

Sensitivity of resource selection and connectivity models to landscape definition

Katherine A. Zeller  · Kevin McGarigal · Samuel A. Cushman · Paul Beier · T. Winston Vickers · Walter M. Boyce

Received: 19 August 2016 / Accepted: 22 January 2017 / Published online: 22 February 2017
© Springer Science+Business Media Dordrecht 2017

Abstract

Context The definition of the geospatial landscape is the underlying basis for species-habitat models, yet sensitivity of habitat use inference, predicted probability surfaces, and connectivity models to landscape definition has received little attention.

Objectives We evaluated the sensitivity of resource selection and connectivity models to four landscape definition choices including (1) the type of geospatial layers used, (2) layer source, (3) thematic resolution, and (4) spatial grain.

Methods We used GPS telemetry data from pumas (*Puma concolor*) in southern California to create

multi-scale path selection function models (PathSFs) across landscapes with 2500 unique landscape definitions. To create the landscape definitions, we identified seven geospatial layers that have been shown to influence puma habitat use. We then varied the number, sources, spatial grain, and thematic resolutions of these layers to create our suite of plausible landscape definitions. We assessed how PathSF model performance (based on AIC) was affected by landscape definition and examined variability among the predicted probability of movement surfaces, connectivity models, and road crossing locations.

Results We found model performance was extremely sensitive to landscape definition and identified only seven top models out of our suite of definitions (<1%). Spatial grain and the number of geospatial layers selected for a landscape definition significantly affected model performance measures, with finer grains and greater numbers of layers increasing model performance.

Conclusions Given the sensitivity of habitat use inference, predicted probability surfaces, and connectivity models to landscape definition, our results indicate the need for increased attention to landscape definition in future studies.

Electronic supplementary material The online version of this article (doi:[10.1007/s10980-017-0489-8](https://doi.org/10.1007/s10980-017-0489-8)) contains supplementary material, which is available to authorized users.

K. A. Zeller (✉) · K. McGarigal
Department of Environmental Conservation, University of
Massachusetts, Amherst, MA 01003, USA
e-mail: kathyzeller@gmail.com

S. A. Cushman
U.S. Forest Service Rocky Mountain Research Station,
2500 S Pine Knoll Dr., Flagstaff, AZ 86001, USA

P. Beier
School of Forestry, Northern Arizona University,
Flagstaff, AZ 86011, USA

T. W. Vickers · W. M. Boyce
Wildlife Health Center, School of Veterinary Medicine,
University of California Davis, Davis, CA 95616, USA

Keywords Species-habitat relationships · Resource selection function · Spatial grain · Thematic resolution · *Puma concolor* · Connectivity · Corridor · Resistance

Introduction

Assessing species-habitat relationships and modeling connectivity requires creating a spatially explicit landscape model to formally represent (1) the types of landscape features that may affect habitat use and movement, (2) the spatial heterogeneity of those landscape features, and (3) the spatial scale of those landscape features. This ‘landscape definition’ is the basis for all habitat use and connectivity models (Cushman et al. 2013), yet the sensitivity of these models to landscape definition has received scant attention in the literature.

Landscape ecologists have long been aware that observed pattern-process relationships are highly sensitive to the spatial scale of the landscape model (Wiens 1989; Wu 2004). Spatial scale is the marriage of two components: extent and grain. Wildlife biologists have traditionally been focused on the former of these two scale components, the spatial extent, for wildlife-habitat inference (McGarigal et al. 2016). Here the most common approach is to summarize landscape features within buffers or kernels of varying size (‘ecological neighborhoods’ sensu Addicott et al. 1987) in order to determine the characteristic spatial scale of selection for a landscape feature (e.g., Holland et al. 2004). It is increasingly recognized that failure to identify the characteristic spatial scale of selection with regards to both grain and extent may bias wildlife-habitat inference (McGarigal et al. 2016).

Wildlife biologists are just beginning to examine the effects of spatial grain (also often referred to as “resolution”) on species-habitat models. Varying the grain size of the geospatial layers has been used to both determine the characteristic scale of selection for a landscape feature (Thompson and McGarigal 2002) and to determine how spatial grain affects overall model performance (Karl et al. 2000; Seoane et al. 2004; Venier et al. 2004; Guisan et al. 2007; Cushman and Landguth 2010a; Gottschalk et al. 2011). In general, these studies have found model performance decreases with increasing grain size, though the effect of grain size on model predictive performance remains equivocal (Tobalske 2002; Seoane et al. 2004; Guisan et al. 2007).

Thematic resolution of the geospatial layers used for a landscape definition has received far less attention than spatial scale in modeling species-habitat relationships. Thematic resolution refers to the level of

heterogeneity of the geospatial layers. In many cases, landscape features can be represented as a continuous gradient, which is thought to more closely mimic real world landscapes and reduce subjectivity (McGarigal and Cushman 2005; Cushman et al. 2010b). With continuous gradients, the thematic resolution is at its greatest given the precision of the raw data. However, landscape features can also be represented categorically, as in the classic patch-mosaic model of landscape structure (Forman 1995). When categorical layers are used, decisions must be made regarding the number and breakpoints of the classes. Lawler et al. (2004) found that species distribution models had similar fit between different thematic resolutions, but that predictions differed in various geographic locations. Seoane et al. (2004) found that finer thematic resolutions resulted in better predictive performance of species distribution models. Cushman and Landguth (2010a) found that the strength of the relationship between gene flow and landscape features increased as the thematic resolution of the layers increased from the lowest resolution of two classes to the highest resolution (continuously scaled layers). They also found that thematic resolution was the dominant factor over spatial grain and spatial extent in defining the landscape for landscape genetic analyses.

Selecting the spatial grain and thematic resolution of the geospatial layers are not the only decisions one must contend with when defining the landscape. The layers themselves must be selected (i.e., thematic content; e.g., elevation, land cover type, etc.) along with the data source of each layer. Selection of geospatial layers is often determined a priori given previous knowledge of the target species or through model selection procedures such as Akaike’s Information Criterion (AIC; Burnham and Anderson 2002). Furthermore, multiple data sources may be available for the chosen layers. For example, there may be many data sources for land cover type that have similar accuracy, but trade-offs may exist across the study area such that one data source may be very good at representing riparian areas but another is better at differentiating scrub from grassland. Source of the data layers in species-habitat models is often not discussed, though when layers from different sources have been compared (e.g., Chust et al. 2004; Seoane et al. 2004; Cushman et al. 2010a), the comparison is often confounded with spatial grain. For example, one

data source available at a 30 m grain (e.g., Landsat) is compared with another at a 250 m grain (e.g., CORINE; European Environmental Agency 2010). Therefore, the effect of these choices on species-habitat models remains unclear.

Defining the landscape for modeling species-habitat relationships is not a straightforward task and one is faced with many choices, all of which may affect the resulting models and conclusions about species-habitat use. We are not aware of any studies that have looked at how species-habitat models are affected by all four of these landscape definition choices: (1) spatial grain, (2) thematic resolution, (3) which and how many geospatial layers to include in a definition, and (4) which data source to include for each geospatial layer.

Because our collective interest is in modeling wildlife movement and connectivity, we used GPS data from pumas (*Puma concolor*) in southern California to explore the sensitivity of multi-scale path selection function (PathSF) models (Cushman and Lewis 2010; Cushman et al. 2010b; Zeller et al. 2015) to landscape definition. We hypothesized that model performance (as defined below) would be sensitive to landscape definition and, specifically, that model performance would increase with (1) decreasing spatial grain, (2) increasing thematic resolution, and (3) increasing number of geospatial layers, provided all the layers are true drivers of habitat selection. We also predicted that some data sources would improve model performance measures more than others. We used the Temecula Corridor region as a vignette to show the results of different landscape definitions for estimating the relative probability of movement, connectivity and road crossing locations.

Methods

Study area and puma data

Our study area (4089 km²) included the Santa Ana Mountains of southern California and surrounding lowlands (Fig. 1). This coastal mountain range experiences a Mediterranean climate with hot, dry summers and mild, wetter winters. Between October 2011 and March 2014, we fit 10 pumas (six female and four male) with Lotek 4400S GPS collars programmed at a 5 min acquisition interval (Lotek Wireless, Inc.

Canada). Collar duration ranged from nine to 71 days (median = 29). Manufacturer tests indicate that long-term positional accuracy of the GPS collars is 5–10 m, though this may vary with certain vegetation types and topographical conditions (J. Chang, personal communication). To avoid the use of data that may have large spatial errors, we removed two-dimensional fixes with a position dilution of precision (PDOP) >5 (Lewis et al. 2007). This data filtering resulted in a final data set of 75,716 fixes across the 10 individuals (range 1650–18,464; median = 7147). We pooled sexes in the analysis due to the low number of individuals. Daily paths were constructed for each puma by connecting consecutive 5-min points with straight-line segments over a 24-h period. This resulted in 315 daily paths for use in the PathSFs.

Geospatial data

We used the following seven geospatial layers which have been shown to influence puma habitat use: (1) elevation (Alexander et al. 2006; Burdett et al. 2010; Wilmers et al. 2013; Allen et al. 2014), (2) percent slope (Dickson et al. 2005; Dickson and Beier 2007; Wilmers et al. 2013), (3) terrain ruggedness (represented as total curvature; Burdett et al. 2010), (4) land cover type (Burdett et al. 2010; Wilmers et al. 2013), (5) percent vegetative cover (Holmes and Laundré 2006; Kissling et al. 2009), (6) roads (Dickson et al. 2005; Wilmers et al. 2013; Gray et al. 2016), and (7) human development (represented here as percent impervious surface; Burdett et al. 2010; Wilmers et al. 2013). We derived percent slope and terrain ruggedness from the National Elevation Dataset (30 m; USGS 2009) using the percent slope and total curvature tools in the DEM Surface Tools Extension for ArcMap (Jenness 2013).

Some of the geospatial layers were available across our study area from multiple sources. To examine the possible effect of data source in our analyses, we selected three different sources for land cover type and percent vegetative cover, and two different sources for roads. We assumed most available elevation layers would have little error and be very similar to each other. Therefore, we selected only one data source for elevation and its derived layers (percent slope and terrain ruggedness). All geospatial layers and their sources are provided in Table 1.

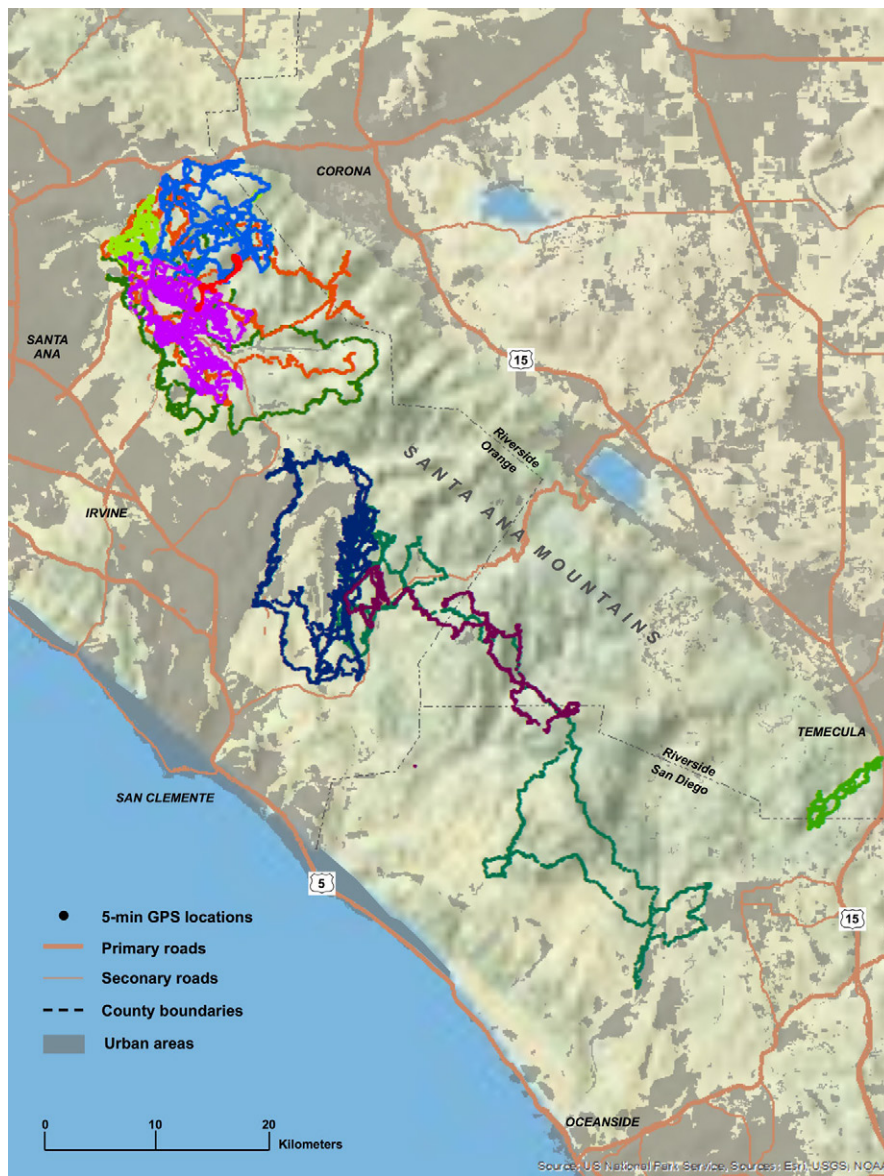


Fig. 1 Santa Ana Mountain study area and surrounding region. Puma 5-min GPS telemetry points used in the analysis are shown. Colors indicate different individuals. (Color figure online)

We represented layers that were available in a continuous format (elevation, percent slope, terrain ruggedness, percent vegetative cover, and percent impervious surface) with four thematic resolutions: continuous, three classes, four classes, and five classes. Class breakpoints were determined using the Jenks optimization method (Jenks 1967). This method identifies breakpoints that minimize the within-class variance and maximize the between-class variance. Classifications of each continuous layer are provided

in Online Appendix 1. We represented land cover type, a categorical-only layer, using five classes (Agriculture/Urban, Barren/Scrub, Chaparral, Grassland, Woodland) or eight classes (Agriculture, Barren, Chaparral, Grassland, Riparian, Scrub, Urban, Woodland). These classes were determined based on the dominant vegetative classes and earlier resource selection functions conducted on pumas in the study area (Zeller et al. 2014, 2015). We represented roads, another categorical-only layer, with two, three, or four

Table 1 Data source and year of geospatial data layers used to model puma movement in southern California

Geospatial data layer	Source	Year	Citation
Elevation	National Elevation Dataset (<i>raster</i>)	2009	USGS (2009)
Percent slope	Calculated from the National Elevation Dataset	–	
Terrain ruggedness	Calculated from the National Elevation Dataset	–	
Percent impervious surface	National Land Cover Database (<i>raster</i>)	2011	Jin et al. (2013)
Land cover type	CalVeg (<i>vector</i>)	2014	USDA (2014)
	LandFire, Existing Vegetation Type (<i>raster</i>)	2012	LandFire (2012b)
	National Land Cover Database (<i>raster</i>)	2011	Jin et al. (2013)
Percent vegetative cover	LandFire, Existing Vegetation Cover (<i>raster</i>)	2012	LandFire (2012a)
	Landsat, Vegetation Continuous Fields (<i>raster</i>)	2005	Sexton et al. (2013)
	National Land Cover Database (<i>raster</i>)	2011	Jin et al. (2013)
Roads	Open Street Map (<i>vector</i>)	2014	OpenStreetMap (2014)
	County Roads Data		
	Orange County (<i>vector</i>)	2011	OCTA (2011)
	Riverside County (<i>vector</i>)	2013	Riverside GIS (2013)
	San Bernadino County (<i>vector</i>)	2014	San Bernadino (2014)
	San Diego County (<i>vector</i>)	2013	SanGIS (2014)

County roads data were merged across the four counties in our study area to create a single coverage. *Raster* or *vector* indicate the original format of the data

classes. This allowed us to represent (1) primary and secondary roads only, (2) primary, secondary and tertiary roads only, and (3) all roads. Classification crosswalks for each categorical layer are provided in Online Appendix 2. We recognize that vector features such as land cover and roads may be represented continuously by using moving windows to summarize each feature within a window. However, our PathSF analysis summarizes data within a weighted kernel around each used and available area, making an initial smoothing or weighting of the surface an extra, unnecessary step.

All raster layers were available at a 30 m spatial grain size and we rasterized all vector layers to a 30 m grain size. To examine a suite of spatial grains, we upscaled each 30 m layer to 120 m and 240 m using the majority rule for categorical layers and the focal mean for continuous layers. Each landscape definition was restricted to a single spatial grain.

Landscape definitions

Varying the data source, thematic resolution, and spatial grain provided multiple representations of each geospatial layer. For example, elevation had a single

data source with four different thematic resolutions and three different spatial grains, for a total of 12 different representations (Fig. 2). Likewise, percent slope, terrain ruggedness, and percent impervious surface also had 12 representations each. Percent vegetative cover had three data sources for the same combination of thematic resolutions and spatial grains, for a total of 36 representations. Roads had two data sources, three thematic resolutions and three spatial grains, for a total of 18 representations. Land cover type had three data sources, two thematic resolutions and three spatial grains, for a total 18 representations.

The ultimate landscape definition for puma could consist of a single geospatial layer or any combination of geospatial layers represented at any of the available spatial grains, thematic resolutions, and data sources, all of which are plausible and realistic alternatives for modeling puma movement. Given the vast number of layer representations and combinations, analyzing a full factorial of landscape definitions was not feasible. Therefore, we repeated a random selection procedure 2500 times. We assumed this would allow us to capture the general patterns of how landscape definition affects inference about puma movement. To generate a single landscape definition we:

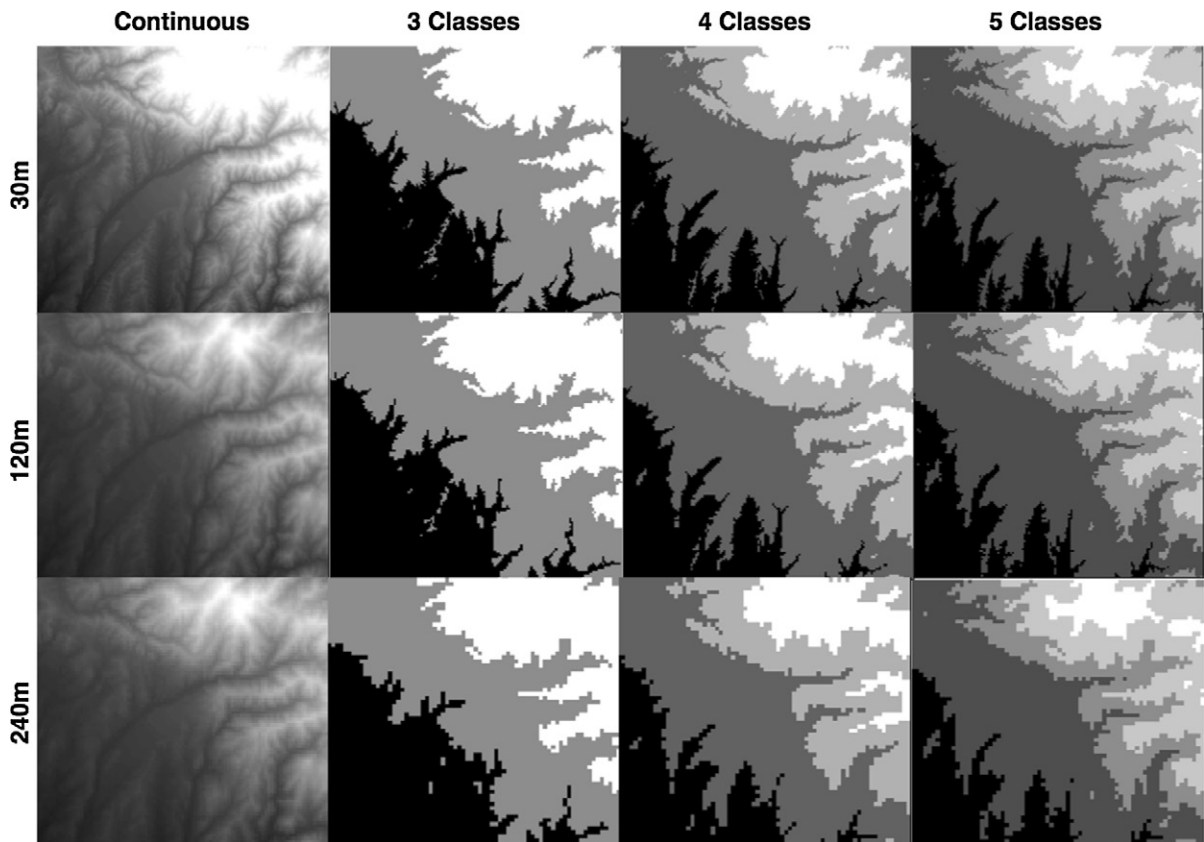


Fig. 2 Subset of the study area showing the 12 possible representations of elevation. Elevation was represented at four thematic resolutions (continuous; three classes, four classes, and

five classes) and each thematic resolution was represented at three different spatial grains (30, 120, 240 m)

- (1) randomly selected the spatial grain (30, 120, 240 m) for all layers included in the landscape definition;
- (2) randomly selected the number of layers to include (one to seven);
- (3) randomly selected, without replacement, which layer(s) to include (elevation, percent slope, terrain ruggedness, land cover type, percent vegetative cover, roads, percent impervious surface);
- (4) randomly selected the data source for each layer, as appropriate; and
- (5) randomly selected the thematic resolution of each layer.

We repeated this process until 2500 unique landscape definitions were generated.

We also derived a model to compare how we would define the landscape given previous analyses and knowledge of this population (hereafter referred to as

the expert model). The expert model consisted of CalVeg Land Cover represented with eight classes, OpenStreetMap Roads represented with three classes, and the following continuous layers: elevation, percent impervious surface, percent slope, and terrain ruggedness (all at a 30 m grain size).

Statistical analyses

We conducted all statistical analyses in the R software environment (R Core Team 2013). We used multi-scale PathSFs as described in Zeller et al. (2015) to model landscape use and connectivity for pumas in our study area. PathSF's are analyzed in the paired 'used'/'available' framework typical of conditional resource selection functions where the proportion of used to available for a landscape feature indicates preference or avoidance of that feature (e.g., Compton et al. 2002; Cushman and Lewis 2010). For each of our layer

representations, we calculated the used data within a 30 m fixed-width buffer around each daily path. Available data were calculated using a Pareto-weighted kernel around each daily path (Zeller et al. 2015). The Pareto distribution was chosen as fits the empirical distribution of puma movement distances well (Zeller et al. 2014). If a layer had a categorical representation, we calculated the proportion of each category within the used or available area. If a layer had a continuous representation, we calculated the mean. Therefore, all predictor variables in the statistical models were continuous in nature. If a geospatial layer was continuous, it was included in the model as a single variable. If a geospatial layer was categorical, the number of categories equaled the number of variables it contributed to the model (since each category was treated as a separate variable). This is worth noting since we later take a closer look at the number of layers used in each landscape definition, which varied from 1 to 7 (elevation, percent slope, terrain ruggedness, percent vegetative cover, percent impervious surface, land cover type, and roads) versus the number of variables in each landscape definition, which varied from 1 to 24.

Pumas in our study area select different landscape features at different scales (Zeller et al. 2014); therefore, we developed multi-scale PathSF models using a pseudo-optimized two-step approach (Zeller et al. 2014, 2015; McGarigal et al. 2016). The first step is to run univariate models to identify the best scale and the second step is to combine the best scales in a multiple regression model. To represent different scales, we varied the radii of the Pareto kernel at 10 different spatial extents from 500 to 7500 m (Zeller et al. 2014). To identify the characteristic scale of selection we identified the scale with the lowest corrected AIC value (AICc; Burnham and Anderson 2002; Online Appendix 3). This scale was then used in building the multivariate conditional logistic regression models for each of our 2501 landscape definitions.

For a single landscape definition, we took each layer that comprised that definition and calculated used and available data (at the appropriate scale) for the puma paths. If there were multiple layers for a definition and correlations ≥ 0.7 were found between layers, we dropped the variable with the lowest AIC weight from each pair of correlated layers. We then ran a conditional logistic regression model. Occasionally models produced complete separation warnings or

convergence errors. When this was encountered we dropped the model from the analysis, a new unique landscape definition was generated, and the model was re-run. This was repeated for all landscape definitions, resulting in 2500 fitted models. For the expert model, we created a global model with all uncorrelated layers at their selected representations. Because we thought all the variables would have some influence on puma habitat use, we then used the ‘dredge’ function in the *MuMIn* package (Barton 2016) to create all possible combinations of model variables. We used AICc to identify the top models and used model averaging to average top models within 4 AICc units of each other.

We calculated AICc for each model to compare overall model performance and to select the top models. We calculated percent deviance explained (D2) to compare the strength of the overall fit of the models (Franklin 2009), and we used the concordance correlation coefficient (CCC; Lin 1989) to evaluate the calibration of the predictive models. For a well-calibrated model, the predicted observations should fall close to the expected observations on a line originating at 0 with a slope of 1 (Johnson et al. 2006). The CCC statistic measures how correlated two points are based on their deviance from this 45-degree line, with CCC values closer to 1 indicating better-calibrated models.

We determined the sensitivity of model selection, model fit, and prediction calibration to landscape definition by modeling the corresponding performance criteria (AICc, D2, and CCC) as a function of the following landscape definition options: (1) spatial grain (30, 120, 240 m), (2) number of layers used in the model (1–7), (3) number of variables (1–24), and (4) whether the variables were all represented continuously, all represented categorically, or whether a mix of continuous or categorical representations were present. Specifically, we conducted Likelihood Ratio Tests comparing the full model with each of these four definition options left out in turn. Because roads and land cover class could only be represented categorically, we omitted landscape definitions with these variables to assess thematic resolution. We wanted to compare only models where a layer could be represented both categorically and continuously. We also produced mean and standard error plots to assess the relative influence of each of these four definition options on each of the performance criteria.

Using AICc values, we identified the top landscape definitions for pumas in our study area. We calculated odds ratios for the top model variables by predicting the probability of movement for each variable at the 25th and 75th percentile of the variable distribution while keeping the other variables in the model at their means and taking the ratio of the 75th percentile predicted probability to that of the 25th percentile.

To assess the importance of layer representations and determine whether some layer representations influenced model performance more than others; we identified paired models with and without each layer representation. We subtracted the AICc of the model with the layer representation from that of the model without the layer representation and took the mean of this difference across model pairs. The greater the mean value, the more important that layer representation is for modeling pumas in our study area. We also used this layer importance metric to determine whether some layer sources influenced model performance more than others. We found it uninformative to use AIC weights to assess variable importance due to only having three models with any AIC weight.

To determine the sensitivity of probability of movement values (and thus inferred habitat selection) obtained from the different landscape definitions, we randomly sampled 1000 pixels throughout the study area and predicted the probability of movement at each pixel for each of our 2501 models. We then used various data exploration metrics (standard deviation, coefficient of variation, range, interquartile range) to determine the sensitivity of predicted values to landscape definition.

Modeling connectivity and road crossing locations

To provide a cursory example of how landscape definition may affect connectivity and corridor modeling, we selected landscape definitions across the model performance continuum at the 0th, 25th, 50th, 75th and 100th percentile of AICc values plus the expert model. We predicted the relative probability of movement from each of these six models across our study area in the paired prediction framework as described in Zeller et al. (2015). This approach requires ‘used’ and ‘available’ to be calculated for every pixel in the study area. We first calculated the proportion of each variable within a 30 m fixed-width buffer at each pixel in our study area. We then

calculated the proportion of each variable at each pixel weighted by the Pareto kernel at the appropriate scale for that variable. This is akin to the ‘used’ and ‘available’ in the paired regression framework and allows for a unique relative probability of movement to be identified for each pixel in the study area. We assumed that this relative probability of movement could be used as a proxy for landscape conductance (McRae et al. 2008). For example, a pixel with a high probability of movement would have high conductance and vice versa. We visually examined the probability of movement/conductance surfaces to highlight differences.

We modeled connectivity across the Temecula corridor region, which is a subset of our study area. This region has been identified as the last viable, though highly threatened, link between the Santa Ana Mountain and the Palomar Mountain puma populations (Ernest et al. 2014; Vickers et al. 2015). An eight-lane highway, Interstate 15 (I-15), bisects the two mountain ranges and recent conservation attention has been paid to identifying locations for road crossing structures along its length. Therefore, as an example of how landscape definition might affect conservation agendas, we sought to identify road-crossing locations for each of the six selected models. First, we used CircuitScape (McRae et al. 2013) to create current density maps (McRae et al. 2008) across each conductance surface between nationally protected lands on either side of I-15. We then identified the top 10 pixels along I-15 with the most current flow, which might be considered preferred locations for constructing road-crossing structures. In this context, ‘current flow’ represents the number of random walkers that would move through a pixel as they passed between protected areas. We noted the location of the road-crossing pixels for each of our six landscape definitions as well as differences in these locations among landscape definitions.

Results

Delta AICc, D2, and CCC of the 2501 conditional logistic regression models varied widely with landscape definition (Fig. 3). The top model had only two competing models within 4 AICc units and four others within 10 AICc units; all other models were greater than 10 delta AICc units from the top model. Thus,

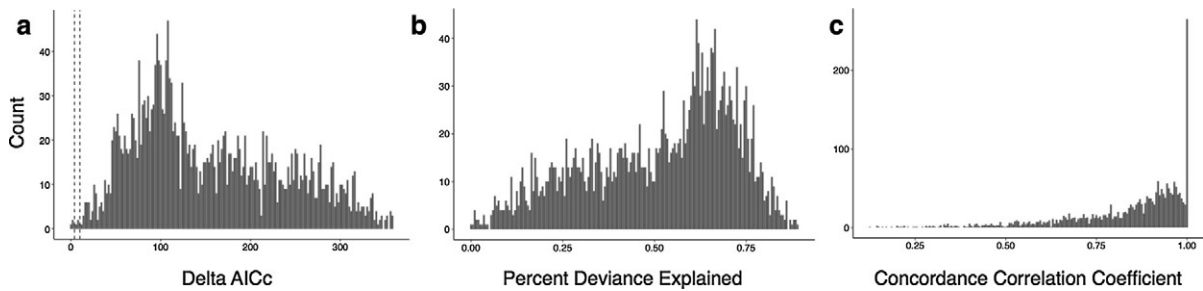


Fig. 3 Histograms of **a** delta AICc, **b** percent deviance explained (D2), and **c** concordance correlation coefficient (CCC) across our 2501 landscape definitions associated with

modeling puma movement in southern California. The *left* and *right* vertical dashed lines in **a** represent a delta AICc of 4 and 10, respectively

only seven (<1%) of the models were at all competitive out of the 2501 evaluated, and the great majority of models were vastly inferior, with delta AICc values greater than 100. The best expert model was not among the four top models (it was ranked as the 31st model) and differed from the top model by 20.75 AICc units. D2 followed similar patterns, and even though the vast majority of models explained an ecologically meaningful percent of the deviance, the range in absolute explanatory power among models extended from 0.002 to 0.892. CCC was much less sensitive to landscape definition than AICc and D2, with most models having well-calibrated predictions.

We had expected to observe many more top models from the 2501 landscape definitions and were concerned about spurious model results. Therefore, we bootstrapped the data 1000 times, by drawing from the data set with replacement, for each landscape definition to produce a mean bootstrapped AICc value with a confidence interval for each model. These bootstrap results, presented in Online Appendix 4, confirm the top seven models.

Likelihood ratio tests indicated that spatial grain and number of layers significantly influenced the three model performance measures (Table 2). Spatial grain was the most significant, with finer spatial grain resulting in better model performance values (Figs. 4a, 5a, Online Appendices 5a–6a). This was followed by the number of layers, with a greater number of layers resulting in better values (Figs. 4c, 5b, Online Appendices 5c–6c). The number of variables in a landscape definition was significant for D2, but not AICc or CCC, with a greater number of variables resulting in better values (Fig. 4d; Online Appendices 5d–6d). Model form was significant for AICc and D2 where a mix of continuous and

categorical layers increased model performance (Figs. 4b, 5c, Online Appendices 5b–6b).

The top three models (delta AICc < 4) their associated landscape definitions and variable odds ratios are provided in Table 3 (all seven models with a delta AICc < 10 are provided in Online Appendix 7). The top three models had D2 > 0.86 and CCC > 0.95, and thus had exceptionally strong explanatory power and well-calibrated predictions. These models included 4–5 geospatial layers and a mix of continuous and categorical layer representations for a total of 5–10 variables per model. All three of these models were at a spatial grain of 30 m and included percent slope defined continuously. All three of these models also included roads defined with two classes (primary and secondary roads); however, the source of the roads data varied among models. Elevation, terrain ruggedness, percent impervious surface and percent vegetative cover were included in various forms in some but not all of these models. Only land cover type was not included in any of the top models. The exact interpretation of these models requires coupling the characteristic scale of each variable (Online Appendix 3) with the corresponding regression coefficient or, odds ratio (Table 3). Briefly, across these top three models, pumas strongly avoided areas with steep slopes evaluated over intermediate scales (1500 m). Pumas showed weak preference for lower elevations evaluated over intermediate scales (1500–2000 m). Pumas generally avoided less rugged terrain at intermediate scales (1500 m) and strongly avoided more rugged terrain at very coarse scales (7500 m). Pumas showed very weak preference for areas with moderate vegetative cover evaluated over intermediate scales (3500 m), but an avoidance of areas with increasing vegetative cover over coarse scales (7500 m). Pumas

Table 2 Likelihood ratio test results

	AICc			D2			CCC		
	<i>Log-likelihood</i>	χ^2	<i>p value</i>	<i>Log-likelihood</i>	χ^2	<i>p value</i>	<i>Log-likelihood</i>	χ^2	<i>p value</i>
<i>Spatial grain</i> (<i>df</i> = 2)	−2377			576			912		
	−2610	465	<2.2e−16	344	464	<2.2e−16	903	19	7.5E−05
<i>Number of layers</i> (<i>df</i> = 6)	−2377			576			912		
	−2387	19	0.0042	568	17	0.0093	904	16	0.0138
<i>Variable form</i> (<i>df</i> = 2)	−2377			576			912		
	−2380	6	0.0493	573	7	0.0302	910	5	0.0821
<i>Number of variables</i> (<i>df</i> = 23)	−2377			576			912		
	−2393	31	0.1228	547	57	0.0001	910	5	0.9999

AICc, D2, and CCC for the 2500 landscape definitions were modeled as a function of four landscape definition options (1) spatial grain, (2) number of layers used in the model, (3) whether the variables were all represented continuously, categorically, or whether a mix of forms were present, and (4) number of variables. Likelihood ratio tests were performed comparing the full model with each of these definition options left out in turn

strongly selected areas with the lowest percentage of impervious surfaces (0–12%) evaluated over coarse scales (7500 m). Lastly, pumas showed strong avoidance of primary and secondary roads evaluated over coarse scales (7500 m).

Our layer importance results indicated that all geospatial layer representations improved model performance (positive values, Mean difference in AICc, Table 4). We expected the layer representations in our top three models to also have high importance as judged by our criterion, but this expectation was not consistently supported. For example, the most important layer representation (county roads represented with four classes) was not present in any of our top three models. In addition, one layer representation that was present in our top three models had one of the lowest variable importance values (terrain ruggedness represented categorically with three classes). We believe these results are due to the fact that our top-ranked models were relative outliers in the distribution of our 2501 model definitions, and that these layer importance results more closely reflect the importance of layer representations in the center of the model distribution. These results indicated percent impervious surface and roads provided the most

marked improvement in model performance (>50; Table 4). However, variable importance results from the other layers were inconsistent (Table 4).

Similarly, we expected some geospatial data sources to improve AICc values more than others. However, we only found a consistent improvement for data source with the land cover data sources. The CalVeg land cover layers consistently outperformed the other land cover data sources (Table 4).

Probability of movement/conductance values at the 1000 randomly selected pixels varied widely across the 2501 landscape definitions (Fig. 6). The interquartile range at a pixel ranged from 0.2 to 1, with most of the pixels varying between 0.5 and 0.65. This indicates that probability of movement/conductance values are highly sensitive to the landscape definition used in the model.

Probability of movement/conductance surfaces for the top model, the expert model, and the 25th, 50th, 75th, and 100th percentile of AICc values also varied spatially (Fig. 7). Probability of movement/conductance was relatively evenly distributed across the study area in the top model. The expert model showed stark contrasts between areas with high and low conductance, and the 25th percentile model resulted in relatively higher conductance values across the study

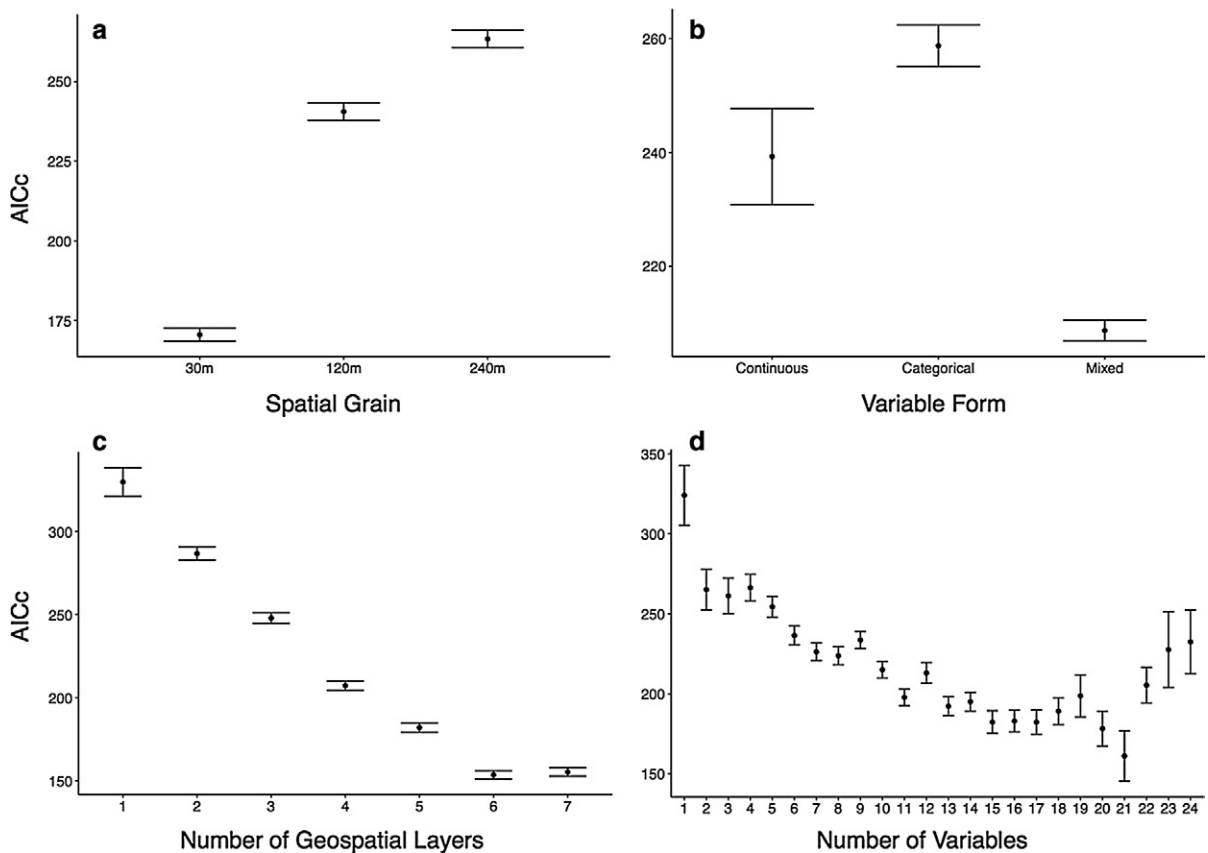


Fig. 4 Mean and standard error in model AICc as a function of spatial grain, variable form, number of geospatial layers, and number of variables in a landscape definition associated with modeling puma movement in southern California

area. The 75th percentile surface was primarily driven by major roads, while most of the 100th percentile surface showed consistently low conductance throughout the study area except in a few locations. The model results of these six example landscape definitions are included in Online Appendix 8.

CircuitScape current density surfaces and the locations of the top 10 pixels along I-15 that had the most current, or greatest inferred flow of individuals, differed markedly according to landscape definition (Fig. 8). To determine if the crossing locations were biologically less probable between the top model and the other five models, we calculated the percent change in current from the 10 crossing locations of the top model and the 10 crossing locations of the other models, measured on the current surface of the top model. These values indicated a percent decrease in current of 71, 88, 87, 95, and 88% for the expert

model, and 25th, 50th, 75th, and 100th percentile models, respectively.

Discussion

As landscape ecologists we are keenly aware of the importance of scale in wildlife-habitat inference and the preeminent importance of landscape definition in any landscape ecological analysis. As such, we conducted this study with the full expectation that puma habitat selection during within-home range movements in southern California would be somewhat sensitive to landscape definition. What our study revealed in this regard, however, was quite startling—namely that inferred habitat selection, probability of movement/conductance surfaces, and resultant connectivity models were all exceptionally sensitive to

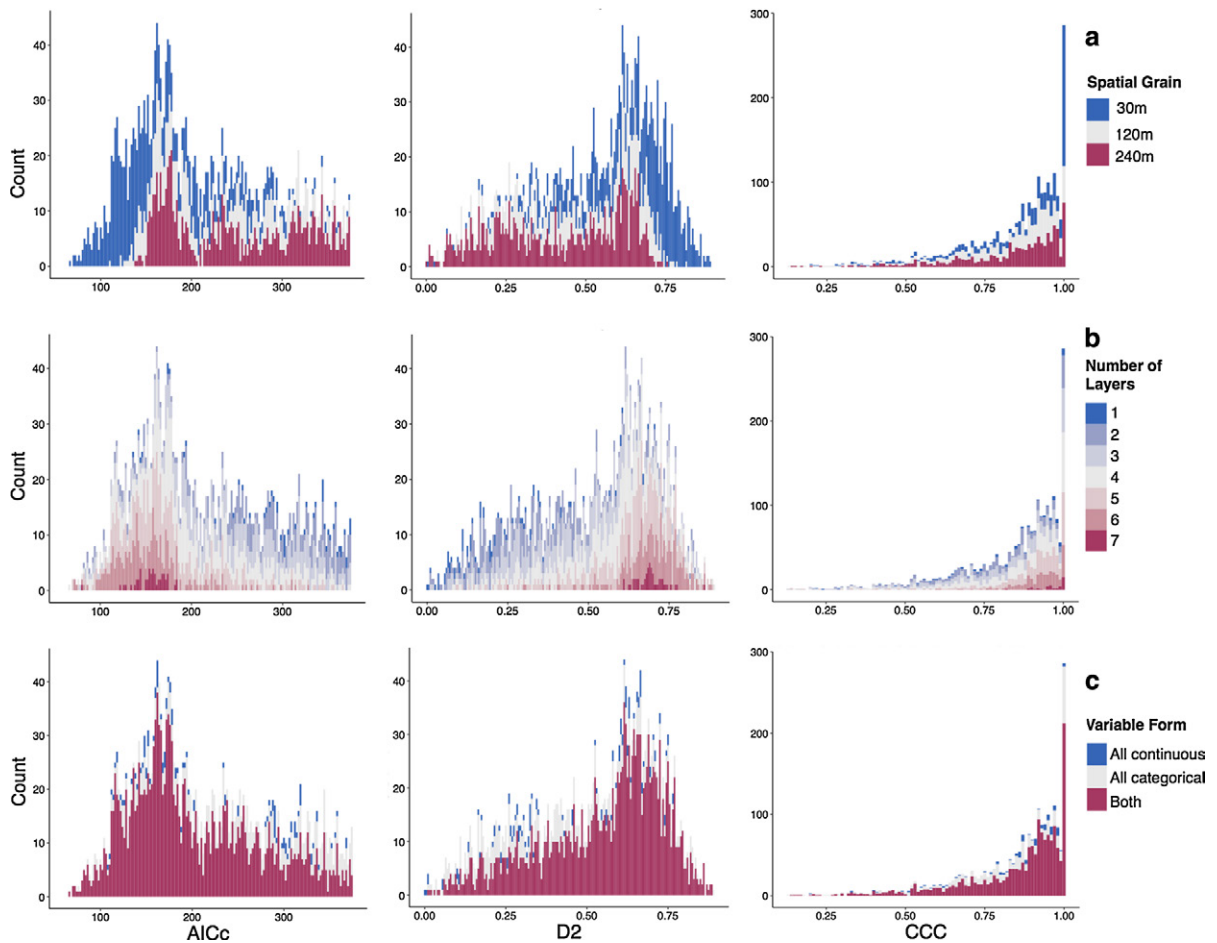


Fig. 5 AICc, percent deviance explained (D2), and concordance correlation coefficient (CCC) values for path selection function models derived with 2501 different landscape definitions associated with modeling puma movement in southern California. Histograms are color-coded according to **a** the

spatial grain of the landscape definition (30, 120, 240 m), **b** the number of geospatial layers included in a landscape definition (1–7), and **c** the form of the variables in the landscape definition (whether they are represented continuously, categorically, or both). (Color figure online)

landscape definition. All 2501 of the alternative landscape definitions were plausible and realistic given our current understanding of pumas in southern California, yet, the weight of empirical evidence (based on AICc) was overwhelmingly in support of only a few of the alternative models. Moreover, the absolute explanatory power (based on percent deviance explained, D2) of the alternative models varied widely. Despite the dramatic differences among alternative models in their relative and absolute explanatory power, the vast majority of the alternative models produced predictive surfaces very well calibrated to the landscape (based on the concordance correlation coefficient, CCC). Overall, these results suggest that while there may be many ways to define

the landscape and produce well-calibrated predictive surfaces, there may be very few clearly superior ways to define the landscape. Indeed, there are vastly more ways to define the landscape relatively poorly than there are to “get it right” even though all definitions seem plausible a priori.

These findings extend to probability of movement/conductance surfaces and connectivity modeling. In our case study of pumas, the relative predicted probability of movement/conductance values, for most pixels, ranged nearly from 0 to 1 across alternative landscape definitions, indicating that different landscape definitions may result in polar opposite conclusions regarding probability of movement/conductance for the same location. This can

Table 3 Top three landscape definitions, as indicated by AICc values ($\Delta \text{AICc} < 4$), for modeling puma movement in southern California and the associated geospatial layers, thematic resolutions, thematic class (number of variables in the model) and associated odds ratio

Layer	Thematic resolution	Class	Model 1 odds ratio	Model 2 odds ratio	Model 3 odds ratio
Percent slope	Continuous	–	0.004	0.07	0.0005
Elevation	Continuous	–		0.99	0.0869
Elevation	four classes	1	1.16		
		3	1.0		
		4	0.99		
Terrain ruggedness	Three classes	1	0.64		0.727
		2	0.99		0.990
		3	0.0007		0.023
Landfire percent vegetative cover	Four classes	1			0.99
		2			1.0
		3			0.186
		4			–
Percent impervious surface	Four classes	1		17.16	
OSM roads	Two classes	1	0.0005	0.098	
		2	0.0011	0.077	
County roads	Two classes	1			0.002
		2			0.17

Odds ratios express the relative odds of preference or avoidance; values greater than one indicate a preference, values less than one indicate an avoidance. All geospatial layers were at a 30 m spatial grain. OSM = Open Street Map. Note, some landscape definitions do not contain the full set of classes for a layer because the absent variables were highly correlated (>0.7) with other variables in the same model

have implications for connectivity modeling that is based on surface conductance, or its inverse, resistance. Indeed, the predicted road crossing locations from various landscape definitions in our connectivity modeling exercise were strikingly different. Our findings align with those of Cushman et al. (2014) who predicted road-crossing locations for black bears with two different resistance surfaces (one derived from genetic data and one derived from path data). Despite the fact that both resistance surfaces included the same variables and were very highly correlated, they produced very different road crossing locations. Consequently, recommended locations for constructing wildlife road crossing structures strongly depend on both data type and the chosen landscape definition.

We conducted this study with the expectation that some geospatial layers and digital representations of them (e.g., at certain spatial grains and thematic resolutions based on one particular data source) would outperform others. Specifically, we hypothesized that model performance would increase with (1) decreasing

spatial grain, (2) increasing thematic resolution, (3) increasing number of geospatial layers, and (4) that some data sources would improve model performance more than others. Consistent with our first hypothesis, our puma model performance was most sensitive to the spatial grain of the landscape definition, with finer spatial grains resulting in better model performance. These findings agree with other studies that have examined the effect of spatial grain on performance of species-habitat models (Karl et al. 2000; Seoane et al. 2004; Venier et al. 2004; Guisan et al. 2007; Cushman and Landguth 2010a; Gottschalk et al. 2011). Though Tobalske (2002) and Seoane et al. (2004) found an increase in predictive performance of species distribution models with decreasing spatial grain, landscape definitions at our coarsest grain size (240 m) generally resulted in poorer performing models, which is noteworthy because this spatial grain is very close to many freely available data platforms such as MODIS and CORINE (both 250 m). Overall, our results indicate that finer-grained geospatial data are superior for

Table 4 Relative importance of geospatial layer representations across spatial grains in generalized linear models predicting puma movement in southern California

Layer	Thematic resolution	Mean difference in AICc
<i>County roads</i>	4 Classes	176
<i>Percent impervious surface</i>	3 Classes	164
<i>Percent impervious surface</i>	Continuous	157
<i>Percent impervious surface</i>	4 Classes ^a	143
<i>CalVeg land cover type</i>	5 Classes	142
<i>Percent impervious surface</i>	5 Classes	140
<i>CalVeg land cover type</i>	8 Classes	111
<i>NLCD land cover type</i>	5 Classes	95
<i>OSM roads</i>	2 Classes ^a	76
<i>County roads</i>	3 Classes	70
<i>LandFire land cover type</i>	5 Classes	69
<i>OSM roads</i>	4 Classes	67
<i>OSM roads</i>	3 Classes	67
<i>NLCD percent vegetative cover</i>	Continuous	63
<i>County roads</i>	2 Classes ^a	59
<i>LandFire percent vegetative cover</i>	4 Classes ^a	55
<i>Landfire percent vegetative cover</i>	Continuous	47
<i>Percent slope</i>	4 Classes	46
<i>Percent slope</i>	3 Classes	46
<i>LandFire land cover type</i>	8 Classes	46
<i>LandFire percent vegetative cover</i>	3 Classes	37
<i>Percent slope</i>	5 Classes	36
<i>Landsat percent vegetative cover</i>	5 Classes	36
<i>LandFire percent vegetative cover</i>	5 Classes	31
<i>Terrain ruggedness</i>	4 Classes	29
<i>Elevation</i>	5 Classes	29
<i>NLCD percent vegetative cover</i>	3 Classes	28
<i>NLCD percent vegetative cover</i>	5 Classes	25
<i>Percent slope</i>	Continuous ^a	25
<i>Elevation</i>	Continuous ^a	23
<i>Landsat percent vegetative cover</i>	4 Classes	22
<i>NLCD percent vegetative cover</i>	4 Classes	18
<i>Elevation</i>	4 Classes ^a	16
<i>Elevation</i>	3 Classes	15
<i>Terrain ruggedness</i>	5 Classes	15
<i>Terrain ruggedness</i>	3 Classes ^a	11
<i>Landsat percent vegetative cover</i>	3 Classes	10
<i>Terrain ruggedness</i>	Continuous	3

Importance was determined by finding paired models, one model with the layer representation of interest and one model that was the same, but without the layer representation of interest. The AICc of the model with the layer of interest was subtracted from the AICc of the model without the layer of interest. If the layer improved the model, this difference would be positive. The mean difference across paired models for a layer representation was calculated. The greater the value, the more that layer improved upon model performance

^a Indicates a layer representation that was present in one of the top three models

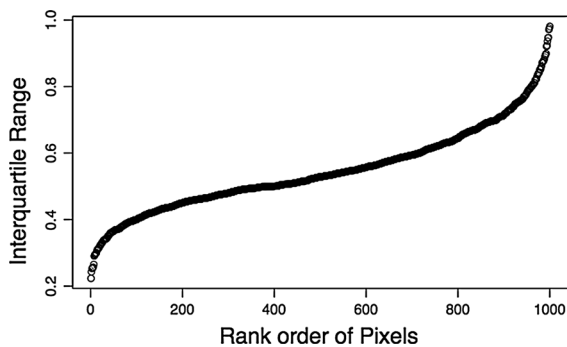


Fig. 6 Empirical distribution plot of the interquartile range of probability of movement (or conductance) values across the 2501 models of puma movement in southern California for 1000 randomly selected pixels across the study area

modeling puma movement in southern California, and we encourage others to investigate whether this finding is generalizable to other contexts.

Support for our second hypothesis regarding thematic resolution was equivocal. The best performing models tended to have a mix of layers defined continuously and categorically. Based on our previous work we expected models with all continuous layers to outperform categorical-only and mixed definition models (McGarigal and Cushman 2005; Cushman and Landguth 2010a; Cushman et al. 2010a, b). Moreover, we generally expected finer thematic resolutions to have greater model performance than coarser resolutions of that same layer, as in Seoane et al. (2004) and Cushman and Landguth (2010a). Perhaps our results differed from the previous studies because pumas in our study area are responding to more broadly defined landscape patterns, which are sometimes better reflected as categorical layers, than more finely detailed landscape structure. Another possibility is that pumas are responding to some

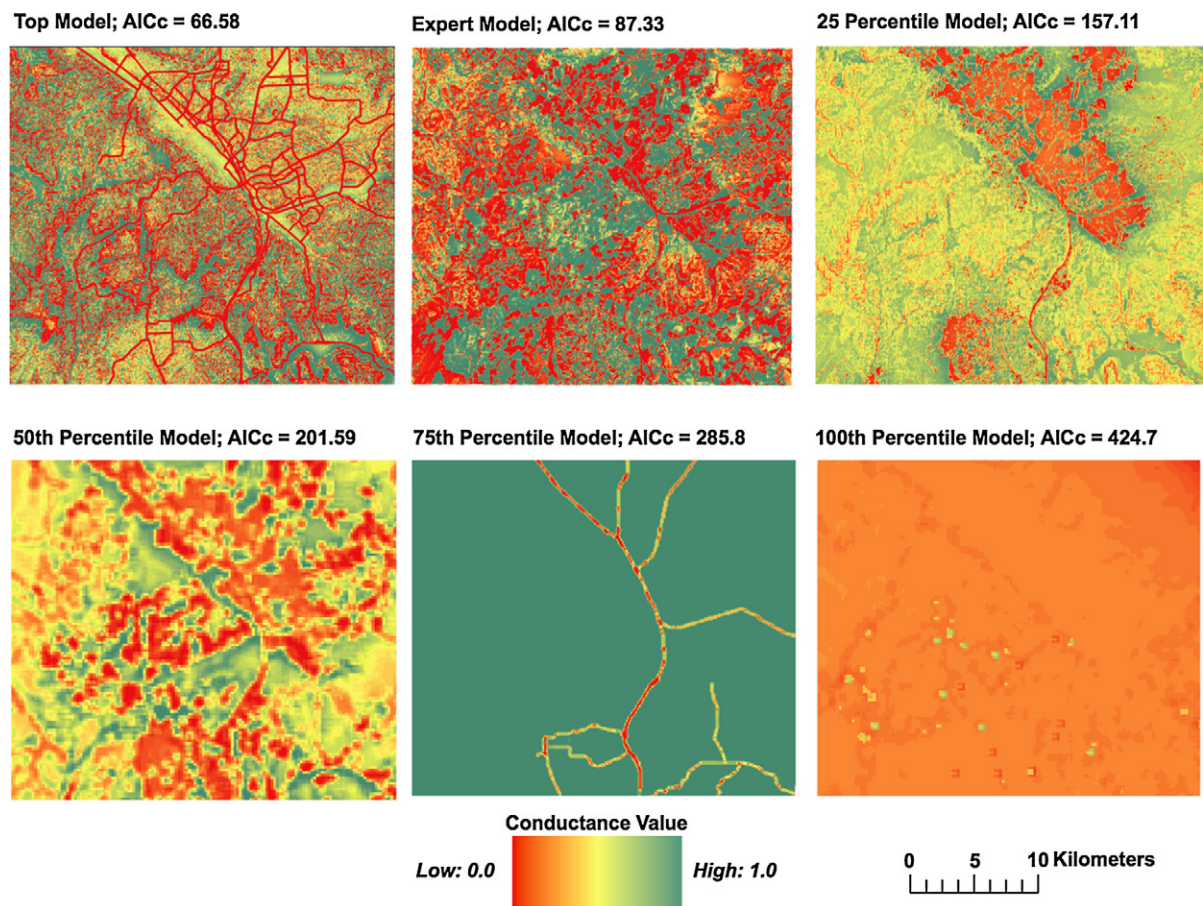


Fig. 7 Probability of puma movement/conductance surfaces for a subset of the study area in southern California for the top-ranked model, the expert model, and models with the 25th, 50th, 75th, and 100th percentile of AICc values. (Color figure online)

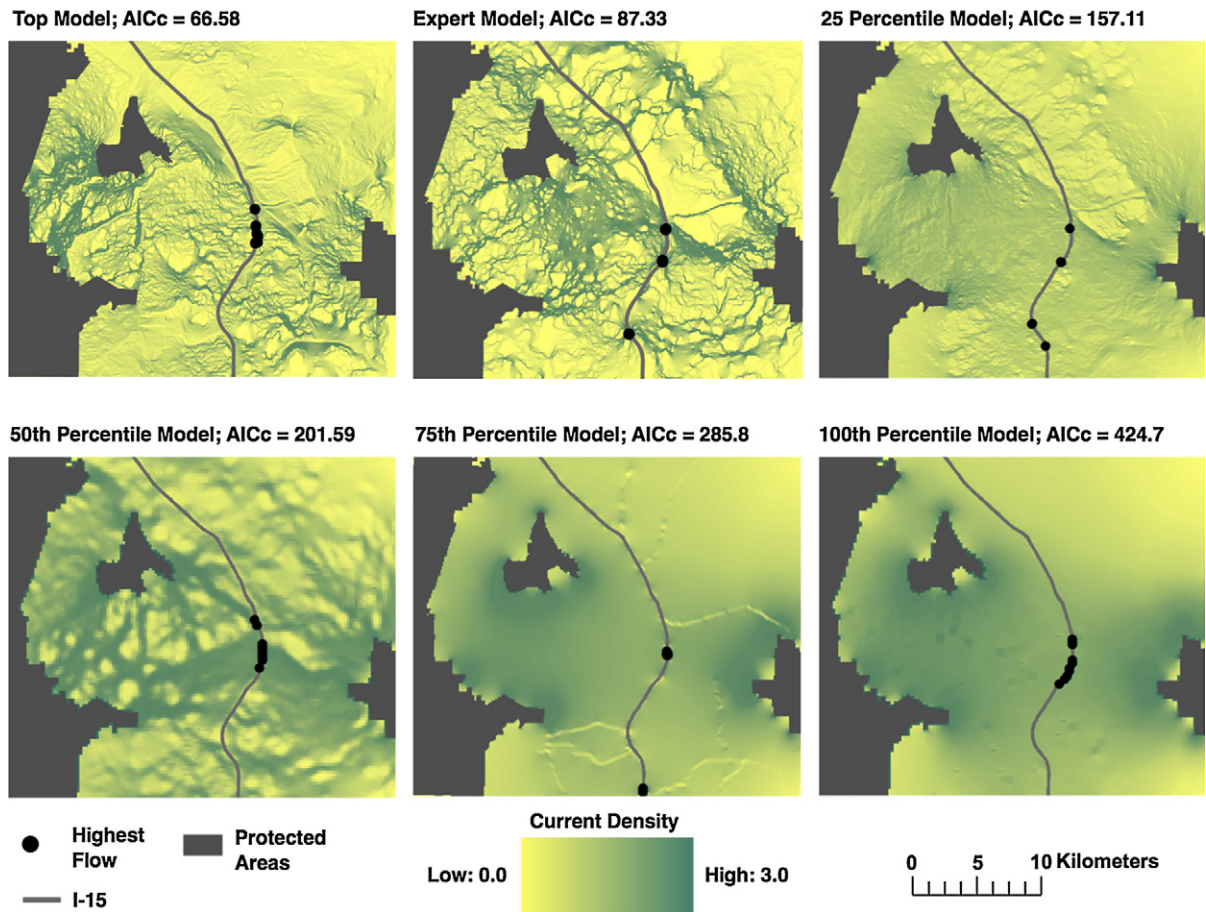


Fig. 8 CircuitScape current density surfaces associated with modeling puma movement in southern California and the point locations along I-15 with the 10 highest current densities. (Color figure online)

variables in a non-linear, step-wise form and the categorical nature of some variables better captured this relationship than did continuous variables. The issue of thematic resolution was first addressed in MacArthur et al. (1966), who found that bird species in one location appeared to respond to a higher number of vegetation classes than bird species in another location. Given the many decades that have passed since this paper was published it is surprising that more research has not been conducted on how thematic resolution of geospatial layers affects species-habitat models. The equivocal results of thematic resolution on model performance found here and by Lawler et al. (2004) indicate that this is an area ripe for further research.

In support of our third hypothesis, our puma model performance was very sensitive to the number of geospatial layers in the landscape definition, with

increasing number of layers resulting in better model performance. However, aside from model improvement with the inclusion of roads and percent impervious surface variables, there was no clarity in our variable importance results in terms of which layers and associated layer representations were better than others. Although the results likely reflected layer importance for the bulk of the model distribution, they did not adequately capture the layers in the top models. These results further highlight that our top models were relative outliers in our suite of landscape definitions.

With regards to our last hypothesis, we only found one data source that outperformed others in terms of model performance (the CalVeg data source for land cover type outperformed NLCD and LandFire). This is in contrast to previous studies that found consistent model improvement with data source (Chust et al.

2004; Seoane et al. 2004). However, these studies confounded data source with spatial grain, indicating that spatial grain may have been more influential than the actual sources of data. Though not wildlife-specific, Cushman et al. (2010a) evaluated the ability of different land cover maps to predict the distribution of plant species. In their analysis, they did not confound grain and data source and found large differences in the explanatory power of the different data sources. For pumas in our study area, other landscape definition options were more influential than data source, though weighing the pros and cons of different data sources is surely an important consideration when selecting geospatial layers.

Our empirical evaluation of alternative landscape definitions for modeling puma movement in southern California allowed us to identify clearly superior landscape definitions among the pool of viable candidates, which led to stronger inference about puma habitat selection during movement in the study area than our expert a priori landscape definition. We learned from our top-ranked models that slope was better represented as a continuous variable and that pumas showed a strong avoidance of steep slopes, as has been previously documented (Beier 1995; Dickson et al. 2005; Dickson and Beier 2007; Wilmers et al. 2013). All top-ranked models also included at least one of the other two topographic variables, elevation (represented either continuously or categorically) and terrain ruggedness (better represented categorically), indicating that pumas are strongly influenced by topography, as has been previously documented (Alexander et al. 2006; Burdett et al. 2010; Wilmers et al. 2013; Allen et al. 2014). Our top-ranked models also indicated a strong avoidance of primary and secondary roads by pumas, echoing previous research (Dickson et al. 2005; Wilmers et al. 2013). One of our top-ranked models also included percent impervious surface represented categorically and indicated that pumas strongly selected for the lowest class of imperviousness (0–12%; Burdett et al. 2010).

One of our most noteworthy findings with regards to puma habitat selection was the apparent lack of strong selection for vegetation composition and structure in the top models, which we included in our expert model and was the sole basis for defining the landscape in our previous modeling work in this system (Zeller et al. 2014, 2015). Only one of our top-ranked models included any type of vegetative characteristic, with

pumas avoiding areas with very high vegetative cover at coarse scales. Land cover type (represented as the relative abundance of individual land cover types) was not in any of the top-ranked models. Overall, our results suggest pumas in our study area respond more strongly to topographic variables and human development in the form of roads and other impervious surfaces than to other landscape characteristics related to the composition and structure of vegetation. This finding aligns somewhat with recent findings by Gray et al. (2016), who showed that distance from roads (as a proxy for human development) could be used to accurately model puma occurrence and landscape permeability. The similarities to our findings may be due to the fact that both our study areas had relatively high levels of human development. However, these results are also similar to other studies on large felids. Elliot et al. (2014) showed vegetation was much weaker than roads, towns, and agricultural lands in predicting lion movement, and Krishnamurthy et al. (2016) showed agricultural areas and villages were more important for predicting tiger movement than was natural vegetation. Our findings do not indicate that pumas do not select for vegetation composition and structure during within-home range movement, but rather that, comparatively, selection is much stronger for terrain and human development than for vegetation. It is also possible that vegetation cover attributes were not selected in the top-ranked models because topographic variables served as a proxy for vegetation cover types, as has been observed in previous studies (Burrough et al. 2001; Beier and Brost 2010). To examine this hypothesis further, we conducted a variance partitioning using the *ecospat* package in R (Broennimann et al. 2015) to portion the explained variance in the top model (with land cover type added as a predictor) between the terrain and human development predictors and land cover type predictors. Terrain and human development independently accounted for 8.7% of the explained variance, land cover independently accounted for 5.7% of the explained variance, and these predictors jointly accounted for 86% of the explained variance. Thus, the vast majority of the explained variance was confounded between the two sets of predictors and we are therefore unable to say whether terrain is acting as a proxy for vegetative characteristics or vice versa.

In summary, our findings may have implications for both research and conservation planning. First, we had

expected to find that our alternative landscape definitions would produce only slight differences in model performance in our case study on pumas in southern California. Instead, we found massive differences in model performance among the alternatives, with only a handful of competing landscape definitions among the 2501 models evaluated. If we were able to run the full factorial of landscape definitions, we may have found that there were indeed a greater number of competing models and perhaps an even better performing top model. However, like us, most researchers will only be able to compare a limited set of landscape definitions. Our results indicate that, at least for PathSFs, researchers may need to evaluate many different landscape definitions to find the optimal landscape representation for a study area and target species. Evaluating habitat and movement relationships with thematic resolution, thematic content, and grain that do not match the organisms' ecology and perceptions can greatly reduce model performance and the interpretations gained from models of landscape conductance.

This finding is relevant in consideration of Type I and Type II errors and the issue of affirming the consequent (Cushman and Landguth 2010b), in which a result that is consistent with a hypothesis is incorrectly accepted as demonstration that the hypothesis is true. Specifically, high sensitivity of model performance to variable grain and thematic resolution that we observed suggest elevated risk of Type II errors (failing to see an effect when it is present) when using variables at suboptimal definition. In addition, the high inherent correlation among variables increases the difficulty of distinguishing effects such that the risk of affirming spurious correlations and making Type I errors is elevated. The practice of dropping variables from correlated pairs reduces variance inflation and Type I error rate in the full model, but can result in affirming spurious correlations if the incorrect variable is dropped. The high dependence of variable influence on landscape definition compounds the challenge of resolving this.

Second, given that many resource selection functions are used to predict the relative probability of use across a study area to identify resource use areas for conservation purposes, the wide differences among predicted values at the same pixels across landscape definitions is very disconcerting and indicates different landscape definitions result in large differences in

predicted quality of locations for movement. In our case study on pumas we were specifically modeling probability of within-home range movement, which may be more sensitive to landscape definition than modeling probability of use, but further research is needed to determine this. Additionally, previous studies have found that dispersal of individuals is less constrained by landscape features than is home-range use (Elliot et al. 2014; Mateo Sánchez et al. 2014), which might suggest that connectivity estimates derived from dispersal data would be less sensitive to landscape definition than within home-range data.

Differences in the predicted probability of movement/conductance surfaces also translated into differences among modeled connectivity surfaces (derived using CircuitScape) and the optimal road crossing locations, again in agreement with Cushman et al. (2014). It was reassuring that our top model identified road crossing locations that have been approached by pumas in our study area and that were also identified by a consensus of road ecology and puma experts (Vickers et al. unpublished report). However, the alternative landscape definitions were highly variable in identifying optimal road crossing locations, in some cases agreeing with the top model and in other cases indicating different locations. Our expert model also identified road crossing locations that have been approached by pumas in our study area, though not to the same extent as the top model indicating we may have missed some important crossing locations.

The way the landscape is represented is at the heart of all species-habitat models. Landscape definition will ultimately affect inference about species-habitat relationships, probability of use surfaces, and connectivity estimates. Therefore, defining the landscape to the best of our ability is of utmost importance. To our knowledge, this is the first study to assess model performance across all of the following four landscape definition choices: (1) spatial grain, (2) thematic resolution, (3) number of geospatial layers, and (4) source of geospatial layers. However, we would like to note that our analysis, results, and conclusions are based on a single data set, which is a limitation of this study. More research is needed to determine the effect of landscape definition with other data sets, study areas and taxa as well as with other species-habitat models such as point and step selection functions, species distribution models, and occupancy models. Research is also needed to more effectively tease apart the

effects of thematic resolution and layer source on species-habitat models. Without further studies, generalizable conclusions and supporting landscape ecological theory, we are left with an important, but rather unsatisfactory recommendation for the time being: put a great deal of thought into the landscape definition of choice and if possible, and examine a very wide variety a landscape definitions.

Acknowledgements We thank B. Compton and E. Plunkett, for assistance with computational capacity, D. Dawn, D. Krucki, C. Bell, P. Bryant, D. Stewart, and K. Krause for field assistance and the two journal reviewers for thoughtful and insightful suggestions on improving this manuscript. We would also like to thank the following landowners/managers: The Nature Conservancy, California Department of Fish and Wildlife, Orange County Parks Department, The New Irvine Ranch Conservancy, Audubon Starr Ranch Reserve, Riverside County Parks Department, and the Cleveland National Forest. This material is based upon work supported by the Robert and Patricia Switzer Foundation, the American Association of University Women, P.E.O., UMASS Graduate School, The Nature Conservancy, Orange County Transportation Corridor Agency, The Nature Reserve of Orange County, and the McBeth Foundation.

References

- Addicott JF, Aho JM, Antolin MF, Padilla DK, Richardson JS, Soluk DA (1987) Ecological neighborhoods: scaling environmental patterns. *Oikos* 49:340–346
- Alexander SM, Logan TB, Paquet PC (2006) Spatio-temporal co-occurrence of cougars (*Felis concolor*), wolves (*Canis lupus*) and their prey during winter: a comparison of two analytical methods. *J Biogeogr* 33:2001–2012
- Allen ML, Elbroch LM, Casady DS, Wittmer HU (2014) Seasonal variation in the feeding ecology of pumas (*Puma concolor*) in northern California. *Can J Zool* 92:397–403
- Barton K (2016) Package MuMIn: multi-model inference. <https://cran.r-project.org/web/packages/MuMIn/index.html>. Accessed 30 April 2016
- Beier P (1995) Dispersal of juvenile cougars in fragmented habitat. *J Wildl Manag* 59:228–237
- Beier P, Brost B (2010) Use of land facets to plan for climate change: conserving the arenas, not the actors. *Conserv Biol* 24:701–710
- Broennimann O, Petitpierre B, Randin C, Engler R, Di Cola V, Breiner F, D'Amen M, Pellissier L, Pottier J, Pio D, Garcia Mateo R, Hordijk W, Dubuis A, Scherrer D, Salamin N, Guisan A (2015) Spatial ecology miscellaneous methods. R Package available from: <http://www.unil.ch/ecospat/home/menuinst/tools-data/tools.html>
- Burdett CL, Crooks KR, Theobald DM, Wilson KR, Boydston EE, Lyren LM, Fisher RN, Vickers TW, Morrison SA, Boyce WM (2010) Interfacing models of wildlife habitat and human development to predict the future distribution of puma habitat. *Ecosphere* 1:1–21
- Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. Springer, New York
- Burrough PA, Wilson JP, van Gaans PFM, Hansen AJ (2001) Fuzzy k-means classification of topo-climatic data as an aid to forest mapping in the Greater Yellowstone Area, USA. *Landscape Ecol* 16:523–546
- Chust G, Pretus JL, Ducrot D, Ventura D (2004) Scale dependency of insect assemblages in response to landscape pattern. *Landscape Ecol* 19:41–57
- Compton BW, Rhymer JM, McCollough M (2002) Habitat selection by wood turtles (*Clemmys insculpia*): an application of paired logistic regression. *Ecology* 83:833–843
- Cushman SA, Evans JS, McGarigal K, Kiesecker JM (2010a) Toward gleasonian landscape ecology: from communities to species, from patches to pixels. Research paper RMRS-RP-84. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO
- Cushman SA, Gutzweiler K, Evans JS, McGarigal K (2010b) The gradient paradigm: a conceptual and analytical framework of landscape ecology. In: Cushman SA, Huettmann F (eds) Spatial complexity, informatics, and wildlife conservation. Springer, New York, pp 83–110
- Cushman SA, Landguth EL (2010a) Scale dependent inference in landscape genetics. *Landscape Ecol* 25:967–979
- Cushman SA, Landguth EL (2010b) Spurious correlations and inference in landscape genetics. *Mol Ecol* 19:3592–3602
- Cushman SA, Lewis JS (2010) Movement behavior explains genetic differentiation in American black bears. *Landscape Ecol* 25:1613–1625
- Cushman SA, Lewis JS, Landguth EL (2014) Why did the bear cross the road? Comparing the performance of multiple resistance surfaces and connectivity modeling methods. *Diversity* 6:844–854
- Cushman SA, McGarigal K, McKelvey K, Vojta CD, Reagan CM (2013) Landscape analysis for habitat monitoring. In Rowland MM, Vojta CD (tech. eds) A Technical guide for monitoring wildlife habitat. Gen. Tech. Rep. WO-89. U.S. Department of Agriculture, Forest Service, Washington, DC, pp 6-1–6-18
- Dickson BG, Beier P (2007) Quantifying the influence of topographic position on cougar (*Puma concolor*) movement in southern California, USA. *J Zool* 271:270–277
- Dickson BG, Jenness JS, Beier P (2005) Influence of vegetation, topography, and roads on cougar movement in southern California. *J Wildl Manag* 69:264–276
- Elliot NB, Cushman SA, Macdonald DW, Loveridge AJ (2014) The devil is in the dispersers: predictions of landscape connectivity change with demography. *J Appl Ecol* 51:1169–1178
- Ernest HB, Vickers TW, Morrison SA, Buchalski MR, Boyce WM (2014) Fractured genetic connectivity threatens a southern California puma (*Puma concolor*) population. *PLoS ONE* 9(10):e107985. doi:10.1371/journal.pone.0107985
- European Environmental Agency (2010) CORINE (Coordination of information on the environment). The European Topic Centre on Land Use and Spatial Information. <http://>

- www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster#tab-gis-data
- Forman RTT (1995) Land mosaics: the ecology of landscapes and regions. Cambridge University Press, Cambridge
- Franklin J (2009) Mapping species distributions: spatial inference and prediction. Cambridge University Press, Cambridge, pp 1–340
- Gottschalk TK, Aue B, Hotes S, Ekschmitt K (2011) Influence of grain size on species-habitat models. *Ecol Model* 18:3403–3412
- Gray M, Wilmers CC, Reed SE, Merenlender AM (2016) Landscape feature-based permeability models relate to puma occurrence. *Landsc Urban Plan* 147:50–58
- Guisan A, Graham CH, Elith J, Huettmann F, NCEAS Species Distribution Modelling Group (2007) Sensitivity of predictive species distribution models to change in grain size. *Divers Distrib* 13:332–340
- Holland JD, Bert DG, Fahrig L (2004) Determining the spatial scale of species' response to habitat. *BioScience* 54:227–233
- Holmes BR, Laundré JW (2006) Use of open, edge and forest areas by pumas (*Puma concolor*) in winter: are pumas foraging optimally? *Wildl Biol* 12:201–209
- Jenks GF (1967) The data model concept in statistical mapping. *Int Yearb Cartogr* 7:186–190
- Jenness J (2013) DEM surface tools. Jenness Enterprises. http://www.jennessent.com/arcgis/surface_area.htm
- 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 Sens Environ* 132:159–175. Downloaded from the national map viewer: <http://viewer.nationalmap.gov/viewer/>. Accessed 24 August 2014
- Johnson CJ, Nielsen SE, Merrill EH, McDonald TL, Boyce MS (2006) Resource selection functions based on use-availability data: theoretic motivation and evaluation methods. *J Wildl Manag* 70:347–357
- Karl JW, Heglund PJ, Garton EO, Scott JM, Wright NM, Hutto RL (2000) Sensitivity of species habitat-relationship model performance to factors of scale. *Ecol Appl* 10:1690–1705
- Kissling WD, Fernandez N, Paruelo JM (2009) Spatial risk assessment of livestock exposure to pumas in Patagonia, Argentina. *Ecography* 32:807–817
- Krishnamurthy R, Cushman SA, Sarkar MS, Malviya M, Naveen M, Johnson JA, Sen S (2016) Multi-scale prediction of landscape resistance for tiger dispersal in central India. *Landscape Ecol* 31:1355–1368
- LandFire (2012a) Existing vegetation cover. U.S. Department of the Interior, Geological Survey. <http://www.landfire.gov/viewer/>. Accessed 21 August 2014
- LandFire (2012b) Existing vegetation type. U.S. Department of the Interior, Geological Survey. <http://www.landfire.gov/viewer/>. Accessed 21 August 2014
- Lawler JJ, O'Connor RJ, Hunsaker CT, Jones KB, Loveland TR, White D (2004) The effects of habitat resolution on models of avian diversity and distributions: a comparison of two land-cover classifications. *Landscape Ecol* 19:515–530
- Lewis JS, Rachlow JL, Garton EO, Vierling LA (2007) Effects of habitat on GPS collar performance: using data screening to reduce location error. *J Appl Ecol* 44:663–671
- Lin LI (1989) A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1:255–268
- MacArthur R, Recher H, Cody M (1966) On the relation between habitat selection and species diversity. *Am Nat* 100:319–332
- Mateo Sánchez MC, Cushman SA, Saura S (2014) Scale dependence in habitat selection: the case of the endangered brown bear (*Ursus arctos*) in the Cantabrian Range (NW Spain). *Int J Geogr Inf Sci* 28:1531–1546
- McGarigal K, Cushman SA (2005) The gradient concept of landscape structure. In: Wiens J, Moss M (eds) Issues and perspectives in landscape ecology. Cambridge University Press, Cambridge, pp 112–119
- McGarigal K, Wan HY, Zeller KA, Timm BC, Cushman SA (2016) Multi-scale habitat modeling: a review and outlook. *Landscape Ecol* 31:1161–1175
- McRae BH, Dickson BG, Keitt TH, Shah VB (2008) Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology* 89:2712–2724
- McRae BH, Shah VB, Mohapatra TK (2013) Circuitscape 4 user guide. The Nature Conservancy. <http://www.circuitscape.org>. Accessed 19 May 2015
- OpenStreetMap contributors (2014) OpenStreetMap California. <http://download.geofabrik.de/north-america/us/california.html>. Accessed 16 September 2014
- Orange County Transportation Authority (2011) Master plan of arterial highways. Orange County Transportation Authority, Orange, California. <http://www.octa.net/Plans-and-Programs/GIS-Data/GIS-Data-Download/>. Accessed 18 September 2014
- R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Riverside County Geographic Information Systems (2013) County static data. Riverside County Innovation Center, Riverside CA. <http://gis.rivcoit.org/GISData.aspx>. Accessed 24 August 2014
- San Bernadino (2014) San Bernadino county centerlines. http://maps.sanbag.ca.gov/website/Transportation_data.htm. Accessed 18 September 2014
- SanGIS (2014) All roads shape file. SanGIS/SANDAG Data Warehouse. San Diego Geographic Information Source—JPA/San Diego Association of Governments (SANDAG). http://www.sangis.org/Download_GIS_Data.htm. Accessed 20 July 2014
- Seoane J, Bustamente J, Diaz-Delgado R (2004) Are existing vegetation maps adequate to predict bird distributions? *Ecol Model* 175:137–149
- Sexton JO, Song X-P, Feng M, Noojipady P, Anand A, Huang C, Kim D-H, Collins KM, Channan S, DiMiceli C, Townshend JRG (2013) Global, 30-m resolution continuous fields of tree cover: landsat-based rescaling of MODIS Vegetation Continuous Fields with lidar-based estimates of error. *Int J Digit Earth*. doi:10.1080/17538947.2013.786146
- Thompson CM, McGarigal K (2002) The influence of research scale on bald eagle habitat selection along the lower Hudson River, New York (USA). *Landscape Ecol* 17:569–586
- Tobalske C (2002) Effects of spatial scale on the predictive ability of habitat models for the green woodpecker in

- Switzerland. In: Scott JM, Heglund PJ, Samson F, Haufler J, Morrison M, Raphael M, Wall B (eds) Predicting species occurrences: issues of accuracy and scale. Island Press, Covelo, pp 197–204
- USDA Forest Service (2014) CalVeg: FSSDE. Evge-Tile47A_02_v2. Pacific Southwest Region Remote Sensing Lab, McClellan
- U.S. Geological Survey (2009) National elevation dataset. EROS Sioux Falls, SD. <http://viewer.nationalmap.gov/viewer/>. Accessed 20 August 2014
- Venier LA, Pearce J, McKee JE, McKenney DW, Niemi GJ (2004) Climate and satellite-derived land cover for predicting breeding bird distribution in the Great Lakes Basin. *J Biogeogr* 31:315–331
- Vickers TW, Sanchez JN, Johnson CK, Morrison SA, Botta R, Smith T, Cohen BS, Huber PR, Ernest HB, Boyce WM (2015) Survival and mortality of pumas (*Puma concolor*) in a fragmented, urbanizing landscape. *PLoS ONE*. doi:[10.1371/journal.pone.0131490](https://doi.org/10.1371/journal.pone.0131490)
- Wiens JA (1989) Spatial scaling in ecology. *Funct Ecol* 3:385–397
- Wilmer CC, Wang Y, Nickel B, Houghtaling P, Shakeri Y, Allen ML, Kermish-Wells J, Yovovich V, Williams T (2013) Scale dependent behavioral responses to human development by a large predator, the puma. *PLoS ONE*. doi:[10.1371/journal.pone.0060590](https://doi.org/10.1371/journal.pone.0060590)
- Wu J (2004) Effects of changing scale on landscape pattern analysis: scaling relations. *Landscape Ecol* 19:125–138
- Zeller KA, McGarigal K, Beier P, Cushman SA, Vickers TW, Boyce WM (2014) Sensitivity of landscape resistance estimates based on point selection functions to scale and behavioral state: pumas as a case study. *Landscape Ecol* 29:541–557
- Zeller KA, McGarigal K, Beier P, Cushman SA, Vickers TW, Boyce WM (2015) Using step and path selection functions for estimating resistance to movement: pumas as a case study. *Landscape Ecol* 31:1319–1335