

# Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics

MELANIE A. MURPHY,<sup>1,4</sup> JEFFREY S. EVANS,<sup>2</sup> AND ANDREW STORFER<sup>3</sup>

<sup>1</sup>*School of Biology, Colorado State University, Fort Collins, Colorado 80524 USA*

<sup>2</sup>*The Nature Conservancy, Fort Collins, Colorado 80524 USA*

<sup>3</sup>*School of Biological Sciences, Washington State University, Pullman, Washington 99164 USA*

**Abstract.** A major objective of ecology is to understand how ecological processes limit population connectivity and species' distributions. By spatially quantifying ecological components driving functional connectivity, we can understand why some locally suitable habitats are unoccupied, resulting in observed discontinuities in distribution. However, estimating connectivity may be difficult due to population stochasticity and violations of assumptions of parametric statistics. To address these issues, we present a novel application of Random Forests to landscape genetic data. We address the effects of three key ecological components on *Bufo boreas* connectivity in Yellowstone National Park: ecological process, scale, and hierarchical organization. Habitat permeability, topographic morphology, and temperature–moisture regime are all significant ecological processes associated with *B. boreas* connectivity. Connectivity was influenced by growing-season precipitation, 1988 Yellowstone fires, cover, temperature, impervious surfaces (roads and development), and topographic complexity (56% variation explained). We found that habitat permeability generally operates on fine scales, while topographic morphology and temperature–moisture regime operate across multiple scales, thus demonstrating the importance of cross-scale analysis for ecological interpretation. In a hierarchical analysis, we were able to explain more variation within genetic clusters as identified using Structure (a Bayesian algorithm) (74%; dispersal cover, growing-season precipitation, impervious surfaces) as opposed to between genetic clusters (45%; ridgelines, hot, dry slopes, length of hot season, and annual precipitation). Finally, the analytical methods we developed are powerful and can be applied to any species or system with appropriate landscape and genetic data.

**Key words:** *algorithmic models; boreal toad; Bufo boreas; ecoinformatics; ecological process; landscape ecology; landscape genetics; multiple scales; Random Forests; species connectivity; Yellowstone National Park, USA.*

## INTRODUCTION

A major objective of ecology is to understand how ecological processes that underlie dispersal affect functional connectivity. Species' ranges are generally limited by biotic and abiotic factors (Morin et al. 2007). However, within a species range, underlying ecological processes influence functional connectivity among occupied habitat patches (Pulliam 2000, Ovaskainen and Hanski 2004). Quantifying connectivity among these habitats can elucidate why some locally suitable habitats are unoccupied, resulting in discontinuities of species distribution (Hanski and Gaggiotti 2004).

Amphibians are highly suitable focal organisms for studies of connectivity as they generally have patchy distributions (Wagner and Fortin 2005) that may result in limited connectivity (Smith and Green 2005). Yet reliable estimates of amphibian connectivity based on demographics may require decades of data due to

population stochasticity (Pechmann et al. 1991, Green 2003). Alternatively, a landscape genetics approach can be used to assess functional connectivity among occupied habitats with genetic distances using neutral markers (Keyghobadi et al. 2005, Storfer et al. 2007), an effective approach in amphibian systems (see, for example, Beebe 2005). However, ecological data used to assess connectivity are often spatially autocorrelated with nonlinear relationships between dependent and independent variables thus violating assumptions of standard parametric models (Wagner and Fortin 2005). In addition, these data and corresponding measures of genetic distance may have a low statistical signal-to-noise ratio, resulting in models with little predictive power.

Application of alternative analytical approaches to landscape genetic questions is needed to address these issues. One possible solution is the implementation of ecoinformatic approaches (Chon and Park 2006) such as classification and regression trees. Random Forests (Breiman 2001) is a type of classification and regression tree that is insensitive to autocorrelation, makes no

Manuscript received 10 May 2008; revised 20 January 2009; accepted 8 April 2009. Corresponding Editor: A. Sih.

<sup>4</sup> E-mail: melanie.murphy@colostate.edu

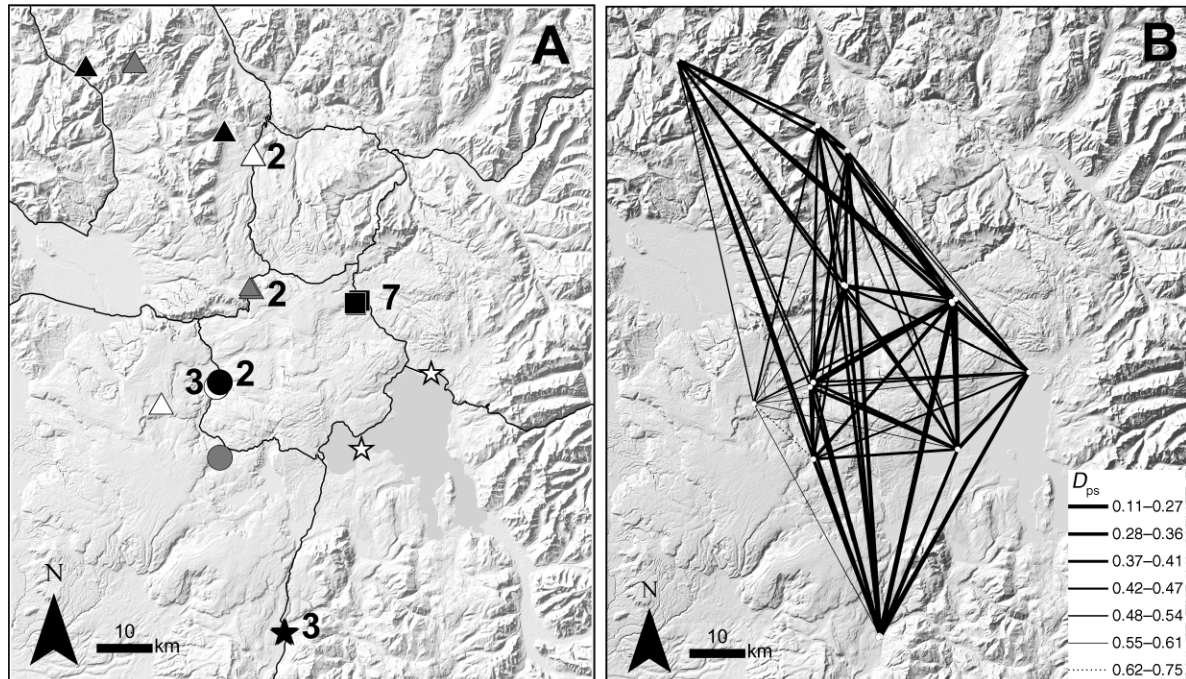


FIG. 1. (A) Study area (Yellowstone National Park, USA) with sample locations and genetic groups as identified using Structure (Pritchard et al. 2000). We identified four levels of genetic structure (circles, stars, triangles, and squares) with substructure in three of those groups (circles, stars, triangles) displayed by symbol shade (white, gray, or black). Numbers indicate areas where more than one site was present (see enlargement of these areas in Appendix D). (B) Pairwise connections between sites overlaid on a shaded relief. Line width is scaled by the genetic distance measure  $D_{ps}$ . The figure can be seen in color in Appendix D.

distributional assumptions, has the ability to incorporate complex interactions among independent variables, and has a high degree of power (De'ath and Fabricius 2000, Cutler et al. 2007). Random Forests (RF) grows many classification or regression trees based on a bootstrap of the data. For each tree, variation in the response variable is partitioned based on the predictor variables. Random Forests then combines the trees into a single prediction and provides measures of variable importance. While RF is a very powerful tool, as currently implemented ecological interpretation may be incomplete as it does not provide a framework for variable selection or testing overall model significance (Cutler et al. 2007).

To address these issues, we develop tools to increase ecological interpretability of RF and apply them to a novel landscape genetics study of *Bufo boreas* (boreal toads) in Yellowstone National Park. We address three major questions. First, what ecological process(es) limit *B. boreas* connectivity? Habitat permeability, topographic morphology, and temperature–moisture regimes are all ecological processes (Turner 1989) that may limit *B. boreas* connectivity. *Bufo boreas* is known to be sensitive to habitat permeability as measured by canopy cover (Bartelt et al. 2004), roads (Arens et al. 2007), and thermal influence (Koch and Peterson 1995). In addition, topographic morphology is often a barrier to dispersal (Funk et al. 1999), and temperature–moisture

regimes likely limit connectivity due to species' lack of physiological control over water loss (Palo et al. 2003, Bartelt and Peterson 2005). Second, on what scale(s) do these processes limit connectivity? Ecological data in landscape genetic studies are generally measured at a single scale. However, it is unlikely that the landscape surrounding areas between sites does not influence connectivity. Third, is connectivity hierarchical based on genetic clusters? Fine-scale processes with localized effects (for example, roads and dominant vegetation) likely influence connectivity within groups of more related individuals while broad-scale processes (major ridgelines and global trends in temperature and moisture) result in genetic discontinuities.

## METHODS

### Field data collection

We stratified the National Wetlands Inventory data for Yellowstone National Park (USFWS 2002) by elevation (two classes) and precipitation (three classes) with at least 20 wetlands randomly sampled per stratum. To address low site occupancy (3%), we visited known *B. boreas* breeding sites in addition to the stratification to ensure adequate sampling (D. Patla, *unpublished manuscript*). Geographic site locations (Fig. 1) were recorded with a Garmin 12XL GPS (Garmin, Olathe, Kansas, USA). Tissue samples were collected by buccal swabs from adults (Goldberg et al. 2003) and tail clips

TABLE 1. Ecological processes and respective metrics.

Metric by process	Abbreviation	Source	Metric explanation
Isolation by distance (IBD)			
Topographic distance (m)	distance	SRTM	topographically corrected distance
Habitat			
Water	water	NLCD	class: open water
Forest	forest	NLCD	classes: deciduous, evergreen, mix
Wetlands	wetland	NLCD	classes: emergent, herbaceous
Meadow	meadow	NLCD	classes: 71, 81, 82
Percentage of canopy	canopy	NLCD	NLCD canopy percentage product
Impervious surfaces	imperv	NLCD	classification of impervious surface
Thermal†	therm	USGS	percentage of thermal cells for a window around each cell in vector
1988 fires	burn	USGS	1988 fire perimeter
Topographic morphology			
Elevation	elev	SRTM	elevation in meters
Topographic roughness‡	rough	SRTM	variance in elevation
Elevation relief ratio‡	err	SRTM	elevation complexity
Relative slope position	rsp	SRTM	position between valley (0) and ridge (1)
Hierarchical slope position§	hsp	SRTM	measure of topographic exposure
Ridges	ridge	SRTM	ridgelines derived from rsp (1)
Temperature–moisture regime			
Heat load index	hli	SRTM	cool/hot values (slope-aspect transformation)
Compound topographic index	cti	SRTM	flow accumulation by catchment size (wetness)
Slope temperature–moisture	ssina	SRTM	indicates temperature–moisture
Curvature	crv	SRTM	convex or concave topography
Solar insolation	inso	SRTM	measure of solar intercept
Growing-season precipitation	gsp	CM	mean growing-season precipitation
Mean annual precipitation	map	CM	mean annual precipitation
Frost-free period	ffp	CM	date of last freeze to date of first freeze
Degree-days >5°C	dd5	CM	sum of degrees for all days >5°C

Notes: “Process” is the process of which the metric is a measure (isolation by distance, habitat permeability, topographic morphology, or temperature–moisture regime). Sources of data that contained the variable or from which the variable was derived are provided. Sources are all 30-m resolution and are as follows: NLCD, National Land Cover Database (2001); USGS, coverages developed for Yellowstone National Park research unit; SRTM, Shuttle Radar Topography Mission digital elevation model (DEM); and CM, climate model. Metric explanations are descriptions of the variables. The last two columns provide information on calculation and the ecological justification, along with references in cases of a previously published metric. Metrics are calculated along vectors for each pairwise comparison. Definitions for newly derived metrics (rsp, hsp) are included in Appendix B. Some metrics were calculated at multiple window sizes.

† Percentage of landscape (PLAND) thermal at a circular window size of 0, 15, and 27 cell (30-m cell size) radii.

‡ Calculated at  $3 \times 3$ ,  $15 \times 15$ , and  $27 \times 27$  30-m cell window sizes.

§ Decomposed from 0 to  $810 \times 810$  m window size (see Appendix B).

from tadpoles (2–5 mm) for DNA analysis following sterile field protocol.

#### Genetic data collection

DNA was extracted using the Qiagen DNeasy96 tissue protocol (QIAGEN, Valencia, California, USA). Fifteen microsatellite loci (Simandle et al. 2006) were genotyped with an ABI 3730 automated sequencer and scored with GeneMapper 3.7 (Applied Biosystems, Foster City, California, USA). (For loci, optimized conditions, quality control, and basic validation, see Appendix A.) To avoid potential allele frequency bias due to sampling tadpoles, we used a maximum likelihood algorithm (Colony) to identify full siblings (Wang 2004) and subsample sites at the size of the smallest sibling cluster (C. Goldberg, *personal commu-*

*nication*). We represented the pairwise genetic relationship between sample sites by connecting each site to every other site (Fig. 1B) in ArcInfo (ESRI 2000, Environmental Systems Research Institute, Redlands, California, USA; user-designed Arc Macro Language [AML] program). We estimated *B. boreas* functional connectivity (response variable) for each vector by  $F_{ST}$  (Wright 1951, Weir and Cockerham 1984),  $D_{ps}$  (Bowcock et al. 1994) as distance,  $D_c$  (Cavalli-Sforza and Edwards 1967), and Nei's genetic distance (Nei 1978).

#### Ecological process

To determine which ecological processes limit connectivity (habitat permeability, topographic morphology, and/or temperature–moisture regime), we calculated a set of metrics for each process along each vector



TABLE 1. Extended.

Calculation	Ecological justification
total vector length	isolation by distance has been documented in <i>Bufo boreas</i> (Manier and Arnold 2006)
count of water cells	open water is a likely dispersal barrier
count of forest cells	forest provides cover, which is important for dispersal (Bartelt et al. 2004)
count of wetland cells	proximal sites needed for sustaining populations (Stevens et al. 2007)
count of meadow cells	provides shrub cover, important for dispersal (Bartelt et al. 2004)
average along vector	mid-range cover provides dispersal habitat (Bartelt et al. 2004)
count of impervious cells	roads and development may limit dispersal ability (Mazerolle 2004, Arens et al. 2007)
mean PLAND (McGarigal and Marks 1995)	breeding sites are associated with thermal runoff (Koch and Peterson 1995)
count of burned cells	increased site occupancy has been observed post-fire (Hossack and Corn 2007)
mean elevation	elevation regulates breeding phenology, thus gene flow (Giordano et al. 2007)
mean rough (Riley et al. 1999)	high energy cost for dispersal in topographically complex areas
mean err (Evans 1972)	high energy cost for dispersal in topographically complex areas
mean rsp (Appendix B)	similar slope position may indicate similar breeding phenology
mean hsp (Appendix B)	exposed areas may impede dispersal (Bartelt et al. 2004)
count of ridge cells	ridges have been shown to limit amphibian gene flow (Funk et al. 1999)
mean hli (McCune and Keon 2002)	hot areas impede dispersal, little physiological control over water loss (Bartelt and Peterson 2005)
mean cti (Moore et al. 1993)	<i>Bufo boreas</i> may disperse through wet areas to prevent desiccation (Bartelt and Peterson 2005)
mean ssina (Stage 1976)	hot, dry slopes areas impede dispersal (Bartelt and Peterson 2005)
(Zeverbergen and Thorne 1987)	high curvature may retain moisture while low curvature may be exposed
mean inso (Fu and Rich 1999)	breeding-season inso relates energy needed for tadpole rearing
mean gsp (Rehfeldt 2006)	rainfall increases dispersal (Bartelt and Peterson 2005)
mean map (Rehfeldt 2006)	precipitation, including snowpack, may influence dispersal (Corn 2003)
mean ffp (Rehfeldt 2006)	short breeding season may result in less gene flow (Palo et al. 2003)
mean dd5 (Rehfeldt 2006)	high temperature increases desiccation risk (Bartelt and Peterson 2005)

connecting sites (see Table 1 for metric descriptions, references, and ecological basis). We assessed habitat permeability by amount of each cover class (forest, meadow, water, wetland), mean percent cover, amount of impervious surfaces (roads and development), and mean thermal influence along each vector (Table 1).

We measured mean topographic morphology between sites using a set of previously published metrics: elevation, topographic complexity (topographic roughness, elevation relief ratio), and ridges. In addition, we derived two new topographic metrics: relative slope position (relative position between valley bottom [0] and ridge top [1]) and hierarchical slope position, which measures topographic exposure (see Table 1, Appendix B).

We evaluated temperature–moisture with metrics derived from both a digital elevation model (DEM) and a spline-based climate model based on 30-year climate norms (Rehfeldt 2006, Rehfeldt et al. 2006) (Table 1). From the DEM, we calculated means of heat load index, compound topographic index, slope temperature–moisture, curvature, and solar insolation (Table 1). Using the Rehfeldt climate model, we calculated a mean for each vector of growing-season precipitation, mean annual precipitation, frost-free period, and degree-

days  $>5^{\circ}\text{C}$  (Table 1). Due to the pairwise nature of the data, we included topographic distance in all models to account for any distance effect.

Using these metrics as the predictor variables (Table 1), we built a model of connectivity for each ecological process in RF as implemented in R (Breiman 2001, Liaw and Wiener 2002, R Development Core Team 2007) using genetic distance as the response variable ( $F_{ST}$ ,  $D_{PS}$ ,  $D_C$ , or Nei's genetic distance). We executed RF as a regression with the following parameters: 5000 trees (number of bootstrap iterations), 34% data withheld for each tree (out of bag [OOB] sample), and  $m$  (number of independent metrics permuted at each tree node) optimized to the OOB error estimate (following Liaw and Wiener 2002).

To increase the ecological interpretability and predictive power of this algorithmic statistic, we developed methodology for removing redundant metrics, metric selection, evaluating model fit, and testing overall model significance. Random Forests is generally insensitive to collinearity among metrics. However, we found including metrics overlapping in multidimensional statistical space deflated variable importance and overall variation explained (data not shown). Therefore, we removed multivariate redundant variables using qr matrix de-

composition (threshold = 0.05; Becker et al. 1988). In the case of multivariate redundancy, we used a parsimony rule to retain the smallest number of nonredundant metrics.

Next we developed methodology for metric selection based on optimal model fit with the fewest number of metrics. While RF can operate with large numbers of variables, extensive metric lists make ecological interpretation difficult, introduce noise, and may decrease explanatory power. As an output from RF, metrics are ranked in order of importance ( $I$ ) based on the number of times a given metric decreased mean squared error (MSE). We ran an initial model with all metrics for a given ecological process and then calculated a model improvement ratio (MIR) for each metric. Unlike the raw  $I$  score, which can be influenced by the total number of metrics in the model, the MIR is comparable among models. It is calculated as  $[I_n/I_{\max}]$ , where  $I_n$  is the importance of a given metric and  $I_{\max}$  is the maximum model improvement score. We then iterated through MIR thresholds (0–1, in 0.1 increments), retaining all metrics above the given threshold. We then selected the threshold that minimized retained metrics, minimized model MSE, and maximized percentage of variation explained.

We assessed model fit by standard regression diagnostics: percentage of variation explained (RF pseudo- $R^2$ ), observed vs. predicted connectivity (median accuracy of predicted connectivity for OOB sample across all trees), residual distribution, and MSE. Next we wanted to test overall model significance. Although RF is a powerful statistic, it is possible that a given model is not significant from random chance. We created a null distribution by randomizing observed genetic distances (response variable) and calculating percentage of variation explained for each randomization ( $n = 1000$ ). We then calculated whether the percentage of variation explained for a given model was >95th percentile of the null distribution ( $P < 0.05$ ), a statistically significant model.

While all ecological processes may be significant, metrics from each process may not be retained in a model including all ecological processes. Therefore, we used the metrics selected in each significant ecological process model as the potential set of metrics for an overall model. We then repeated the above methodology for metric selection, model fit, and model significance.

#### Multiple scales

To assess the scale(s) at which ecological process(es) influence *B. boreas* connectivity, we iteratively increased the vector widths between sites and recalculated all metrics (60, 120, 240, 480, and 960 m diameter; ESRI 2000). For each scale, we built a model for each ecological process, tested overall model significance, and created a model with all significant ecological processes as in the previous RF methods. To create a single cross-scale model, we then combined the selected metrics from each scale in a single RF run and repeated

the metric selection, tested overall model significance, and evaluated model fit. When removing multivariate redundant metrics, we retained the finest scale if multiple scales of the same metric were identified using qr decomposition.

#### Hierarchical effect

We identified genetic clusters using Structure (Pritchard et al. 2000), a Bayesian algorithm that assigns individuals to the most likely genetic groups with no a priori definition of cluster number or their delineation (see Appendix A for conditions). To determine whether processes influencing connectivity are hierarchical based on genetic clusters, we partitioned the data into vectors within and between genetic clusters developing separate models for each. Within each hierarchical level, it is possible that not all ecological processes are significant. In addition, metrics selected at each scale may vary. Therefore, we tested the significance of each ecological process, built a connectivity model including all significant ecological process, and repeated the procedure at each scale (0, 60, 120, 240, 480, and 960 m) following the same RF methods. Using the metrics selected at each scale, we then ran RF to build a model of connectivity at each hierarchical level (within and between genetic clusters). We then performed metric selection, evaluated model fit, and tested overall model significance.

### RESULTS

#### Genetic results

We collected *B. boreas* samples at 26 breeding locations for a total sample size of 805, reduced to 685 after screening tadpoles for full siblings (Fig. 1; see Appendix A for basic population genetic tests and statistics). We were consistently able to explain more genetic variation with fewer metrics using the allele frequency distribution-based measure  $D_{ps}$  than  $F_{ST}$ ,  $D_C$ , or  $D_n$ . Therefore, *B. boreas* connectivity is estimated using  $D_{ps}$  unless otherwise noted (0.1396–0.6437). We list model results for all genetic distance measures in Appendix C.

#### Ecological process

We found a significant relationship between each of the ecological processes and connectivity ( $P < 0.05$ , 52–55% variation explained; Table 2). In comparison, topographically corrected distance alone explained only 21% of the variation in genetic distance ( $P < 0.05$ ; Table 2). Although we are unable to test significance among RF models, all tested models were significant. In addition, the temperature–moisture models of connectivity explained slightly more variation in genetic distance than habitat permeability or topographic complexity (55% vs. 52–54%; Table 2), suggesting the temperature–moisture process may disproportionately influence connectivity. However, when we combined all ecological processes into a single model of connectivity,

TABLE 2. Landscape genetic models of *Bufo boreas* connectivity in Yellowstone National Park.

Group by question	PVE	MSE	Model
Distance			
Distance	21.07	0.0169	distance
Process			
Habitat	51.80	0.0051	burn, distance, canopy, water, forest, meadow, imperv, wetland, therm15
Topographic morphology	53.57	0.0055	hsp, elev, rough15, err27, distance, rough27, err3, err15, rough3
Temperature–moisture regime	55.31	0.0052	gsp, distance, ssina
All	56.28	0.0050	gsp, distance, ssina, rough27/15, burn, canopy, hsp, err27, dd5, err15, forest, ffp
Scale			
Cross-scale	59.09	0.0047	ridge_960, gsp, ssina_480, canopy_60, ssina_240, ssina_960, imperv_240, err27_60
Hierarchical			
Within	74.03	0.0039	meadow_480, gsp_960, imperv_480, gsp_120, gsp_480, gsp
Between	44.85	0.0048	ssina_240, ssina_480, map, ssina_960, ridge_960, canopy_480, map_240, meadow_960, canopy_240, imperv_240, map_120, imperv_480, imperv_120, dd5_960

Notes: “Question” is the underlying question that a particular model addresses, while “group” defines the data used to build the model. “Distance” is genetic distance as  $D_{ps}$ . “PVE” is percentage of variation explained, a pseudo- $R^2$  derived from Random Forests regression. “MSE” is overall model mean squared error. “Model” is the final model with selected metrics in descending order of importance; metric abbreviations are given in Table 1. Values preceding an underscore or lacking an underscore (with the exception of dd5) indicate window size in number of 30-m cells for multiscale variables; numbers following an underscore indicate buffer size (scale). All models were significant at  $P < 0.05$ . Full model results and additional validation statistics are available in Appendix C.

metrics of all the tested ecological processes were retained: temperature–moisture (growing-season precipitation, slope temperature–moisture, degree-days  $>5^{\circ}\text{C}$  [dd5], and frost-free period), topographic morphology (topographic roughness, hierarchical slope position, and elevation relief ratio), and habitat permeability (burned, canopy, and forest; Table 2).

#### Multiple scales

Overall, we found that habitat permeability operated at fine scales, scale of topographic morphology was metric dependent, and temperature–moisture regime tended to operate across scales (Fig. 2). We observed distinct scale affects around 240 m (Fig. 2), possibly delineating localized (fine) vs. regional (broad) influenc-

es on connectivity. Two temperature–moisture metrics (dd5 and frost-free period), habitat permeability (burned, canopy), and topographic morphology metrics (roughness, elevation relief ratio, and hierarchical slope position) were retained for all scales  $\leq 240$  m (Fig. 2). Two temperature–moisture metrics (growing-season precipitation and slope temperature–moisture) were nonredundant and retained across scales (Fig. 2). Conversely, ridge lines, which may be regional barriers to gene flow, were only retained at scales  $\geq 240$  m (Fig. 2). We were able to explain the most variation in genetic distance by creating a single cross-scale model (59.09%; Table 2). Although we were unable to test significant differences among models, the cross-scale model had an overall improvement in model fit as assessed by MSE,

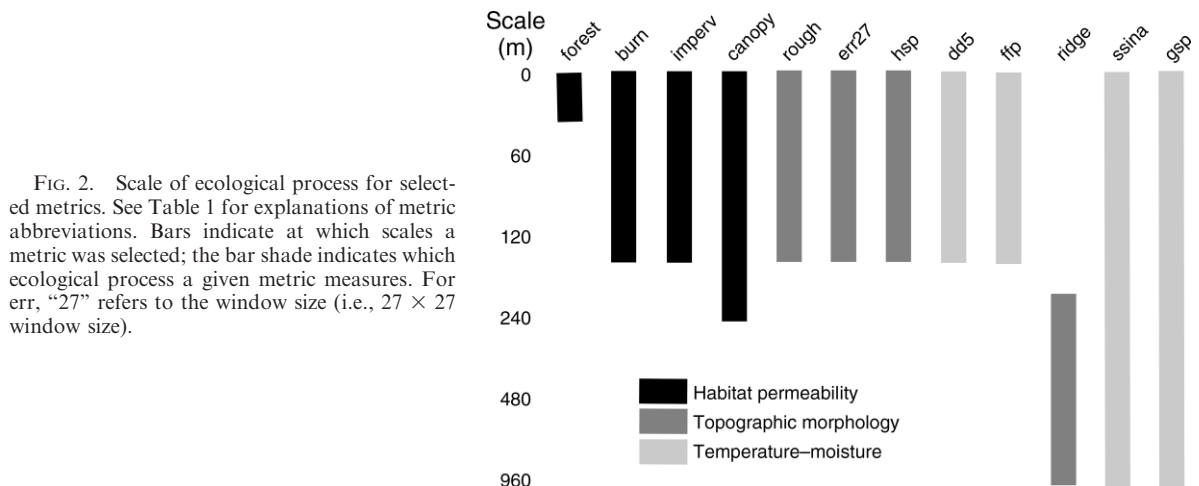


FIG. 2. Scale of ecological process for selected metrics. See Table 1 for explanations of metric abbreviations. Bars indicate at which scales a metric was selected; the bar shade indicates which ecological process a given metric measures. For err, “27” refers to the window size (i.e.,  $27 \times 27$  window size).

observed vs. predicted plots, and residual distribution (Appendix C). The cross-scale model included: topographic morphology (ridges at 960 m, elevation relief ratio at 60 m), temperature–moisture regime (growing-season precipitation and slope temperature–moisture), and habitat permeability (canopy at 60 m, impervious surfaces at 240 m; Table 2).

#### *Hierarchical effect*

We identified four genetic clusters using Structure, three of which had supported substructure (Fig. 1; Pritchard et al. 2000). As predicted, different ecological processes explain within-cluster variation compared to between-cluster variation (Table 2). In addition, at a given hierarchical level (within or between genetic clusters) distance was not retained as an important variable. Within clusters, we were able to explain 74.03% of the variation in  $D_{ps}$ , the most variation explained in any tested model, with little model error (Table 2). Meadows (which provide shrub habitat), growing-season precipitation (moisture during amphibian active season), and impervious surfaces (fragmentation effects of roads and development) were all selected metrics (Table 2). Between clusters, we were able to explain 45.32% of the variation in genetic distance (Table 2) with temperature (slope temperature–moisture and  $dd5$ ), ridges, and mean annual precipitation as selected metrics (Table 2).

#### DISCUSSION

Due to spatial autocorrelation, nonlinear relationships, the pairwise nature of the data, and overall signal-to-noise ratio, our analysis would be implausible using standard parametric approaches. Using a novel application of RF to landscape genetic data, we were able to elucidate effects of ecological processes, scale, and hierarchical effects on *B. boreas* connectivity. Habitat permeability, topographic morphology, and temperature–moisture regimes are all strongly correlated with *B. boreas* connectivity in Yellowstone, with a relatively small proportion of variation explained by distance alone. In addition, processes driving connectivity operated at multiple spatial scales. Habitat permeability tended to operate on fine scales, topographic morphology at broad scales, and temperature–moisture regime across spatial scales. We found hierarchical structure, with localized processes driving connectivity within genetic clusters while regional spatial and temporal processes drove connectivity between genetic clusters. Finally, this study puts the Yellowstone distribution of *B. boreas* in an ecological context. Species distribution may be limited by growing-season precipitation, ridges, and impervious surfaces, resulting in the observed patchy distribution (Fig. 1, Table 2).

#### *Genetic metrics*

Our results demonstrate that genetic distance based on allele frequency distribution ( $D_{ps}$ ) is more powerful

and detects more recent landscape changes than heterozygosity-based genetic distance ( $F_{ST}$ ). Methods based on allele frequency distribution are more likely to detect recent events (Miller and Waits 2003), identifying landscape genetic structure within five generations post-change (Murphy et al. 2008). Conversely,  $F_{ST}$  may reflect past landscape condition (Holzhauer et al. 2006) and assumes population equilibrium, which is often violated in natural populations.

#### *Ecological process*

In contrast to other studies (Arnaud 2003, Manier and Arnold 2006), topographic distance was not the strongest predictor of connectivity as reflected by the comparatively small amount of variation explained (Table 2; Appendix C). Models of habitat permeability, topographic morphology, and temperature–moisture regime are all significant, explaining more than twice the variation in *B. boreas* connectivity compared to distance alone (Table 2). In addition, all ecological processes are represented when metrics are combined into a single model (Table 2), further supporting the inference that all three processes influence *B. boreas* connectivity.

Our results based on a novel application of RF are well supported in the literature. Habitats with mid-range canopy cover (such as shrubby meadow areas) are preferentially selected during dispersal, thus increasing connectivity (Bartelt et al. 2004). Connectivity of amphibians in Yellowstone may be fire dependent. Increased site occupancy by *B. boreas* has been observed in the years immediately following fire, possibly due to stable temperature of burrowing sites in burned areas (Hossack and Corn 2007). In addition, a study of tiger salamanders in northern Yellowstone found a positive correlation between burned habitats and gene flow (Spear et al. 2005). High topographic complexity (as measured by elevation relief ratio and topographic roughness) is associated with watershed boundaries and their associated geomorphology, which may be barriers for amphibians (Funk et al. 1999). Growing-season precipitation regulates breeding-site persistence (Corn 2003) and habitat moisture, both vital for *B. boreas* presence and dispersal. Metamorphosed *B. boreas* have little physiological control over evaporation (Bartelt and Peterson 2005); therefore, individuals likely prevent desiccation behaviorally by avoiding hot, exposed habitats. Finally, length of the growing season (frost-free period) affects time to metamorphosis in amphibians and may limit gene flow among populations (Giordano et al. 2007).

#### *Multiple scales*

Although landscape genetic studies generally only test a single scale, we demonstrate metrics of ecological processes influencing *B. boreas* connectivity operate across multiple scales. Fine-scale metrics include measures of local habitat condition, microclimate, and



topography in which local variability influences *B. boreas* connectivity (Table 2, Fig. 2). Two key metrics, growing-season precipitation and slope temperature–moisture, operate across scales suggesting a more general relationship with *B. boreas* connectivity (Fig. 2).

Each scale, analyzed independently, results in a slightly different inference about *B. boreas* connectivity. Incorporating all of these scales into a single analysis captures a more complete picture of the effects of ecological processes on *B. boreas*. The cross-scale model includes metrics measured at fine scales (canopy, elevation relief ratio), mid-scale (impervious surfaces), broad scale (ridges), and across multiple scales (growing-season precipitation and slope temperature–moisture; Fig. 2). However, cross-scale models only improved the percentage of variation explained by ~3–5%, indicating that the straight-line model (scale 0) had high predictive power. This model of connectivity may be performing well for two major reasons: inclusion of multi-scale metrics and long-term stability of the study system due to its protection status. Multi-scale metrics (i.e., multiple window sizes; Table 1) may capture the surrounding landscape, even at smaller bandwidths. In addition, although Yellowstone has areas of high human use, it is a relatively intact landscape. For this reason, historic and current gene flow may be very similar, strengthening the overall signal for models of connectivity. The influence of the surrounding landscape and subsequent scale effect may be more notable in highly fragmented or rapidly changing systems.

#### *Hierarchical effect*

Within-cluster variation in connectivity represents a finer spatial and temporal scale than between-cluster variation. Between clusters, genetic variation represents historic or long-term conditions. Therefore, it is not surprising that we are able to explain more variation in connectivity within clusters (74.03%) as compared to between clusters (44.85%; Table 2) using current landscape condition. Topographic distance is notably absent as an important variable in predicting *B. boreas* connectivity both within and between genetic clusters. These results reinforce the cross-scale analyses and suggest two things: (1) distance may be acting as a surrogate for the hierarchical effect in the ecological process and scale models and (2) landscape features can have a strong effect on connectivity, independent of distance.

Overall, connectivity within genetic clusters is governed by habitat permeability and moisture; areas with high growing-season precipitation, mid-range cover, and low density of impervious surfaces are associated with increased connectivity (Table 2). The overwhelming importance of growing-season precipitation within genetic clusters may indicate fragmentation due to drought conditions over the past 8–12 years (Table 2).

These effects are also consistent with apparent drought-induced genetic bottlenecks in tiger salamanders throughout northern Yellowstone (Spear et al. 2006). In addition, growing-season precipitation is important at multiple nonredundant scales, suggesting it represents several mechanisms. Local growing-season precipitation may represent average humidity and soil moisture. At the 960-m scale, growing-season precipitation may represent regional differences in temperature–moisture regime across Yellowstone.

Connectivity between genetic clusters can be explained by more general, mostly broad-scale metrics. Major topographic features, such as ridgelines, can explain the observed genetic groups, suggesting that topography is a barrier for *B. boreas* in Yellowstone. In contrast to the multiscale models that employ the whole data set, slope temperature–moisture only operates at mid to broad scales and degree days >5°C only at broad scales (Table 2). Mean annual precipitation replaces growing-season precipitation as the prominent measure of moisture across scales, reinforcing the more long-term nature of processes driving connectivity between genetic groups.

#### CONCLUSIONS

Our study presents a novel algorithmic approach for assessing connectivity and has several important implications for landscape genetics, *B. boreas* conservation, and understanding mechanisms of fine-scale species distributions. Isolation by distance and autocorrelation are important considerations in landscape genetics. However, even in relatively intact landscapes such as Yellowstone, distance may only explain a small portion of the variation in genetic connectivity compared to the variation explained in models including other potential explanatory processes. By using an algorithmic approach, we are able to simultaneously include multiple metrics of ecological processes to explain a high proportion of variation in genetic distance (up to 74%).

Because *B. boreas* is thought to be in decline throughout large portions of its range (Muths et al. 2003), it is essential to assess species' connectivity and vulnerability in protected areas where populations remain. At all scales (Table 2; Appendix C), precipitation and temperature are major drivers of *B. boreas* connectivity. As Yellowstone becomes warmer and drier as predicted due to climate change (Bartlein et al. 1997), functional connectivity in this landscape will likely degrade. In addition, niche space for species limited by temperature–moisture regimes will likely shift (Rehfeldt et al. 2006), potentially resulting in much of the presently occupied habitat being unsuitable. Knowledge of these threats may help inform management decisions; e.g., an increase in impervious surfaces due to development, also shown to limit this species' connectivity, may have a greater impact in a context of climate change than it would otherwise. Finally, implementation of landscape genetic approaches such as the one presented



here can provide robust estimates of population connectivity to assist in development of science-based conservation and management decisions.

#### ACKNOWLEDGMENTS

Research was permitted by Yellowstone National Park (Permit YELL-05452; C. Smith and C. Hendrix) and IACUC (ASAF 3378). We thank M. J. Murphy, M. Flisher, and D. Murphy for field assistance; C. Corkan, D. Corkan, D. Patla, and C. Peterson provided additional tissue samples. A. Gallant, P. Bartelt, and R. Klaver assisted in compiling spatial data. A. Cutler provided valuable discussion of RF. We thank S. Cushman, R. Dezanini, C. Goldberg, D. Patla, C. Peterson, D. Pilliod, S. Spear, M. Webser, L. Waits, and two anonymous reviewers for valuable comments and advice. Funding for M. A. Murphy was provided by EPA-STAR fellowship (FP-916695), NSF (grant DEB-0608458 to A. Storfer and M. A. Murphy), James King fellowship, Theodore Roosevelt Memorial Fund, Society for Wetland Science (2004, 2005), WSU Zoology alumni scholarship, Canon National Parks, Sigma-Xi grants-in-aid (2004), and Graduate Women in Science. A. Storfer was supported by NSF DEB-0548415. The research described in this paper has been funded wholly or in part by the U.S. Environmental Protection Agency (EPA) under the Science to Achieve Results (STAR) Graduate Fellowship Program. EPA has not officially endorsed this publication, and the views expressed herein may not reflect the views of the EPA.

#### LITERATURE CITED

- Arens, P., T. van der Sluis, W. P. van't Westende, B. Vosman, C. C. Vos, and M. J. M. Smulders. 2007. Genetic population differentiation and connectivity among fragmented Moor frog (*Rana arvalis*) populations in The Netherlands. *Landscape Ecology* 22:1489–1500.
- Arnaud, J.-F. 2003. Metapopulation genetic structure and migration pathways in the land snail *Helix aspersa*: influence of landscape heterogeneity. *Landscape Ecology* 18:333–346.
- Bartelt, P., and C. R. Peterson. 2005. Physical modeling operative temperature and evaporative water loss in amphibians. *Journal of Thermal Biology* 30:93–102.
- Bartelt, P., C. R. Peterson, and R. W. Klaver. 2004. Sexual differences in the movements and habitats selected by western toads in southeastern Idaho. *Herpetologica* 60:455–467.
- Bartlein, P. J., C. Whitlock, and S. L. Shafer. 1997. Future climate in the Yellowstone National Park region and its potential impact on vegetation. *Conservation Biology* 11:782–792.
- Becker, R. A., J. M. Chambers, and A. R. Wilks. 1988. The new S language: a programming environment for data analysis and graphics. Wadsworth, Pacific Grove, California, USA.
- Beebee, T. J. C. 2005. Conservation genetics of amphibians. *Heredity* 95:423–427.
- Bowcock, A. M., A. Ruiz-Linares, J. Tomfohrde, E. Minch, J. R. Kidd, and L. L. Cavalli-Sforza. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368:455–457.
- Breiman, L. 2001. Random Forests. *Machine Learning* 45:5–32.
- Cavalli-Sforza, L. L., and A. W. F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics* 19:233–257.
- Chon, T.-S., and V.-S. Park. 2006. Ecological informatics as an advanced interdisciplinary interpretation of ecosystems. *Ecological Informatics* 1:213–217.
- Corn, P. S. 2003. Amphibian breeding and climate change: importance of snow in the mountains. *Conservation Biology* 17:622–625.
- Cutler, D. R., T. C. Edwards, Jr., K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random Forests for classification in ecology. *Ecology* 88:2783–2792.
- De'ath, G., and K. E. Fabricius. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81:3178–3192.
- Evans, I. S. 1972. General geomorphometry, derivatives of altitude, and descriptive statistics. Pages 17–90 in R. J. Chorley, editor. *Spatial analysis in geomorphology*. Harper and Row, New York, New York, USA.
- Fu, P., and P. M. Rich. 1999. Design and implementation of the Solar Analyst: an ArcView extension for modeling solar radiation at landscape scales. In 19th Annual ESRI User Conference. Environmental Systems Research Institute, San Diego, California, USA.
- Funk, W. C., D. A. Tallmon, and F. W. Allendorf. 1999. Small effective population size in the long-toed salamander. *Molecular Ecology* 8:1633–1640.
- Giordano, A. R., B. J. Ridenhour, and A. Storfer. 2007. The influence of altitude and topography on genetic structure in the long-toed salamander (*Ambystoma macrodactylum*). *Molecular Ecology* 16:1625–1637.
- Goldberg, C. S., M. E. Kaplan, and C. R. Schwable. 2003. From the frog's mouth: buccal swabs for collection of DNA from amphibians. *Herpetological Review* 34:220–221.
- Green, D. M. 2003. The ecology of extinction: population fluctuation and decline in amphibians. *Biological Conservation* 111:331–343.
- Hanski, I., and O. E. Gaggiotti, editors. 2004. *Ecology, genetics, and evolution of metapopulations*. Elsevier, Burlington, Massachusetts, USA.
- Holzhauser, S., K. Ekschmitt, A.-C. Sander, J. Dauber, and V. Wolters. 2006. Effect of historic landscape change on the genetic structure of the bush-cricket *Metrioptera roeseli*. *Landscape Ecology* 21:891–899.
- Hossack, B. R., and P. S. Corn. 2007. Responses of pond-breeding amphibians to wildfire: short-term patterns in occupancy and colonization. *Ecological Applications* 17:1403–1410.
- Keyghobadi, N., J. Roland, S. F. Matter, and C. Strobeck. 2005. Among- and within-patch components of genetic diversity respond at different rates to habitat fragmentation: an empirical demonstration. *Proceedings of the Royal Society B* 272:553–560.
- Koch, E. D., and C. R. Peterson. 1995. *Amphibians and reptiles of Yellowstone and Grand Teton National Parks*. University of Utah Press, Salt Lake City, Utah, USA.
- Liaw, A., and M. Wiener. 2002. Classification and regression by Random Forest. *R News* 2:18–22.
- Manier, M. K., and S. J. Arnold. 2006. Ecological correlates of population genetic structure: a comparative approach using a vertebrate metacommunity. *Proceedings of the Royal Society B* 273:3001–3009.
- Mazerolle, M. J. 2004. Amphibian road mortality in response to nightly variations in traffic intensity. *Herpetologica* 60:45–53.
- McCune, B., and D. Keon. 2002. Equations for potential annual direct incident radiation and heat load. *Journal of Vegetation Science* 13:603–606.
- McGarigal, K., and B. J. Marks. 1995. FRAGSTATS. Spatial analysis program for quantifying landscape structure. USDA Forest Service General Technical Report PNW-GTR-351. Pacific Northwest Research Station, Portland, Oregon, USA.
- Miller, C. R., and L. P. Waits. 2003. The history of effective population size and genetic diversity in the Yellowstone grizzly (*Ursus arctos*): implications for conservation. *Proceedings of the National Academy of Sciences (USA)* 100:4334–4339.
- Moore, I., P. Gessler, G. Nielsen, and G. Petersen, editors. 1993. *Terrain attributes and estimation methods and scale effects*. Wiley, London, UK.
- Morin, X., C. Augspurger, and I. Chuine. 2007. Process-based modeling of species' distributions: What limits temperate tree species' range boundaries? *Ecology* 88:2280–2291.

- Murphy, M. A., J. S. Evans, S. Cushman, and A. Storfer. 2008. Evaluation of a novel approach for representing "populations" as continuous surfaces in landscape genetics. *Ecography* 31:685–697.
- Muths, E., P. S. Corn, A. P. Pessier, and D. E. Green. 2003. Evidence for disease-related amphibian decline in Colorado. *Biological Conservation* 110:357–365.
- Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583–590.
- Ovaskainen, O., and I. Hanski. 2004. Metapopulation dynamics in highly fragmented landscapes. Pages 73–104 in I. Hanski and O. E. Gaggiotti, editors. *Ecology, genetics, and evolution of metapopulations*. Elsevier, Burlington, Massachusetts, USA.
- Palo, J. U., R. B. O'Hara, A. T. Laugen, A. Laurila, C. R. Primmer, and J. Merila. 2003. Latitudinal divergence of common frog (*Rana temporaria*) life history traits by natural selection: evidence from a comparison of molecular and quantitative genetic data. *Molecular Ecology* 12:1963–1978.
- Pechmann, J. H. K., D. E. Scott, R. D. Semlitsch, J. P. Caldwell, L. J. Vitt, and J. W. Gibbons. 1991. Declining amphibian populations: the problem of separating human impacts from natural fluctuations. *Science* 253:892–895.
- Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Pulliam, H. R. 2000. On the relationships between niche and distribution. *Ecology Letters* 3:349–361.
- R Development Core Team. 2007. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rehfeldt, G. E. 2006. A spline model of climate for the western United States. General Technical Report 165. USDA Forest Service, Rocky Mountain Research Station, Fort Collins, Colorado, USA.
- Rehfeldt, G. E., N. L. Crookston, M. V. Warwell, and J. S. Evans. 2006. Empirical analysis of plant–climate relationships for western United States. *International Journal of Plant Sciences* 167:1123–1150.
- Riley, S. J., S. D. DeGloria, and R. Elliot. 1999. A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences* 5:23–27.
- Simandle, E. T., M. M. Peacock, L. Zirelli, and C. R. Tracy. 2006. Sixteen microsatellite loci for the *Bufo boreas* group. *Molecular Ecology Notes* 6:116–119.
- Smith, A. M., and D. M. Green. 2005. Dispersal and the metapopulation paradigm in amphibian ecology and conservation: Are all amphibian populations metapopulations? *Ecography* 28:110–128.
- Spear, S. F., C. R. Peterson, M. M. Maticq, and A. Storfer. 2005. Landscape genetics of the blotched tiger salamander (*Ambystoma tigrinum melanostictum*). *Molecular Ecology* 14:2553–2564.
- Spear, S., C. R. Peterson, M. M. Maticq, and A. Storfer. 2006. Molecular evidence for historical and recent population size reductions of tiger salamanders (*Ambystoma tigrinum*) in Yellowstone National Park. *Conservation Genetics* 7:605–611.
- Stage, A. 1976. An expression of the effects of aspect, slope, and habitat type on tree growth. *Forest Sciences* 22:457–469.
- Stevens, C. E., C. A. Paszkowski, and A. L. Foote. 2007. Beaver (*Castor canadensis*) as a surrogate species for conserving anuran amphibians on boreal streams in Alberta, Canada. *Biological Conservation* 134:1–13.
- Storfer, A., M. A. Murphy, J. S. Evans, C. Goldberg, S. Spear, S. Robinson, R. Dezzani, E. Demmelle, L. Vierling, and L. P. Waits. 2007. Putting the 'landscape' in landscape genetics. *Heredity* 98:128–142.
- Turner, M. G. 1989. Landscape ecology: the effect of pattern on process. *Annual Review of Ecology and Systematics* 20:171–197.
- USFWS. 2002. National Wetlands Inventory. (<http://www.fws.gov/wetlands/Data/DataDownload.html>)
- Wagner, H. H., and M.-J. Fortin. 2005. Spatial analysis of landscapes: concepts and statistics. *Ecology* 86:1975–1987.
- Wang, J. 2004. Sibship reconstruction from genetic data with typing errors. *Genetics* 166:1963–1979.
- Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.
- Wright, S. 1951. The genetical structure of populations. *Annals of Eugenics* 15:323–354.
- Zeverbergen, L. W., and C. R. Thorne. 1987. Quantitative analysis of land surface topography. *Earth Surface Processes and Landforms* 12:47–56.

## APPENDIX A

Additional genetic methods and amplification conditions (*Ecological Archives* E091-018-A1).

## APPENDIX B

Definitions for newly derived topographic morphology metrics (*Ecological Archives* E091-018-A2).

## APPENDIX C

Full table of model results (expanded from Table 2) (*Ecological Archives* E091-018-A3).

## APPENDIX D

Color version of Fig. 1 from text and maps of important variables retained in Random Forests models (*Ecological Archives* E091-018-A4).