Ruggedness[b]	Total curvature derived from National Elevation Dataset with DEM Surface Tools (Jenness 2013)	–	–
	Topographic Position Index[b]	Derived from National Elevation Dataset	–	–
	Ridges[a]	Derived from Topographic Position Index values ≥ 8	–	–
	Canyons[a]	Derived from Topographic Position Index values ≤−8	–	–
	Steep Slope[a]	Derived from Topographic Position Index values −8 − 8, slope ≥6°	–	–
	Gentle Slope[a]	Derived from Topographic Position Index values −8 −8, slope ≤6°	–	–
Water	Streams[a]	National Hydrography Dataset streams layer	2011	USGS 2011
	Distance to Water[b]	Derived from National Hydrography Dataset calculated as Euclidean distance to blue line streams	–	–
Vegetation type	Agriculture[a]	Vegetation Data of San Diego County	2014	SANDAG 2014
	Chaparral[a]	Vegetation Data of San Diego County	2014	SANDAG 2014
	Coastal Scrub[a]	Vegetation Data of San Diego County	2014	SANDAG 2014
	Coniferous Forest[a]	Vegetation Data of San Diego County	2014	SANDAG 2014
	Desert Scrub[a]	Vegetation Data of San Diego County	2014	SANDAG 2014
	Hardwood Forest[a]	Vegetation Data of San Diego County	2014	SANDAG 2014
	Herbaceous Grassland[a]	Vegetation Data of San Diego County	2014	SANDAG 2014
	Riparian[a]	Vegetation Data of San Diego County	2014	SANDAG 2014
	Urban[a]	Vegetation Data of San Diego County	2014	SANDAG 2014
	Water and Wetlands[a]	Vegetation Data of San Diego County	2014	SANDAG 2014

[a]Categorical variables.
[b]Continuous variables.

GPS telemetry point and path data were from individuals operating within established home ranges. To estimate resistance and model connectivity from the first data type, we developed seven multi-scale species distribution models (SDMs) based on six different algorithms and an ensemble SDM. From the second data type, we developed two multi-scale point selection functions—one with
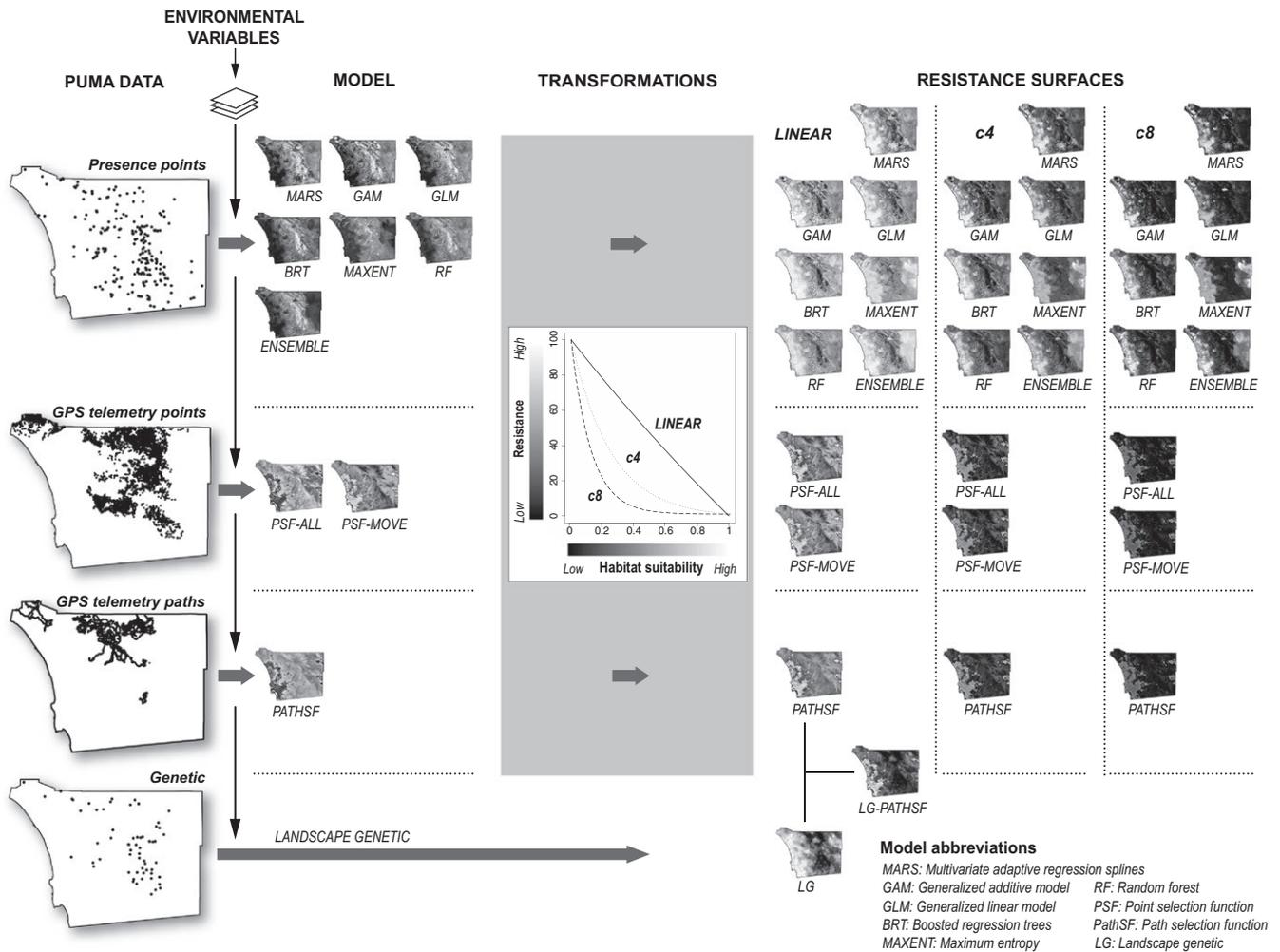
FIGURE 2 Conceptual diagram of the methods used to estimate resistance from each data type. A data type was combined with environmental variables to create a habitat model. The habitat model was then predicted across the study area. Three transformations were applied to the habitat surfaces to convert habitat suitability (or probability of use or movement) to resistance. These transformations, negative linear, c4 and c8, were based on the following function, provided by Keeley et al. (2016): $R = 100 - 99 * ((1-\exp(-c*h))/(1-\exp(-c)))$ where R is resistance, h is habitat suitability and c is either 4 or 8. The landscape genetic approach estimated resistance directly; therefore, no transformation was used with the genetic data. Each resistance surface was then fed into two connectivity models, CircuitScape and cost distance. Details on puma data and analytical approaches are provided in Appendix S1

all puma GPS data points (PSF-All) and one with GPS points from when a puma was in a movement state (PSF-Move; Zeller et al., 2014). From the third data type, we developed a multi-scale path selection function (Zeller et al., 2016; Figure 2). We then predicted habitat suitability across the study area from the SDMs, relative probability of resource use from the PSFs and relative probability of movement from the PathSFs. The genetic data were collected from captured or deceased pumas across the study area and were used in a multi-scale landscape genetic analysis, which estimated resistance directly. Information detailing the puma data and multi-scale analytical approaches is provided in Appendix S1.

2.3 | Resistance surfaces

Beier, Majka, and Spencer (2008) challenged the assumption that resistance is the linear negative inverse of habitat suitability.

Furthermore, Mateo-Sánchez et al. (2015a,b) and Keeley et al. (2016) proposed that dispersing individuals may be more tolerant of subpar habitat types than individuals in home ranges or established territories. They found more support for relaxing the relationship between habitat suitability and resistance so that only lower values of habitat suitability would have a high resistance (Figure 2). Therefore, we explored three different transformations to convert habitat suitability, or predicted probability of use/movement, to resistance: a negative linear transformation and two negative exponential transformations (c4 and c8 following Keeley et al., 2016; Figure 2). The negative exponential transformations assign high resistance values to only the lowest habitat suitability values with the c8 being the most extreme of the two transformations (Figure 2).

Habitat suitability surface values from the seven SDMs ranged from 1 to 1,000. We rescaled these surfaces from 0 to 1 and then

applied the three different transformations to obtain resistance values from 1 to 100. From the PSF-All, PSF-Move and PathSF models, we obtained relative probability-of-use surfaces ranging from 0 to 1 and used the three transformations to convert these to resistance.

We derived our resistance surface from the landscape genetic analysis by summing the resistance surfaces for the variables in the final model and rescaling from 1 to 100. We did not use any transformations since the landscape genetic approach estimates resistance directly. We also developed one additional resistance surface for testing; a multilevel resistance surface (described in Zeller, Vickers, Ernest, & Boyce, 2017) that was the product (rescaled from 1 to 100) of the linearly scaled PathSF resistance surface and the landscape genetic resistance surface (Figure 2). This multilevel resistance surface, which represents both broad-scale dispersal and successful breeding as well as fine-scaled movement decisions, may be superior to either one of these surfaces in isolation (Zeller, Vickers, et al., 2017).

2.4 | Puma dispersal data

Pumas typically disperse from their natal territories as subadults. From our population of collared pumas, we identified 11 subadult dispersers that travelled from their natal home range to a new home range in the study area using net squared displacement curves (Bunnefeld et al., 2011). We extracted GPS locations from the dispersal period and visually identified the maximally displaced start and end points of this dispersal period. For each individual, we used these start and end points as the source and destination points for each connectivity algorithm. We used all the dispersal points (n = 1,044) between these start and end points as our validation data (Figure 1).

2.5 | Connectivity algorithms and model validation

Across each of our 32 resistance surfaces, we ran CS (using CS software; Shah & McRae, 2008) and CD (using the *gdistance* R package) connectivity algorithms between the start and end points of each puma dispersal event. The CS approach produces a cumulative current flow surface between the start and end points, and the CD approach adds together the cost distance kernels from the start and end points, resulting in a surface ranging from least to maximal cost (e.g., Appendix S2). We quantile-rescaled the resultant connectivity surfaces so they were comparable across models and algorithms.

To assess the performance of each connectivity surface, we used three connectivity validation methods as recommended by McClure et al. (2016). For the first validation method, we extracted the values on each connectivity surface at the dispersal points between the start and end points, and calculated the mean and standard error of those values. We assumed better performing surfaces had higher connectivity values at dispersal points than other surfaces. For the second validation method, we

obtained null connectivity surfaces by running the two connectivity algorithms between the start and end points of each dispersal path on a resistance surface where every pixel had a value of one. We then differenced the null connectivity value from the empirical connectivity value at each dispersal point and calculated the mean and standard error of these values. Empirical connectivity surfaces outperform the null connectivity surface when values at the dispersal points are higher on the empirical surfaces than on the null surface and a positive difference is obtained. This method directly assessed the performance of the resistance surfaces by identifying surfaces that significantly outperformed a null, isolation-by-distance model. For the third validation method, we calculated the proportion of dispersal points for each puma that fell within the top 10% of values on the connectivity surface. For management and implementation purposes, corridors are often delineated from a connectivity surface by taking an upper percentage of connectivity values. Therefore, this method determined how well a corridor model derived from these connectivity models captured the dispersal process. We also performed likelihood ratio tests to determine the effect of model type, transformation to resistance and corridor algorithm on each of these three performance measures.

3 | RESULTS

3.1 | Models

Based on the results of our univariate testing of environmental variables (Appendix S3), different puma data types and modelling approaches sometimes selected different scales for the same variable. However, puma preference or avoidance of variables remained relatively consistent across data types and models. Puma SDMs indicated an avoidance of grassland, urban areas, and higher ruggedness and topographic positions, a slight avoidance of agriculture and areas further from water, and a preference for chaparral, forest, elevation, riparian areas and streams (e.g., Appendix S4). All six individual SDMs had an AUC ≥ 0.80, indicating moderate model performance (Manel, Williams, & Ormerod, 2001), and we included all six of these individual SDMs in the ensemble model, which had an AUC of 0.85 (Appendix S5).

Although the final model variables and scales differed slightly among models, the final averaged multi-scale PSF-All, PSF-Move and PathSF models all indicated an avoidance of agriculture, grassland, steep slopes, and human development and a preference for higher elevations, forests, and streams and riparian areas (Appendices S6–S8).

The multi-scale landscape genetics model indicated higher resistance values with higher proportions of agriculture, coastal scrub, grassland, urban, and higher values of distance to water and lower resistance values with higher proportions of canyons, and higher values of elevation and topographic position (Appendix S3). Lower resistance values for forests, streams and ridges were at intermediate proportions of those variables.
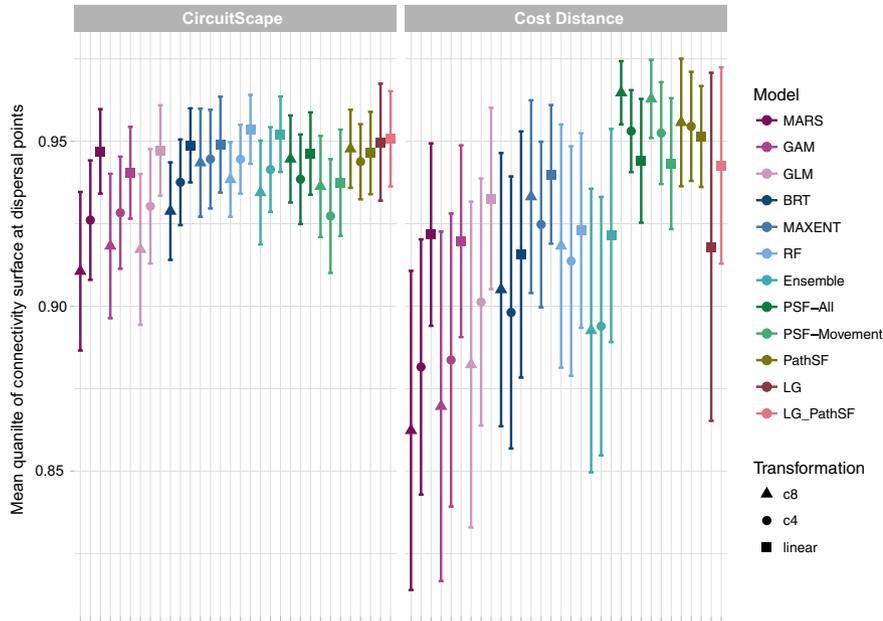
FIGURE 3 Mean and bootstrapped standard errors of the value of each connectivity surface at the dispersal points. Dispersal points were from an independent data set consisting of eleven dispersal events from 2005 to 2012 (*n* = 1,044). The two facets correspond to the connectivity algorithm used, model indicates the data/model type used to estimate resistance, and transformation refers to the transformation used to convert habitat suitability to resistance. [Colour figure can be viewed at wileyonlinelibrary.com]

3.2 | Validation

The first validation method indicated that, in general, all connectivity surfaces performed well, and all had a mean connectivity value at the dispersal points >0.85 (bootstrapped standard errors ranged from a minimum of 0.81 to a maximum of 0.975; Figure 3). All CS surfaces had a mean value at the dispersal points >0.91, while CD PathSFs had mean values >0.95. The results indicated PSF and PathSF models outperformed SDMs when CD was used, and the CD models generally outperformed the CS models. The performance of PSF and PathSF models was very similar, but PSF models were more sensitive to the type of transformation from relative probability of use to resistance with the more negative exponential transformation resulting in higher model performance. The landscape genetic model did not outperform the SDM models, but was improved when combined with the PathSF model. The linear transformation to resistance outperformed the other transformations for the SDM models. The likelihood ratio tests indicated that all three factors—model type, transformation to resistance and connectivity algorithm—influenced connectivity surface performance results with model type having the most influence, followed by connectivity algorithm, and then transformation to resistance (Table 2).

The second validation method indicated that the PathSF model, regardless of transformation, consistently outperformed the null isolation-by-distance model, but only with the CD algorithm (Figure 4). Similarly, the PSF-All model with a c8 resistance transformation and the PSF-Movement model under both inverse exponential transformations also significantly outperformed the null model with the CD algorithm. None of the resistance surfaces derived from SDM or genetic models outperformed the null model with either connectivity algorithm. When the CS algorithm was used, no combination of model type or resistance surface was better than the null model. Likelihood ratio tests again indicated all three factors—model

type, transformation to resistance and connectivity method—influenced model performance with model type having the most influence (Table 2).

The top 10% of each connectivity surface captured at least 65% of the puma dispersal points regardless of model type, transformation to resistance or connectivity algorithm (Figure 5). As a group, the GPS collar and genetic surfaces with the CD algorithm performed the best, capturing at least 83% of the dispersal points and outperforming the SDM surfaces. Although the CD outperformed the CS algorithm for the GPS collar and genetic surfaces, CS tended to outperform the CD algorithm for the SDM models. Transformation to resistance had similar effects on connectivity surface performance to the first validation method. Likelihood ratio test results were similar to the previous two validation methods (Table 2).

4 | DISCUSSION

To our knowledge, this is the first study to assess the performance of multiple data types and connectivity algorithms in representing the dispersal process for a species. Although our results were sometimes noisy, we were able to reach general conclusions about data type, connectivity algorithm and the relationship between habitat suitability and resistance. Overall, our results indicated that the choice of data type for estimating resistance and modelling connectivity was the most influential factor in capturing the dispersal process. In general, we found that models based on GPS collar data, genetic data or a combination of the two data types outperformed those based on opportunistic presence-only data. The PathSF models had consistently good performance across the three validation methods and were the only models that, regardless of the transformation to resistance, outperformed the null isolation-by-distance

TABLE 2 Likelihood ratio test results. The three validation response variables (quantile values at dispersal points, differences in values at dispersal points between empirical and null models, and proportion of dispersal points in the top 10% of the connectivity surface) modelled as a function of (1) model type, (2) transformation to resistance and (3) connectivity algorithm. Likelihood ratio tests were performed comparing the full model with each of these left out in turn

	Quantile value at dispersal points			Comparison with distance-only model			Proportion of dispersal points in corridor		
	Log-Likelihood	χ^2	*p*-value	Log-Likelihood	χ^2	*p*-value	Log-Likelihood	χ^2	*p*-value
Model type	46,235			64,170			86		
(*df* = 11)	45,406	1,659	<2.2e-16	62,818	2,704	<2.2e-16	61	49	8.1e-07
Transformation to	46,235			64,170			86		
resistance (*df* = 2)	46,140	190	<2.2e-16	64,014	311	<2.2e-16	82	7.1	.028
Connectivity	46,235			64,170			86		
algorithm (*df* = 1)	45,648	1,176	<2.2e-16	63,977	384	<2.2e-16	82	7	.008

FIGURE 4 Mean difference between the value of each connectivity surface at the dispersal points and the value of the null isolation-by-distance model at the dispersal points. Dispersal points were from an independent data set consisting of eleven dispersal events from 2005 to 2012 (*n* = 1,044). Bootstrapped standard error bars are also shown. Values greater than zero (horizontal dotted line) indicate statistically significant improvement over the null model. The two facets correspond to the connectivity algorithm used, model indicates the data/model type used to estimate resistance, and transformation refers to the transformation used to convert habitat suitability to resistance. [Colour figure can be viewed at wileyonlinelibrary.com]
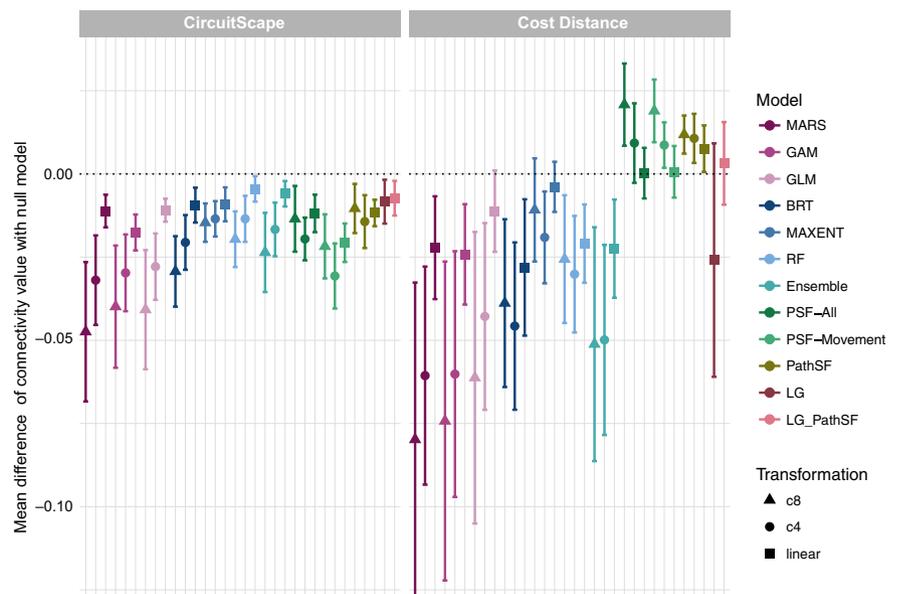
model. Choice of connectivity algorithm also influenced results with the CD algorithm producing higher model performances than the CS algorithm for the GPS collar and genetic data types. Generally, the CS algorithm was less sensitive than CD to the data type used to build the resistance surface, but only CD methods showed improvement over null models. Using a negative exponential relationship between habitat suitability and resistance as opposed to a direct linear transformation had the least effect on model performance.

Our first prediction—that all data types would produce qualitatively similar resistance surfaces with many of the same variables having the same relationship to resistance—was supported. We found that all models, regardless of whether they were derived from opportunistic presence-only data, GPS collar data or genetic data, had six common variables that have been shown to influence puma habitat use and movement in California (Burdett et al., 2010; Dickson, Jenness, & Beier, 2005; Wilmers et al., 2013; Zeller et al., 2014, 2016). All models indicated pumas avoided agricultural, grassland and urban areas, and preferred forests, higher elevations, and

riparian areas or streams. Despite the fact that all these models qualitatively tell the same story and produce similar patterns of resistance, they resulted in different predictions of connectivity. The higher performance of the surfaces derived from the GPS collar data indicates that these resistance surfaces more closely matched the factors pumas were responding to during dispersal than the other data types (Cushman et al., 2014). These results are similar to what Cushman and Lewis (2010) and Cushman et al. (2014) found for black bears in northern Idaho. PathSF and genetic resistance surfaces for black bears both indicated movement was facilitated at mid-elevations with high forest cover and hampered by human development. However, their PathSF connectivity models more effectively captured movement than the landscape genetic models.

Our second expectation—that connectivity models based on movement data would outperform other data types—was only partially supported. Our second validation method addressed the performance of the resistance surfaces directly and indicated that PathSF and PSF models were a significant improvement over the null
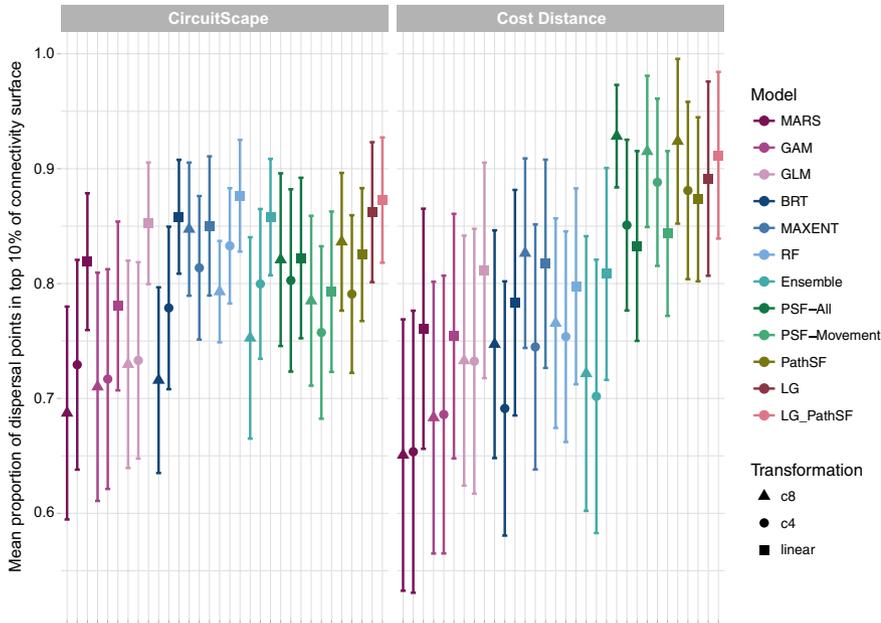
FIGURE 5 Mean proportion of dispersal points in the top 10% of each connectivity surface. Dispersal points were from an independent data set consisting of eleven dispersal events from 2005 to 2012 (*n* = 1,044). Bootstrapped standard error bars are also shown. The two facets correspond to the connectivity algorithm used, model indicates the data/model type used to estimate resistance, and transformation refers to the transformation used to convert habitat suitability to resistance. [Colour figure can be viewed at wileyonlinelibrary.com]

model when a CD algorithm was used, but the resistance transformation differentially affected their performance. PathSF was least sensitive to the transformation, performing equally well under the three transformations whereas PSF-Move and PSF-All only outperformed the null model when an inverse exponential transformation was used—either c4 or c8 for PSF-Move or c8 for PSF-All. The SDMs never significantly outperformed the null model. We expected the PSF-All models to perform worse than the PSF-Movement and PathSF models, but results between these data types were extremely similar, indicating all are adequate for capturing the dispersal process for pumas. This is encouraging as home range use data already exist for many imperilled species, such as large carnivores, and therefore, functional connectivity may be assessed relatively rapidly (Fattebert et al., 2015). These results agree with those found by Fattebert et al. (2015), Keeley et al. (2016) and Keeley et al. (2017) for leopards, desert bighorn sheep and kinkajous, respectively, and Newby (2011) for pumas. However, these results contrast those found by Elliot et al. (2014), Gastón et al. (2016) and Jackson et al. (2016) for African lions, Iberian lynx and African wild dogs. This may be due to differences in taxa or in methodologies as these studies did not compare the ability of corridors to capture dispersal movement, but instead compared corridors derived from dispersal data with corridors derived from other data types. Resistance surfaces derived solely from genetic data performed poorly compared with those derived from GPS collar data. Cushman et al. (2014) found similarly poor performance of genetic data for predicting within home-range movement for black bears compared with PathSFs.

Our third prediction—that we would observe higher model performance with CD than CS—was supported. CD outperformed CS when GPS collar and genetic data were used and, unlike CS, was able to outperform the null isolation-by-distance models for these data types. In comparing the proportion of dispersal points captured in the corridors, CD tended to outperform CS when resistance

surfaces were derived from GPS collar or genetic data. However, for the first validation method, our results are consistent with those found by McRae and Beier (2007) indicating higher performance of CS when compared with CD for predicting genetic relatedness. Across both validation methods, our results contrast those found by McClure et al. (2016) who found CS models outperformed CD models when using a PSF with wolverine dispersal data. The PSF they used was derived from VHF, not GPS telemetry data, and points were collected every 10 days. With this long acquisition interval, it is possible these data were more akin to the opportunistic presence-only data used in this study, which may explain the difference in results. LaPoint, Gallery, Wikelski, and Kays (2013) found that CS outperformed CD for predicting within home-range movements of fishers, although neither method performed well.

The good performance of the CD algorithm may also be explained by the similarities of CD to resistant kernels in the context of our analysis (Compton, McGarigal, Cushman, & Gamble, 2007). Resistant kernels are built on the concept of cost distance, but instead of pairing source and destination points, cost distance is allowed to flow from multiple source points (or every pixel in the landscape) until a biologically relevant cost distance (based on known dispersal distances for a species) has been reached. The cost distance kernels from all source points are then summed to obtain a single surface. In our analysis, we summed the cost distances from the start and the end points of a dispersal path, and therefore, the CD algorithm we used are equivalent to resistant kernels. However, for connectivity modelling across a landscape of interest, resistant kernels have many advantages over CD algorithms: resistant kernels are not constrained by source–destination points, they do not assume individuals have complete knowledge of the landscape, and the flow can be limited by biologically based dispersal distances (Cushman & Landguth, 2012; Cushman et al., 2013). Given the performance of CD models and the advantages of resistant kernels, we would recommend resistant kernels as the connectivity algorithm of choice.

The shape of the transformation from habitat suitability to resistance was also influential in the ability of a connectivity surface to capture the dispersal process, but the results were mixed. General trends indicate that the linear transformation from habitat suitability to resistance resulted in higher model performance for the SDM models, but the c8 transformation resulted in higher model performance for the models based on GPS collar data. However, within a model type, using one transformation over another rarely resulted in significant differences in model performance. We would have assumed that the habitat suitability derived from SDMs reflected habitat selection by individuals to maximize fitness within the home range and not movement per se and that a more relaxed transformation to resistance would have more adequately captured the dispersal process. However, we did not find this to be the case. Our results agreed with those of Mateo-Sánchez et al. (2015a,b), Keeley et al. (2016), and Keeley et al. (2017) indicating a negative exponential transformation to resistance from habitat use derived from GPS collar data is more appropriate than a negative linear transformation and support the notion by Beier et al. (2008) and Cushman et al. (2013) that within home-range habitat use does not directly translate to resistance.

Additional research is needed to determine whether different modelling approaches or connectivity algorithms could improve upon the model performances we observed. Exploring connectivity surface performance by sex would be an additional component that may offer more insights into model performance. For example, studies modelling African lion and brown bear movement found substantially different corridor locations when males and females were modelled separately (Elliot et al., 2014; Maiorano, Boitani, Chiaverini, & Ciucci, 2017). This is more likely to be a factor in species where dispersal behaviour and distances vary with sex, such as in large carnivores. More studies such as this, in addition to simulation modelling with species of different vagilities in different landscapes with varying habitat characteristics, would aid in identifying generalizable conclusions regarding data type and connectivity algorithm for modelling wildlife corridors. Including occupancy models and camera trap data would also be very informative as these data types are becoming increasingly widespread and available.

Our study was focused on a single, generalist species in one geographic location and thus was limited in scope. Nonetheless, we are able to offer general recommendations regarding modelling connectivity for pumas and an analytical paradigm to be tested for other species. Resistance surfaces estimated from GPS collar data (either via PSFs or PathSFs) are recommended over those estimated with opportunistic presence-only or genetic data. If GPS collar data can be obtained at an adequate acquisition interval for capturing movement paths (every 5–30 min for puma; Zeller et al., 2016), we recommend PathSF models over PSF models as, unlike PSF models, all the PathSF models outperformed the null model. We also recommend genetic data over opportunistic presence-only data, given our results and those of Mateo-Sánchez et al. (2015a). Genetic data directly measure not only successful dispersal, but also dispersal

that resulted in breeding, which is the most biologically relevant issue for functional connectivity (Zeller, Vickers et al., 2017). For connectivity approaches, we recommend the CD algorithm, specifically the improved resistant kernel algorithm over CS. However, if only opportunistic presence-only data are available, the dispersal process may be better captured with the CS algorithm as we found the CS algorithm to be less sensitive to data type. Identifying or assessing functional connectivity from more readily available empirical presence-only data is an improvement over expert opinion (Clevenger & Wierzchowski, 2002; Poor, Loucks, Jakes, & Urban, 2012; Shirk, Wallin, Cushman, Rice, & Warheit, 2010). When SDMs are applied with thoughtful selection of scales, modelling approach, transformation and connectivity algorithm, they may be able to capture the dispersal process relatively well. We found that SDMs captured approximately 65%–87% of the dispersal process across connectivity algorithms.

As the goal of most connectivity studies is to identify conservation corridors, capturing the dispersal process within corridors is paramount for providing functional connectivity for wildlife. We were encouraged to find that, when GPS collar or genetic data were used, 75%–87% of the dispersal process was captured within the CS corridors and 83%–93% of the dispersal process was captured within the CD corridors. This indicates that for both GPS collar and genetic data, and regardless of transformation to resistance or connectivity algorithm, functional corridors can be robustly identified for terrestrial species using currently available methods.

DATA AVAILABILITY

Genetic data are available from the Dryad database (https://doi.org/10.5061/dryad.1kh2n). Telemetry data for building the resource selection functions are available on the Figshare database (https://doi.org/10.6084/m9.figshare.4983095). All geospatial data are freely available.

ORCID

Katherine A. Zeller (iD) http://orcid.org/0000-0002-2913-6660

Megan K. Jennings (iD) http://orcid.org/0000-0002-3707-851X

Holly B. Ernest (iD) http://orcid.org/0000-0002-0205-8818

REFERENCES

Abrahms, B., Sawyer, S. C., Jordan, N. R., McNutt, J. W., Wilson, A. M., & Brashares, J. S. (2017). Does wildlife resource selection accurately inform corridor conservation? *Journal of Applied Ecology*, *54*: 412-422. https://doi.org/10.1111/1365-2664.12714

Adriaensen, F., Chardon, J. P., De Blust, G., Swinnen, E., Villalba, S., Gulinck, H., & Matthysen, E. (2003). The application of "least-cost" modelling as a functional landscape model. *Landscape and Urban Planning*, *64*, 233–247. https://doi.org/10.1016/s0169-2046(02)00242-6

Beier, P., Majka, D. R., & Spencer, W. D. (2008). Forks in the road: Choices in procedures for designing wildland linkages. *Conservation Biology*, *22*, 836–851. https://doi.org/10.1111/j.1523-1739.2008.00942.x

Bunnefeld, N., Borger, L., van Moorter, B., Rolandsen, C. M., Dettki, H., Solberg, E. J., & Ericsson, G. (2011). A model-driven approach to quantify migration patterns: Individual, regional and yearly differences. *Journal of Animal Ecology*, *80*, 466–476. https://doi.org/10.1111/j.1365-2656.2010.01776.x

Burdett, C. L., Crooks, K. R., Theobald, D. M., Wilson, K. R., Boydston, E. E., & Lyren, L. M., ... Boyce, W. M. (2010). Interfacing models of wildlife habitat and human development to predict the future distribution of puma habitat. *Ecosphere*, *1*, 1–21. https://doi.org/10.1890/es10-00005.1

Chen, I. C., Hill, J. K., Ohlemuller, R., Roy, D. B., & Thomas, C. D. (2011). Rapid range shifts of species associated with high levels of climate warming. *Science*, *333*, 1024–1026. https://doi.org/10.1126/science.1206432

Clevenger, A., & Wierzchowski, J. (2002). GIS-generated, expert-based models for identifying wildlife habitat linkages and planning mitigation passages. *Conservation Biology*, *16*, 503–514. https://doi.org/10.1046/j.1523-1739.2002.00328.x

Compton, B. W., McGarigal, K., Cushman, S. A., & Gamble, L. R. (2007). A resistant-kernel model of connectivity for amphibians that breed in vernal pools. *Conservation Biology*, *21*, 788–799. https://doi.org/10.1111/j.1523-1739.2007.00674.x

Cushman, S. A., & Landguth, E. L. (2012). Multi-taxa population connectivity in the Northern Rocky Mountains. *Ecological Modelling*, *231*, 101–112. https://doi.org/10.1016/j.ecolmodel.2012.02.011

Cushman, S. A., & Lewis, J. S. (2010). Movement behavior explains genetic differentiation in American black bears. *Landscape Ecology*, *25*, 1613–1625. https://doi.org/10.1007/s10980-010-9534-6

Cushman, S., Lewis, J., & Landguth, E. (2014). Why did the bear cross the road? Comparing the performance of multiple resistance surfaces and connectivity modeling methods. *Diversity*, *6*, 844–854. https://doi.org/10.3390/d6040844

Cushman, S. A., McRae, B., Adriaensen, F., Beier, P., Shirley, M., & Zeller, K. A. (2013). Biological corridors and connectivity. In D. MacDonald & K. Willis (Eds.), *Key topics in conservation biology 2* (pp. 384–404). Hoboken, NJ: Wiley-Blackwell.https://doi.org/10.1002/9781118520178.ch21

Dickson, B. G., Jenness, J. S., & Beier, P. (2005). Influence of vegetation, topography, and roads on cougar movement in southern California. *Journal of Wildlife Management*, *69*, 264–276. https://doi.org/10.2193/0022-541x(2005)069<0264:iovtar>2.0.co;2

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

Fagan, W. F., & Calabrese, J. M. (2006). Quantifying connectivity: Balancing metric performance with data requirements. In K. R. Crooks & M. Sanjayan (Eds.), *Connectivity conservation* (pp. 297–317). New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9780511754821

Fattebert, J., Robinson, H. S., Balme, G., Slotow, R., & Hunter, L. (2015). Structural habitat predicts functional dispersal habitat of a large carnivore: how leopards change spots. *Ecological Applications*, *25*, 1911–1921. https://doi.org/10.1890/14-1631.1

Gastón, A., Blázquez-Cabrera, S., Garrote, G., Mateo-Sánchez, M. C., Beier, P., Simón, M. A., & Saura, S. (2016). Response to agriculture by a woodland species depends on cover type and behavioral state: Insights from resident and dispersing Iberian lynx. *Journal of Applied Ecology*, *53*, 814–824. https://doi.org/10.1111/1365-2664.12629

Gilbert-Norton, L., Wilson, R, Stevens, J. R., & Beard, K. H. (2010). A meta-analytic review of corridor effectiveness. *Conservation Biology*, *24*, 660–668. https://doi.org/10.1111/j.1523-1739.2010.01450.x

Gilleland, E. (2013). Two-dimensional kernel smoothing: Using the R package smoothie. NCAR Technical Note, TN-502 + STR, 17 pp. Retrieved from- http://opensky.library.ucar.edu/collections/TECH-NOTE-000- 000-000-869

Haddad, N. M., Brudvig, L. A., Clobert, J., Davies, K. F., Gonzalez, A., Holt, R. D., & Lovejoy, T. E. (2015). Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science Advances*, *1*, e1500052. https://doi.org/10.1126/sciadv.1500052

Hilty, J. A., Lidicker, W. Z. Jr, & Merenlender, A. M. (2012). *Corridor Ecology: The science and practice of linking landscapes for biodiversity conservation*. Washington, DC: Island Press.

Jackson, C. R., Marnewick, K., Lindsey, P., Roskaft, E., & Robertson, M. P. (2016). Evaluating habitat connectivity methodologies: A case study with endangered African wild dogs in South Africa. *Landscape Ecology*, *31*, 1433–1447. https://doi.org/10.1007/s10980-016-0342-5

Jin, S., Yang, L., Danielson, P., Homer, C., Fry, J., Xian, G. (2013). A comprehensive change detection method for updating the National Land Cover Database to circa 2011. *Remote Sensing of Environment*, *132*, 159–175.

Keeley, A. T. H., Beier, P., & Gagnon, J. W. (2016). Estimating landscape resistance from habitat suitability: Effects of data source and nonlinearities. *Landscape Ecology*, *31*, 2151–2162. https://doi.org/10.1007/s10980-016-0387-5

Keeley, A. T. H., Beier, P., Keeley, B. W., & Fagan, M. E. (2017). Habitat suitability is a poor proxy for landscape connectivity during dispersal and mating movements. *Landscape and Urban Planning*, *161*, 90–102. https://doi.org/10.1016/j.landurbplan.2017.01.007

LaPoint, S., Gallery, P., Wikelski, M., & Kays, R. (2013). Animal behavior, cost-based corridor models and real corridors. *Landscape Ecology*, *28*, 1615–1630. https://doi.org/10.1007/s10980-013-9910-0

Maiorano, L., Boitani, L., Chiaverini, L., & Ciucci, P. (2017). Uncertainties in the identification of potential dispersal corridors: The importance of behavior, sex, and algorithm. *Basic and Applied Ecology*, https://doi.org/10.1016/j.baae.2017.02.005

Manel, S., Williams, H. C., & Ormerod, S. J. (2001). Evaluating presence-absence models in ecology: The need to account for prevalence. *Journal of Applied Ecology*, 38, 921–931.

Mateo-Sánchez, M. C., Balkenhol, N., Cushman, S., Pérez, T., Domínguez, A., & Saura, S. (2015a). A comparative framework to infer landscape effects on population genetic structure: Are habitat suitability models effective in explaining gene flow? *Landscape Ecology*, 30, 1405–1420. https://doi.org/10.1007/s10980-015-0194-4

Mateo-Sánchez, M. C., Balkenhol, N., Cushman, S., Pérez, T., Domínguez, A., & Saura, S. (2015b). Estimating effective landscape distances and movement corridors: Comparison of habitat and genetic data. *Ecosphere*, 6, 1–16. https://doi.org/10.1890/ES14-00387.1

McClure, M. L., Hansen, A. J., & Inman, R. M. (2016). Connecting models to movements: Testing connectivity model predictions against empirical migration and dispersal data. *Landscape Ecology*, 31, 1419–1432. https://doi.org/10.1007/s10980-016-0347-0

McGarigal, K., Wan, H. Y., Zeller, K. A., Timm, B. C., & Cushman, S. A. (2016). Multi-scale habitat modeling: A review and outlook. *Landscape Ecology*, 31, 1161–1175. https://doi.org/10.1007/s10980-016-0374-x

McRae, B. H. (2006). Isolation by resistance. *Evolution*, 60(8), 1551–1561. https://doi.org/10.1111/j.0014-3820.2006.tb00500.x

McRae, B. H., & 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, 19885–19890. https://doi.org/10.1073/pnas.0706568104

Newby, J. R. (2011). *Puma dispersal ecology in the Central Rocky Mountains*. Theses, Dissertations, Professional Papers. Paper 786.

Nicholson, E., Westphal, M. I., Frank, K., Rochester, W. A., Pressey, R. L., Lindenmayer, D. B., & Possingham, H. P. (2006). A new method for conservation planning for the persistence of multiple species. *Ecology Letters*, 9, 1049–1060. https://doi.org/10.1111/j.1461-0248.2006.00956.x

Poor, E. E., Loucks, C., Jakes, A., & Urban, D. L. (2012). Comparing habitat suitability and connectivity modeling methods for conserving pronghorn migrations. *PLoS One*, 7, e49390. https://doi.org/10.1371/journal.pone.0049390

R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.

Shah, V. B., & McRae, B. (2008). Circuitscape: A tool for landscape ecology. In G. Varoquaux, T. Vaught & J. Millman (Eds.), *Proceedings of the 7th Python in Science Conference* (pp. 62–66).

Shirk, A. J., Wallin, D. O., Cushman, S. A., Rice, C. G., & Warheit, K. I. (2010). Inferring landscape effects on gene flow: A new model selection framework. *Molecular Ecology*, 19, 3603–3619. https://doi.org/10.1111/j.1365-294X.2010.04745.x

Theobald, D. (2006). Exploring the functional connectivity of landscapes using landscape networks. In K. R. Crooks & M. Sanjayan (Eds.), *Connectivity conservation* (pp. 416–443). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511754821

Wilmers, C. C., Wang, Y., Nickel, B., Houghtaling, P., Shakeri, Y., Allen, M. L., … Williams, T. (2013). Scale dependent behavioral responses to human development by a large predator, the puma. *PLoS One* 2013; https://doi.org/10.1371/journal.pone.0060590

Zeller, K. A., McGarigal, K., Beier, P., Cushman, S. A., Vickers, T. W., & Boyce, W. M. (2014). Sensitivity of landscape resistance estimates based on point selection functions to scale and behavioral state: Pumas as a case study. *Landscape Ecology*, 29, 541–557. https://doi.org/10.1007/s10980-014-9991-4

Zeller, K. A., McGarigal, K., Beier, P., Cushman, S. A., Vickers, T. W., & Boyce, W. M. (2016). Using step and path selection functions for estimating resistance to movement: Pumas as a case study. *Landscape Ecology*, 31, 1319–1335. https://doi.org/10.1007/s10980-015-0301-6

Zeller, K. A., McGarigal, K., Cushman, S. A., Beier, P., Vickers, T. W., & Boyce, W. M. (2017a). Sensitivity of resource selection and connectivity models to landscape definition. *Landscape Ecology*, 32, 835–855. https://doi.org/10.1007/s10980-017-0489-8

Zeller, K. A., McGarigal, K., & Whiteley, A. R. (2012). Estimating landscape resistance to movement: A review. *Landscape Ecology*, 27, 777–797. https://doi.org/10.1007/s10980-012-9737-0

Zeller, K. A., Vickers, T. W., Ernest, H. B., & Boyce, W. M. (2017b). Multi-level, multi-scale resource selection functions and resistance surfaces for conservation planning: Pumas as a case study. *PLoS One* 12(6): e0179570. https://doi.org/10.1371/journal.pone.0179570

BIOSKETCH

Dr. Katherine Zeller is a landscape ecologist/wildlife biologist with interests in movement and connectivity modelling for wildlife conservation.

Author contributions: K.Z., M.J. and S.C. conceived the study and designed the methodology; K.Z. and M.J. analysed the data; T.W.V. and W.B. collected the GPS telemetry data; H.E. collected and genotyped the genetic data; K.Z. led the writing of the manuscript; and all authors contributed to drafts and gave final approval for publication.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.