Model application niche analysis: assessing the transferability and generalizability of ecological models

Jessica B. Moon,1,2 Theodore H. DeWitt,3 Melissa N. Errend,4 Randall J. F. Bruins,5 Mary E. Kentula,6 Sarah J. Chamberlain,7 M. Siobhan Fennessy,8 and Kusum J. Naithani2

1Oak Ridge Institute for Science and Education (ORISE) Post-doctoral Participant, National Health and Environmental Research Effects Laboratory, Western Ecology Division, Pacific Coast Ecology Branch, U.S. Environmental Protection Agency, Newport, Oregon 97365 USA
2Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas 72701 USA
3National Health and Environmental Research Effects Laboratory, Western Ecology Division, Pacific Coast Ecology Branch, U.S. Environmental Protection Agency, Newport, Oregon 97365 USA
4College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331 USA
5National Exposure Research Laboratory, Systems Exposure Division, U.S. Environmental Protection Agency, Cincinnati, Ohio 45268 USA
6National Health and Environmental Effects Research Laboratory, Western Ecology Division, U.S. Environmental Protection Agency, Corvallis, Oregon 97333 USA
7Department of Geography, Riparia, The Pennsylvania State University, University Park, Pennsylvania 16802 USA
8Department of Biology, Kenyon College, Gambier, Ohio 43022 USA

Citation: Moon, J. B., T. H. DeWitt, M. N. Errend, R. J. F. Bruins, M. E. Kentula, S. J. Chamberlain, M. S. Fennessy, and K. J. Naithani. 2017. Model application niche analysis: assessing the transferability and generalizability of ecological models. Ecosphere 8(10):e01974. 10.1002/ecs2.1974

Abstract. The use of models by ecologists and environmental managers, to inform environmental management and decision-making, has grown exponentially in the past 50 yr. Due to logistical, economical, and theoretical benefits, model users frequently transfer preexisting models to new sites where data are scarce. Modelers have made significant progress in understanding how to improve model generalizability during model development. However, models are always imperfect representations of systems and are constrained by the contextual frameworks used during their development. Thus, model users need better ways to evaluate the possibility of unintentional misapplication when transferring models to new sites. We propose a method of describing a model's application niche for use during the model selection process. Using this method, model users synthesize information from databases, past studies, and/or past model transfers to create model performance curves and heat maps. We demonstrated this method using an empirical model developed to predict the ecological condition of plant communities in riverine wetlands of the Appalachian Highland physiographic region, USA. We assessed this model's transferability and generalizability across (1) riverine wetlands in the contiguous United States, (2) wetland types in the Appalachian Highland physiographic region, and (3) wetland types in the contiguous United States. With this methodology and a discussion of its critical steps, we set the stage for further inquiries into the development of consistent and transparent practices for model selection when transferring a model.

Key words: ecological models; model application niche analysis; model generalizability; model selection; model transferability.

Received 11 April 2017; revised 20 August 2017; accepted 23 August 2017. Corresponding Editor: Debra P. C. Peters.
Copyright: © 2017 Moon et al. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
† E-mail: jbmoon@me.com
INTRODUCTION

The use of quantitative models (e.g., statistical, mechanistic, indices) in ecology has proliferated since the early 1970s, with an increasing number developed to inform environmental management and policy (NRC 2007, Jørgensen 2008, Jørgensen and Nielsen 2012). Due to significant logistical, economical, and theoretical benefits, model users are not only implementing models at the sites for which they are developed (herein, development sites), but are increasingly transferring models to alternative sites (herein, application sites) where data can be scarce. Models are transferred across at least one of five dimensions of context, including spatial scale (e.g., plot, watershed, region, continent), temporal scale (e.g., minutes, hours, months, years), geographic location (e.g., watershed to watershed), time (e.g., year to year), and ecological organizational unit (e.g., species, habitat type, ecosystem type; Wu and Li 2006).

Model developers are always looking for ways to improve model generalizability. For instance, some modelers have highlighted the importance of process-based model equations (Goodall 1972, Vanreusel et al. 2007, Amano 2012, Wenger and Olden 2012), or the incorporation of equations that describe physical laws (Baird et al. 2003), functional or trait-based relationships (Pont et al. 2006, Pease et al. 2015), or causal rather than correlational predictors (Bamford et al. 2009, Estes et al. 2011). Others have emphasized the importance of correctly identifying the scale(s) over which the predictor-response relationships operate (De Marco et al. 2008, Heikkinen et al. 2012), and the appropriate level of model complexity (Meynard and Quinn 2007, Fukuda and Hiramatsu 2008, McDonald and Urban 2010, Falk and Mellert 2011). Despite these efforts, model developers cannot anticipate all of the contexts that model users might transfer their models to, and it is incumbent on model users to carefully select models for their intended purpose and system.

Ecological model developers and users do not have a consistent methodology for informing model selection based on the past performance of candidate models (Miller et al. 2004, NRC 2007). In a 30-yr review of the legal challenges of predictive models used in regulatory decision-making, McGarity and Wagner (2003) found that transferring models to contexts that differed from the model development contexts was one of the biggest vulnerabilities to their legal defensibility. Those who have argued successful legal challenges (e.g., Ohio vs. USEPA in McGarity and Wagner 2003) have done so by arguing that the models in question were used outside their respective “application niche[s],” that is, “...a particular prototype physical system or set of defining conditions for which the model code is scientifically defensible” (USEPA ATFERM 1994: II-2). Establishing rigorous model transfer criteria is important not only for legal proceedings, but also for model users to demonstrate transparency in their decision-making process.

We developed a method for assessing an ecological model’s transferability and generalizability, which we call a model application niche analysis. Our methodology is conceptually based on Tobler’s first law of spatial autocorrelation. This law states that the closer two points are in space or time the more similar they are likely to be in terms of their structure and function (Tobler 1970). Economists have adopted this approach for benefit transfer models, expanding the law of spatial autocorrelation to contextual similarity. They suggest that the more contextually similar an application site is to the model development site, the more likely it is that a model transfer will be successful (Rosenberger and Phipps 2007). The importance of contextual similarity has also been recognized in species distribution modeling (also known as ecological niche modeling). These modelers have created tools to identify when application sites (also known as projection or calibration domains) have model predictor values outside the bounds of predictor values found across sites used in model development (also known as reference or transfer domains; Elith et al. 2010, Zurell et al. 2012, Owens et al. 2013, Mesgaran et al. 2014). However, these modelers also warn that the identification of model extrapolation does not circumvent the need for model evaluation (Elith et al. 2010, Mesgaran et al. 2014).

Our method evaluates the transferability and generalizability of a model by testing the importance of contextual similarity using model validation results (also known as operational validation results; Rykiel 1996) from model application sites. We acknowledge that the need to gather and evaluate model validation data could be considered a logistical challenge of this methodology. Most
concerning, model validations are not performed regularly in ecological modeling. This issue has been highlighted in reviews of oceanographic models (Stow et al. 2009), aquatic biogeochemical models (Arhonditsis and Brett 2004), and individual-based ecological models (Kubicek et al. 2015). In addition to the risks associated with using model outputs that have not been validated, there are additional costs of not reporting successful and unsuccessful model validations. It can hinder one’s ability to identify a model’s application niche for use during model selection.

We begin this paper with an overview of our methodology for evaluating a model’s application niche. We then provide a case study, where we assess a riverine wetland plant condition model’s transferability and generalizability across spatial locations (i.e., transfers across the United States), across organizational units (i.e., transfers across wetland types), and across both spatial locations and organizational units. We end with a discussion of some critical considerations that affect the utility of a model application niche analysis and offer ideas on future lines of inquiry to make this approach practical for a wider range of ecological models.

**METHODOLOGY**

Tobler’s first law of spatial autocorrelation is an underpinning of geostatistics, a set of spatial statistics used to describe the spatial and/or temporal patterning of a variable (for use in ecology, see Legendre and Fortin 1989). A common tool in geostatistics is the semivariogram—a graph of semivariance between measurements (i.e., pairwise comparisons) of a variable as a function of the physical distance (i.e., lag distance) between the measurement locations (Fig. 1a). By fitting a figure was created using ggplot2 (Wickham 2009) in R (version 3.4.0; R Core Team 2017).

![Diagram](image-url)
curve to the semivariogram, we can estimate model parameters that explain the variables’ spatial structure. Common parameters include the nugget, the sill, the partial sill, and the range. The nugget is the semivariance at a lag distance of zero. It represents unexplained variability due to variance that exists at finer scales than were measured and/or analytical measurement error(s). The sill is the semivariance where the curve reaches an asymptote. The partial sill is the semivariance that is left over after subtracting the nugget from the sill, and represents the spatial variance of the variable. The range is the distance at which the asymptote is reached, and defines the distance (e.g., m and km) after which two measurements of the variable are considered statistically independent from each other.

When model performance is the variable of interest, a semivariogram can be used to interpolate a model’s performance across the domain of the application sites (i.e., measurement sites). This model performance map can be used to identify the model’s application niche and connect model performance to underlying biophysical variables (for similar concepts, see Gardner and performance to underlying biogeophysical vari-

ces). The data requirements to perform a semivariogram, a model can be fitted to the data to estimate parameters that explain a model’s transferability and generalizability. In a model performance plot, there are two nuggets, one associated with the model development error (C_D) and one associated with the model application error (C_A). C_D is known as primary site error or measurement error in economic benefit transfer modeling (Rosenberger and Stanley 2006). It includes the unresolved variability in the process being modeled, and the error related to the methods and model assumptions. C_A includes C_D and the contextual interpolation model error. Fig. 1b shows a scenario where C_A equals C_D. In this case, no model error is added when transferring the model to application sites that are contextually the same as sites used in model development. The sill (C_T) includes C_A and the contextual extrapolation model error. Subtracting C_D from C_T results in the maximum model generalizability error, C_Gmax. The model generalizability errors (C_G) make up the function (i.e., curve shape, slope) that describes the relationship between model error and contextual dissimilarity. This is the model error that arises when transferring the model to conditions that differ from those under which the model was developed (Rosenberger and Stanley 2006, Liu et al. 2011), and includes uncertainties associated with methodological errors or poor model calibration across the application sites. The model’s fit (e.g., adj. R^2, Root Mean Squared Error) describes how much of the variability in C_G can be explained by the contextual dissimilarity metric being evaluated.

Finally, the contextual range parameter (R_C) defines the level of contextual dissimilarity when C_G is greater than zero, or where C_G crosses a predeter
dined level (i.e., user-defined) of acceptable model error. The larger R_C is relative to the contextual gradient evaluated (R_max), the more generalizable the model is expected to be. A fully generalizable model would be expressed as a nugget model (i.e., slope of zero) that intersects the y-axis at the model error associated with C_D (i.e., where C_T ~ C_A ~ C_D and R_C ~ R_max). Applying the law of spatial autocorrelation to contextual dissimilarity, Fig. 1b depicts an expected function (i.e., exponential), where model error increases as the model is transferred to application sites that are increasingly dissimilar from the context in which the model was developed.

**Case Study**

**Wetland ecological condition model**

To demonstrate this methodology, we developed an empirical model that predicts a wetland’s mean coefficient of conservatism (mean...
mean CoC value is a plant community index that has been used as a measure of wetland ecological condition (Lopez and Fennessy 2002, Bried et al. 2013), including for wetland restoration and creation efforts (Gutrich et al. 2009, Matthews et al. 2009). To calculate a wetland’s mean CoC value, each plant species is assigned a value between 0 and 10 (determined by best professional judgment, e.g., Chamberlain and Ingram 2012) based on the probability of the species’ presettlement occurrence in a region and its relative tolerance to anthropogenic disturbance (Swink and Wilhelm 1994, Nichols 1999). Higher values are given to plants that are sensitive to human disturbances, while lower values are given to widespread, generalist species (Wilhelm and Ladd 1988). The mean CoC value is the average of the CoCs of all plant species found within a wetland.

Species CoC values vary across regions reflecting the geographic differences in plant species tolerances and are not available in some regions of the United States (see section 5.9 in USEPA 2016). As such, a wetland mean CoC model can be useful for transfers when there is a need to prioritize the selection of wetlands for conservation and restoration efforts. We developed this model using data collected from riverine wetlands across Pennsylvania, USA, a part of the Appalachian Highland physiographic province (Bishel-Machung et al. 1996, Miller and Wardrop 2006, Miller et al. 2006, 2009, Cleveland et al. 2011, Chamberlain and Ingram 2012, Drohan and Brooks 2013, Chamberlain and Brooks 2016, R. P. Brooks and D. H. Wardrop, unpublished data). Within this dataset, we identified 107 riverine wetlands for use in model construction (Fig. 2). Twenty-two percent (n = 24) of these wetlands were excluded at random for model validation. We obtained mean CoC values for each wetland using site plant lists and the Pennsylvania State Riparian Floristic Quality Assessment Calculator (i.e., identified as the Total Mean C value in output; http://apps.cei.psu.edu/fqacalc/).

The wetland ecological condition model predicts a wetland’s mean CoC value using on-site soil surface pH, and the landscape development index (LDI) for a 1.26-km² circular area (200 m radius) around the wetland’s center (Table 1). The LDI describes the level of anthropogenic activity in a given area and is measured as a weighted average of land cover types in that area (Brown and Vivas 2005). The coefficients for each land cover type are calculated as the normalized natural log of the energy use per unit area per time (sej ha⁻¹ yr⁻¹) and are defined as the amounts of nonrenewable energy needed to sustain a given land use activity (Brown and Vivas 2005). To calculate an LDI value for each wetland, we used a crosswalk developed by Lane et al. (2012) between the original LDI coefficients (Brown and Vivas 2005) and the land cover categories in the National Land Cover Database (NLCD). We extracted the 2011 NLCD (Homer et al. 2015) data using ArcGIS (version 10.5; Environmental Systems Research Institute, Redlands, California, USA). The data used in the construction and validation of this model can be found in Data S1 (also see Data S1 Column Explanations and Metadata S1).

We assessed our model’s transferability and generalizability to (1) riverine wetlands across the contiguous United States (spatial transfer), (2) wetland types within the Appalachian Highland physiographic region (organizational unit transfer), and (3) wetland types across the contiguous United States (spatial-organizational transfer). Currently, mean CoC values (or the vegetation community compositions lists on which mean CoC values are based) and soil surface pH levels are only available for a limited number of published studies, predominantly across the eastern United States (e.g., Lopez and Fennessy 2002, Reiss 2006, Boughton et al. 2010, Dee and Ahn 2012). However, the 2011 U.S. EPA National Wetland Condition Assessment (NWCA) database (USEPA 2016) reports soil surface pH values (identified as Layer 1 for PH_H2O in the NWCA database) for 825 wetlands across the contiguous United States (Fig. 2), including 194 riverine sites and 112 Appalachian Highland sites. Mean CoC values were provided for these sites upon request, and we calculated LDI values for each wetland based on the wetland assessment area center (identified as AA_CENTER_LAT and AA_CENTER_LON in the NWCA database). Data for the application sites can be found in Data S2 (also see Data S2 Column Explanations and Metadata S2).

**Analyses**

Semivariogram of model performance.—For comparison to the model application niche analysis, we performed a semivariogram analysis on model
error across the full extent of wetlands in the NWCA dataset (n = 825). Model error was measured as the difference between our application sites’ measured mean CoC values and their average modeled mean CoC values. We used weighted least squares to fit a powered exponential model to the semivariance of model errors across distance bins of 200 km. Distance bins ran from 0 to 4000 km, with a minimum of 30 point-pairs per bin. We subsequently used parameter estimates from the semivariogram model in ordinary kriging, with a constant mean and a global neighborhood, to interpolate model error and model error uncertainty (i.e., standard deviation of model error) across the contiguous United States. We performed our analyses in the geoR package (Ribeiro and Diggle 2001) of R (version 3.4.0; R Core Team 2017).

**Model application niche analysis.**—To create model performance plots for the wetland ecological condition model across model application sites, we constructed contextual dissimilarity indices and calculated a measure of model error. We created one contextual dissimilarity index based on variables intrinsic to the model (i.e., model predictors) and two contextual dissimilarity indices.
based on variables extrinsic to the model (i.e., variables not included explicitly in the model) that were expected to affect the model’s spatial or organizational transfer. We used the absolute difference between measured mean CoC values and average modeled mean CoC values as our measure of model error. Both linear and non-linear model performance curves were evaluated using weighted least squares to extract model performance parameters across the contextual dissimilarity gradients. When a significant model fit could not be obtained, a nugget model (i.e., zero-slope model) was evaluated to determine whether there was a significant contextual interpolation error (i.e., \( H_{ij} = C_A = C_D \)). To create the nugget model, we used the mean model error across all application sites within the dataset. We conducted model evaluations in the base package of R (version 3.4.0; R Core Team 2017) and in JMP® Pro (version 12.2.0; 2015 SAS Institute).

1. Intrinsic contextual dissimilarity index.—We created an intrinsic contextual dissimilarity index by combining measures of univariate dissimilarity for model predictors. These included soil pH and the percent land cover categories used to construct the LDI. The LDI has the same emergy value (i.e., their weights in the LDI equation are equal) for percent cover of forests, shrubs, wetlands, and water cover types (Table 1). As such, the percent cover of these land cover classes was combined into one variable before creating the index. To create the intrinsic dissimilarity index, we used equations developed by Mesgaran et al. (2014) with minimal adaptation (i.e., we used the absolute value to obtain a positive gradient of dissimilarity). For each application site \( i \) and contextual variable \( j \) combination, we calculated a univariate distance (\( UD_{ij} \)) index:

\[
UD_{ij} = \left| \min \left\{ A_{ij} - \min(D_j), \max(D_j) - A_{ij}, 0 \right\} \right| \frac{\max(D_j) - \min(D_j)}{\max(D_j) - \min(D_j)}
\]

where \( A_{ij} \) is the value of the contextual variable, \( j \) (e.g., soil pH, grassland, and barren lands), at the model application site, \( i \). The \( \min(D_j) \) and the \( \max(D_j) \) are the minimum and maximum values

| Model variables \((n = 83)\)† | Coefficients | 95% CI‡ | Adj. \( R^2 \) | RSE |
|---|---|---|---|---|
| pH | –0.73 | –0.86, –0.59 | 0.67 | 0.64 |
| LDI | –0.29 | –0.41, –0.18 | 0.67 | 0.64 |
| Intercept | 8.16 | 7.38, 8.93 | 0.67 | 0.64 |
| LDI weights§ | | | | |
| Open water | 1.00 | | | |
| Forest¶ | 1.00 | | | |
| Shrub/scrub | 1.00 | | | |
| Wetlands# | 1.00 | | | |
| Grassland/herbaceous | 1.89 | | | |
| Pasture/hay | 3.31 | | | |
| Cultivated crops | 5.13 | | | |
| Developed, open spaces | 5.22 | | | |
| Developed, low intensity | 6.90 | | | |
| Developed, medium intensity | 7.51 | | | |
| Developed, high intensity | 8.93 | | | |
| Barren land | 4.38 | | | |

† Model datasets come from Bishel-Machung et al. (1996), Miller and Wardrop (2006), Miller et al. (2006, 2009), Cleveland et al. (2011), Chamberlain and Ingram (2012), Drohan and Brooks (2013), Chamberlain and Brooks (2016), and R. P. Brooks and D. H. Wardrop, unpublished data.

‡ Ordinary non-parametric bootstrap confidence intervals using 10,000 replications in the boot package (Davison and Hinkley 1997, Canty and Ripley 2016) of R (version 3.4.0; R Core Team 2017).

§ Landscape development index (LDI) coefficients come from Lane et al. (2012), who created a crosswalk between land cover classifications used by Brown and Vivas (2005) and land cover classifications in the National Land Cover Database (Homer et al. 2015).

¶ Deciduous, evergreen, and mixed forests.

# Woody and herbaceous wetlands.

Table 1. Description of the wetland mean coefficient of conservatism (mean CoC) model variables, coefficients, and model fit, including the 95% confidence intervals (CI), adjusted \( R^2 \) (adj. \( R^2 \)), and residual squared error (RSE).
of contextual variable, \( j \), found across the model development site. When \( UD_{ij} = 0 \), it means that the value for \( j \) at \( i \) is within the range of the value of \( j \) at sites used in model development (i.e., \( \min(D_j) \leq A_{ij} \leq \max(D_j) \)). When \( UD_{ij} > 0 \), it means the value for \( j \) at \( i \) is outside the range of the value of \( j \) at sites used in model development (i.e., \( A_{ij} \leq \min(D_j) \) or \( A_{ij} \geq \max(D_j) \)).

One approach would be to create model performance plots for each univariate dissimilarity index, separately. However, in our case study few application sites fell outside any one predictor’s range. As such, we opted to calculate an additive multivariate dissimilarity index (Intrinsic MD) for each application site, \( i \), by combining the UD values across all contextual variables (\( j \)):

\[
\text{Intrinsic MD}_i = \sum_{j=1}^{p} UD_{ij},
\]

where \( p \) (\( p = 10 \)) is the total number of contextual variables included in the index.

2. Spatial contextual dissimilarity index.—We hypothesized that the model would be more transferable to riverine wetlands in regions of the United States that have naturally derived acidic soil conditions, and less transferable to riverine wetlands in regions of the United States that have naturally derived neutral-to-basic soil conditions. Soil pH varies widely within the Appalachian Highland physiographic region and is known to affect vegetative community composition (Simon et al. 2005). Acidic soils naturally occur in Pennsylvania, due in part to the abundance of rainfall and subsequent leaching of alkaline ions (Blumberg and Cunningham 1982, Fabian and Brittingham 2007). This is particularly true in higher elevation forested wetlands, which have acidic shale-derived soils. Wetlands in these acidic soils have plant communities with higher mean CoC values reflecting a higher proportion of specialist species. Contrastingly, lower mean CoC values are generally found in wetlands located in Pennsylvania’s valleys. Valleys have neutral-to-basic soils, due in part to geology (i.e., Alfisols formed over carbonate and limestone; Miller 1995) and in part to land use activities (e.g., liming).

We expected the model to underestimate the mean CoC values of application sites with naturally derived neutral-to-basic soil, particularly in regions with low human impact. To test this hypothesis, we gathered data for contextual variables associated with soil formation, including mineralogical and geochemical data for the A horizon and C horizon (using the nearest 2007 sampling location, 1 site per 1600 km²; Smith et al. 2013), temperature and precipitation maps (WorldClim 2, average for 1970–2000, 1 km spatial resolution; Fick and Hijmans 2017), spatial prediction maps for surface soil texture and organic matter (SoilGrid, 250-m soil maps; Hengl et al. 2017), elevation maps (100 m resolution; USGS 2012), and associated stream order information (NHDPlus, 1:100,000-scale stream network; McKay et al. 2012). We extracted contextual datasets using ArcGIS (version 10.5; Environmental Systems Research Institute, Redlands, California, USA).

We created a weighted multivariate spatial dissimilarity index using directional univariate distance indices. We chose a weighted index, because we did not expect all univariate distance indices to have an equal effect on model error. We also chose to use directional univariate distance indices, because we did not want to assume that the effects of context on model error would be the same above and below the model development site’s contextual range. For these metrics only values above or below the model development site’s range were considered dissimilar. We used stepwise linear regression (backward and forward) to narrow down the number of directional univariate distance indices that explained most of the variance in model error. We subsequently evaluated combinations of these univariate indices to identify the fewest number that explained the greatest amount of variance in model error (using Akaike’s information criterion values, analyses of variance, and adj. \( R^2 \)). Slope coefficients from the selected linear model were used as weights for each contextual variable in the multivariate dissimilarity index.

The spatial multivariate dissimilarity index (Spatial MD) for application site, \( i \), was composed of directional elevation (DE) and temperature (DT) gradients with the following weights:

\[
\text{Spatial MD}_i = 0.35 \times \text{DE}_i + 0.18 \times \text{DT}_i
\]

In other words, riverine application sites in higher elevation regions and in higher mean annual temperature regions had larger model errors, compared to riverine application sites at
elevations and mean annual temperatures found within or below the model’s intended range. We performed model evaluations using the MASS package (Venables and Ripley 2002) and the base package of R (version 3.4.0; R Core Team 2017).

3. Organizational contextual dissimilarity index.— We evaluated the model’s transferability and generalizability across wetland types within the Appalachian Highland physiographic region by developing a multivariate dissimilarity index based on the commonly used hydrogeomorphic (HGM) classification system (Brinson 1993). This qualitative scheme identifies differences in wetlands based on their water source(s), hydrodynamics, and geomorphic setting. Peterson-Smith et al. (2009) found that HGM classes (i.e., headwater floodplains, riparian depressions, and slope wetlands) were related to soil properties and plant community composition across this region (e.g., New York, Pennsylvania, Virginia, and West Virginia). Further, Bried et al. (2013) found that the relationship between wetland condition and mean CoC values varied between hydrogeomorphic settings (i.e., hydrologically connected and isolated wetlands) of non-forested sites in upstate New York.

First, we created a dissimilarity index across HGM classifications, by identifying the probability (i.e., 0 = not likely, 0.5 = possible, 1.0 = likely) or frequency (i.e., daily = 0, seasonally = 0.5, annual = 1) of occurrence for six attributes related to water sources and hydrodynamics (Table 2). We assigned values to each HGM class, based on best professional judgment. For each HGM class, \( q \), the HGM dissimilarity index (\( HGM_q \)) was calculated as:

\[
HGM_q = \sum_{a=1}^{r} \text{abs}(P[R_a] - P[HGM_{aq}])
\]

where \( P[R_a] \) is the probability (or frequency) of occurrence in a riverine wetland for attribute, \( a \) (\( a = 1 \ldots r \), where \( r \) is the number of attributes), and \( P[HGM_{aq}] \) is the probability (or frequency) of occurrence for attribute, \( a \), for HGM class, \( q \). The absolute differences were summed across all attributes for each HGM class. Hydrogeomorphic classification rankings fell in the order: riverine wetlands, depressions, fringing wetlands, flats, with slopes and tidal wetlands being equally different from riverine sites.

This broke application sites into quantitative groups based on their \( HGM_q \). However, we also considered a few site-level differences. We added a water salinity index, \( WS_{it} \), where \( t \) describes the water salinity class (i.e., fresh, brackish, or saline) for application site, \( i \). These salinity classes were assigned in the field for the application sites (identified as \text{WATER\_SALINITY} in the NWCA database), and classes were given ordinal rankings (i.e., fresh = 0, brackish = 0.5, and saline = 1.0). Similar to the calculations for \( HGM_q \), we calculated this index using the absolute difference between the water salinity class found at sites used in model development (i.e., fresh) and the salinity class at each of the application sites. We also considered using a plant community index (i.e., based on aquatic bed, herbaceous, scrub/shrub, and forested classes). However, the model development dataset included the full range of vegetation categories. As such, this index would have no effect on the relative differences between application sites along the dissimilarity index. It should also be noted that

| Hydrological attribute          | HGM classifications |
|--------------------------------|---------------------|
|                               | Riverine | Depression | Fringe | Flat | Slope | Tidal |
| Dominated by groundwater       | P        | P          | U      | U    | L     | U     |
| Dominated by surface water     | P        | P          | L      | L    | U     | L     |
| Dominated by precipitation     | P        | P          | P      | L    | U     | L     |
| Dominated by vertical hydrodynamics | P    | L          | P      | U    | U     | L     |
| Dominated by lateral hydrodynamics | P    | P          | P      | L    | L     | L     |
| Hydrodynamic period            | S        | S          | S      | S    | A     | D     |

Notes: We used best professional judgment to determine a probability (i.e., L = Likely, P = Possible, and U = Unlikely) or frequency (i.e., D = Daily, S = Seasonal, and A = Annual) category for each HGM classification. Probability (i.e., L = 1, P = 0.5, and U = 0) and frequency (i.e., D = 0, S = 0.5, and A = 1) categories were converted to ordinal ranks, to produce a HGM dissimilarity index that quantified the similarity of wetland types to riverine wetlands.
Bried et al. (2013) did not find a significant effect of vegetative wetland type (i.e., pine barren vernal pond, wet sedge meadow, shrub swamp) on the relationship between wetland condition and mean CoC values. We calculated the wetland type dissimilarity index (Organizational MD), for each application site, \(i\), using:

\[
\text{Organizational MD}_i = \text{HGM}_i + \text{WS}_i
\]

where \(\text{HGM}_i\) is the HGM dissimilarity index for application site, \(i\), and \(\text{WS}_i\) is the water salinity index, given its salinity class, \(t\), for application site, \(i\).

### Results

**Model performance summary.**—While the wetland ecological condition model was suitable for most application sites, the 95% prediction intervals failed to capture the measured mean CoC values at 30% of riverine application sites (Fig. 3a) and 17% of Appalachian Highland application sites (Fig. 3b), compared to 4% of the model validation sites. This was primarily due to under-prediction of the mean CoC values (Table 3). Under-prediction could be problematic if the model is to be used in wetland conservation and restoration decisions; it would suggest these wetlands have a lower level of ecological condition than they actually do.

**Semivariogram of model performance.**—The semivariogram analysis indicated significant spatial structure across model error in the NWCA dataset. Approximately 25% of the variance in model

Table 3. Summary model performance results for validation wetlands in the model development dataset and application wetlands from the National Wetland Condition Assessment (NWCA) dataset.

| Validation datasets | Slope | Adj. \(R^2\) | RSE |
|---------------------|-------|-------------|-----|
| Model validation dataset \((n = 24)\)† | 1.19 | 0.59 | 0.76 |
| NWCA riverine dataset \((n = 194)\)‡ | 0.57 | 0.27 | 0.95 |
| NWCA appalachian dataset \((n = 112)\)‡ | 0.14 | 0.00 | 1.13 |

Notes: These results include the slope, adjusted \(R^2\) (adj. \(R^2\)), and residual squared error (RSE) for the relationship between the measured and modeled mean CoC values. Analyses were conducted in the base package of R (version 3.4.0; R Core Team 2017).

† Model datasets come from Bishel-Machung et al. (1996), Miller and Wardrop (2006), Miller et al. (2006, 2009), Cleveland et al. (2011), Chamberlain and Ingram (2012), Drohan and Brooks (2013), Chamberlain and Brooks (2016), and Riparia, unpublished data.

‡ The National Wetland Condition Assessment (NWCA) dataset can be found at https://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys.

Fig. 3. Plots depict the measured vs. modeled mean coefficient of conservatism (mean CoC) values (unitless) for plot (a): riverine application sites across the contiguous United States (dark gray = 194); and plot (b): Appalachian Highland application sites (light gray, \(n = 112\)). Model validation sites (i.e., dark purple, \(n = 24\)) are displayed for comparison in both plots. Error bars denote the upper and lower 95% prediction intervals for the modeled mean CoC values. The dashed line denotes the 1:1 line. This figure was created using ggplot2 (Wickham 2009) in R (version 3.4.0; R Core Team 2017).
error could be explained by spatial autocorrelation (i.e., sill = 0.62, nugget = 1.88; Fig. 4a). Our interpolated map of model error revealed that the model underestimated the mean CoC values at wetlands along the coastlines of Florida and the Gulf of Mexico (Fig. 4b). This map also revealed lower model performance in the Rocky Mountains (underestimation) and in parts of the Great Lakes region (underestimation), California (underestimation), and the Great Plains (overestimation). However, the semivariogram range was relatively low, estimated at 1151 km. Since most wetlands were further than 1151 km apart, uncertainty in model error predictions (expressed as the standard deviation of model error) was relatively consistent (i.e., ranging between 1.4 and 1.6) across the contiguous United States (Fig. 4c).

**Model application niche analysis: spatial transfer.**—The Intrinsic MD index could not be used to explain variability in model error across riverine application sites (Fig. 5a). One issue was that only 44% of riverine application sites were dissimilar from wetlands used in model development (i.e., Intrinsic MD index > 0). When riverine application sites were dissimilar, it was primarily attributable to the presence of grasslands or the levels of soil pH. Forty-two percent of riverine application sites with an Intrinsic MD index > 0 had higher percentages of grasslands compared to wetlands used in model development (i.e., model development grassland cover: 0–2%). Twenty-nine percent of riverine application sites with an Intrinsic MD index > 0 were outside the soil pH range found in wetlands used in model development (i.e., model development soil pH 4.09–7.47). The majority of these riverine application sites had more basic soil conditions. An additional 12% of riverine application sites with an Intrinsic MD index > 0 were outside the production range of the CoC model.

---

**Fig. 4.** Plot (a) depicts the semivariogram of model error (i.e., modeled mean coefficient of conservatism – measured mean coefficient of conservatism) for wetlands in the National Wetland Condition Assessment dataset (n = 825). We used distance point-pairs between 0 and 4000 km, with equal lag distances of 200 km, and a minimum criterion of 30 point-pairs per bin. Circles depict the mean model errors, and error bars reflect the standard error of the model error across point-pairs in each bin. A powered exponential model was fit to the semivariogram using weighted least squares regression. Parameter estimates, including the nugget, the sill, and the range (i.e., gray dashed lines with labels), were extracted and used in ordinary point kriging. Plot (b) is the interpolation map of model error (gray contour lines delineate breaks of 0.5), and plot (c) is the interpolation map of the standard deviation (SD) of model error (gray contour lines delineate breaks of 0.5) across the contiguous United States. The semivariogram was created in ggplot2 (Wickham 2009), and the final interpolation maps were created in ArcGIS (version 10.5; Environmental Systems Research Institute, Redlands, California, USA).
Fig. 5. Model performance plots. Plot (a) describes model error for riverine application sites (dark gray circles, \( n = 194 \)) as a function of the Intrinsic multivariate dissimilarity (MD) index. Plot (b) describes model error for riverine application sites as a function of the Spatial MD index. Plot (c) describes model error for Appalachian Highland application sites (light gray circles, 112) as a function of the Intrinsic MD index. Plot (d) describes model error for Appalachian Highland application sites as a function of the Organizational (i.e., wetland type) MD index. Model error was expressed as the absolute difference in the mean coefficient of conservatism between measured values and mean modeled values (unitless). The dark purple lines are the same in each plot and describe the mean (solid lines) and standard error (dotted lines) of the model error at model validation wetlands \( (n = 24) \). Model development error was also delineated from model generalizability error using the light purple shading. The black solid lines in each plot depict the significant model fits to the datasets. Parameters, including the model development nugget \( (C_D) \), the application nugget \( (C_A) \), the application sill \( (C_T) \), the contextual range \( (R_C) \), and the contextual maximum \( (R_{\text{max}}) \), are displayed when appropriate. This figure was created using ggplot2 (Wickham 2009) in R (version 3.4.0; R Core Team 2017).
index > 0 differed from wetlands use in model development for both their percentages of grasslands and their levels of soil pH.

We subsequently tested a nugget model to see whether there was a significant contextual interpolation error (i.e., $H_0: C_A = C_D$). The application nugget ($C_A = 0.96 \pm 0.05$) was significantly higher (i.e., Welch two-sample t-test: $t = 3.32$, df = 38.75, $P$-value = 0.001) than the model development nugget (i.e., $C_D = 0.59 \pm 0.10$). Given the lack of model fit across the Intrinsic MD index and the significant interpolation error, we concluded that the riverine application sites’ generalizability error could not be satisfactorily explained by extrapolation of the model to values outside the model’s predictor ranges. Further, the contextual interpolation error could be attributed to a lack of model calibration at riverine application sites, or issues of model structure due to extrinsic contextual variables not considered during model development.

We were able to explain a significant portion of model error (i.e., Model Error$_1 = 0.71 \times \exp\{0.82 \times \text{Spatial MD}\}$, residual squared error [RSE] = 0.66, df = 192) using the Spatial MD index (Fig. 5b). Seventy-four percent of riverine application sites had a Spatial MD index >0. The application nugget ($C_A = 0.71 \pm 0.05$) was no longer significantly different (i.e., Welch two-sample t-test: $t = 1.07$, df = 38.16, $P = 0.291$) from the model development nugget. This lack of contextual interpolation error suggested that the model was generalizable to riverine application sites within or below the elevation (i.e., <638 m above sea level) and mean annual temperature (i.e., <11.6°C) ranges found at sites used in model development. All of the variability in model error that was explained by the Spatial MD index was attributed to the contextual extrapolation error ($C_A = 2.66$, $C_{G_{\text{max}}} = 2.07$). We used the upper standard error ($SE_U$; Fig. 5b upper dark purple dotted line) of the model validation data as a threshold to determine $R_C$ (i.e., $C_G > SE_U$). Using this criterion, $R_C$ was approximately zero. Based on this criterion, the model was not generalizable to riverine application sites above the maximum elevation and mean annual temperature values found at sites used in the model’s development.

**Model application niche analysis: organizational transfer.**—The Intrinsic MD index could also explain variability in model error across Appalachian Highland application sites (Fig. 5c). Thirty-five percent of Appalachian Highland application sites were dissimilar from wetlands used in model development (i.e., Intrinsic MD$_1$ index >0). Similar to riverine application sites, 33% of the Appalachian Highland application sites with an Intrinsic MD$_1$ index >0 had higher percent cover of grasslands than sites used in model development. Fifty-four percent of Appalachian Highland application sites had an Organizational MD index >0 that had soil pH levels outside the range found at sites used in model development. Unlike the more basic soils in riverine application sites, the Appalachian Highland wetlands were more acidic than wetlands used in model development. Again, the application nugget ($C_A = 0.88 \pm 0.10$) was significantly higher (i.e., Welch two-sample t-test: $t = 2.11$, df = 73.51, $P = 0.038$) than the model development nugget (i.e., $C_D = 0.59 \pm 0.10$), representing a significant contextual interpolation error.

We were able to explain a significant portion of model error (i.e., Model Error$_1 = 0.56 + 0.004 \times \exp\{1.61 \times \text{Organizational MD}\}$, RSE = 0.71, df = 109) using the Organizational MD index in the Appalachian Highland region (Fig. 5d). Seventy-one percent of Appalachian Highland application sites had an Organizational MD index >0. The application nugget ($C_A = 0.58 \pm 0.05$) was not significantly different (i.e., Welch two-sample t-test: $t = -0.18$, df = 63.94, $P = 0.859$) from the model development nugget. Not surprisingly, this suggested that the model was generalizable to freshwater riverine wetlands within the Appalachian Highland region. All of the variability in model error that was explained by the Organizational MD index was attributed to contextual extrapolation across wetland types ($C_A = 3.08$, $C_{G_{\text{max}}} = 2.49$). Using the contextual range ($R_C = 2.14$), we concluded that the model was generalizable for 54% of the contextual range assessed ($R_{G_{\text{max}}} = 4.00$). Taken together, we concluded that the model was more generalizable across wetland types within the Appalachian region than across riverine sites within the United States.

**Model application niche analysis: spatial–organizational transfer.**—We then used a model performance heat map to simultaneously assess the spatial and organizational transferability of our model (Fig. 6). As indicated by our model performance plots, the wetland ecological condition...
model was more transferable to model application sites within or below the elevation (i.e., <638 m above sea level) and mean annual temperature (i.e., <11.6°C) ranges found at sites used in model development (Fig. 6a). This held across all but the most dissimilar wetland types (i.e., Organizational MD index >3), where mean model error was relatively high across the full extent of the Spatial MD index. The most dissimilar wetlands included tidal wetlands, slopes, and brackish-to-saline flats. Further, we assessed the generalizability of our model by setting an acceptable mean model error (user-defined), and subsequently counting the number of cells that fell within this criterion. We excluded contextual dissimilarity spaces with fewer than two application sites (i.e., where standard error could not be calculated, shown as gray cells in Fig. 6b) and used a criterion that the model was generalizable in cells with a mean model error ≤1. We concluded that this model was generalizable across 35% of the spatial-organizational contextual dissimilarity space assessed.

**DISCUSSION**

A model’s utility can vary between development and application sites, due to differences in the strength of predictor–response relationships (Whittingham et al. 2007, McAlpine et al. 2008), the direction of the relationships (Murray et al. 2011), and the predictors to be included in the model (Leftwich et al. 2012). With potential limits on resources for data collection and the high stakes associated with transferring models to
new sites, there is value in developing tools that can better inform our model selection decisions. Both economic benefit transfer modelers (Rosenberger and Stanley 2006, Rosenberger and Phipps 2007) and species distribution modelers (Elith et al. 2010, Zurell et al. 2012, Owens et al. 2013, Mesgaran et al. 2014) have advanced our thinking about the importance of contextual dissimilarity when transferring models.

We have expanded on these ideas by constructing a straightforward quantitative methodology to evaluate a model’s application niche across contextual dissimilarity gradients. The power of our methodology lies in its flexibility and ability to leverage preexisting datasets (and/or model results) to inform the model selection process. This methodology can be used for both empirical (e.g., statistical, indices) and process-based models, where the model user chooses what contextual variables to assess and what dissimilarity and validation metrics to use. We discuss some considerations of these choices in the subsections that follow.

**Contextual variables selection**

To perform a model application niche analysis, one must identify the biological, chemical, and/or physical contextual variables to evaluate. As illustrated in our case study, it is useful to organize contextual variables into two major categories, intrinsic and extrinsic. Intrinsic contextual variables define the limits of the model’s predictor hyperspace (also known as covariate space in species distribution modeling)—the space encompassed by all valid combinations of predictors in the model. When evaluating intrinsic contextual variables, two issues should be addressed: extrapolation beyond the extent of the data used to develop the model, and the model’s signal-to-noise ratio (Wiens 1989). While the signal-to-noise ratio can be directly evaluated by looking at the original model data and uncertainty, model performance plots can be used to evaluate the model error associated with the model’s extrapolation across intrinsic dissimilarity gradients.

Intrinsic dissimilarity gradients have been the focus of tools developed in species distribution modeling to assess the risks of model extrapolation (Elith et al. 2010, Zurell et al. 2012, Owens et al. 2013, Mesgaran et al. 2014). Extrapolation outside the predictor hyperspace is less problematic for simple monotonic relationships, in which predictor-response relationships continue in the same trajectory (Miller et al. 2004). For most monotonic relationships, we expect to see a nugget model that intersects the y-axis at the model error associated with $C_D$. Extrapolation is more problematic for non-linear relationships that are truncated (Anderson and Raza 2010) due to environmental limits or sampling bias at sites used in model development (Anderson and Gonzalez 2011). In these cases, we expect the model performance plot to have a non-zero slope function in the extrapolated region of the predictor hyperspace (i.e., $C_A > C_D > C_T$).

As was the case for the wetland ecological condition model, it is also possible for a model to perform poorly at application sites that exist within the model’s predictor hyperspace. This results in a contextual interpolation error (i.e., $C_D < C_A$). Depending on the circumstances, this can indicate a range of issues, including methodological differences between model development and applications sites; the lack of model calibration at the application sites; and/or that the model’s structure might not hold across extrinsic contextual variables. When care has been taken to calibrate the model and minimize methodological errors, the next step is to assess model performance plots across extrinsic contextual variables.

Gradients across extrinsic variables are less likely to be considered or are held constant during model development, potentially limiting model generalizability and transferability. As such, extrinsic variables are commonly cited as the cause of low predictability during model transfers. For example, extrinsic variables (e.g., regional climate, biogeography, and land use) have been cited as causes for low predictability in wetland HGM functional models (Cole et al. 2002), multi-scale models of koala distributions (McAlpine et al. 2008), nitrogen dioxide land use regression models (Poplawski et al. 2009, Allen et al. 2011), and the River Dee-based G-BASH model (Cresser et al. 2006), as well as for poor transfers of model parameters (e.g., Duan et al. 2006, Gan and Burges 2006). Economic benefit transfer modelers have established lists of relevant extrinsic variables, for users to consider during model selection (Spash and Vatn 2006). Few analogous lists (e.g., factors of soil formation; Jenny 1941) currently exist for ecological models. However, geospatial datasets are widely available
and offer a wide range of contextual variables for analyses (e.g., those used in the creation of the Spatial MD index).

**Constructing contextual dissimilarity indices**

Once contextual variables are selected, one must decide what types of dissimilarity metrics are most appropriate for model evaluation. Linear and continuous univariate variables (e.g., temperature, salinity, elevation) are particularly well suited for model performance plots and heat maps. They can be arranged along axes based on their natural gradients and viewed as arithmetic or rank differences. It is also possible to assess multiple contextual variables by collapsing them into additive (e.g., Intrinsic MD index) or weighted (e.g., Spatial MD index) multivariate indices. This is a useful approach when multiple contextual variables are expected to impact model performance within or among application sites (e.g., spatial transfer of the ecological condition model).

Numerous multidimensional metrics (hierarchical clustering, ordination, etc.) are available for the development of contextual dissimilarity indices (McCune and Grace 2002). Species distribution modelers have built indices and complementary tools, including the Multivariate Environmental Similarity Surface tool (Elith et al. 2010), the Mobility-Oriented Parity tool (Owens et al. 2013), and the Extrapolation Detection tool (ExDet; Mesgaran et al. 2014). Methods have also been described for the development of multidimensional metrics that combine mixed categorical and numeric datasets (e.g., Lee and Yun 2003). Each metric has its own nuisances (for discussions, see Owens et al. 2013, Mesgaran et al. 2014), which should be carefully considered on a case-by-case basis before implementation.

Many multidimensional metrics (and validation metrics) are limited by their lack of directionality, which can be important during a model application niche analysis. For example, two application sites that are the same number of units above or below the model’s range for a given contextual variable are given the same dissimilarity value in the ExDet tool (Mesgaran et al. 2014). In our case study, we addressed this issue by separating upper and lower ranges into directional univariate distance metrics before the construction of our Spatial MD index. When developing contextual dissimilarity metrics, one should also consider the distribution of the data used in model development; in most cases, dissimilarity indices are based on a variable’s mean, median, or range. This can be problematic as ecological data are often strongly skewed, and the model error can vary across the predictor hyperspace.

Qualitative variables (many of which are also multivariate) are also commonly used in ecology to explain context, but are challenging to convert into contextual dissimilarity metrics (for an example approach, see Lee and Yun 2003). When the qualitative variable has some hierarchical structure (e.g., the Cowardin wetland classification system, Cowardin et al. 1979; biogeographic regimes, for map, see USEPA 2011), tree distance metrics might be used. However, many of these variables are more complex, including ecological structural or functional classification schemes (e.g., land cover, trophic guild, soil type, phylogenetic, and eco-region). For these variables, best professional judgment practices can be used, including the deconstruction of the classification scheme into component context variables, and subsequently assigning similarity rankings (e.g., Organizational MD index).

**Collection of application site validation data**

The third critical step in this methodology is to collect and synthesize quality validation data from model application sites across the contextual dimension(s) of interest. Validation data can be taken directly from the literature, when a model has already been transferred to application sites, or be calculated, by synthesizing information from databases (e.g., illustrated in our case study) and/or previous studies that contain predictor and response variables. Good candidate models for a model application niche analysis are those that have been applied broadly. Some models that fit this description include CENTURY (Parton et al. 1987; see reference list at https://www.nrel.colostate.edu/projects/century/), DNDC (Li et al. 1992a, b; see reference list at http://www.globaldndc.net/information/publications-i-3.html), PnET (Aber and Federer 1992; see reference list at http://www.pnet.sr.unh.edu/publications.html), SWAT (Engel et al. 1993, Srinivasan and Arnold 1994, Arnold et al. 1998, see model review Arnold et al. 2012), SLAMM (Park et al. 1986; see reference list at http://warrenpinnacle.com/prof/SLAMM/SLAMM_Bibliography.html), and LANDIS-II (Scheller and...
Mladenoff 2004; see reference list at http://www.landis-ii.org/publications). Unfortunately, many of these models are transferred without any formal exploratory niche analysis and without complete validation at the application sites. However, with proper documentation on model calibration and validation, such widely used models are well suited for performing model application niche analyses, with potential long-term benefits to future model users.

It should also be noted that the utility of this methodology can be affected by the quantity and quality of the independent data sources used in the model calibrations and validations (e.g., Rastetter 1996). When methodological errors are large, or model calibration is not performed at application sites, it can hinder the identification of model structural issues during the extrinsic dissimilarity analysis. For example, we did not calibrate the wetland ecological condition model for application sites. While there were clear reasons to doubt the model’s structure for saline and tidal wetland application sites, where other variables (e.g., salinity, tidal hydrology, and storm surges) have already been identified as influential to mean CoC values (Cretini et al. 2012), we could not distinguish calibration issues from model structural issues when evaluating the spatial transfer to riverine sites. Upon further inspection, we found that the riverine dataset was significantly correlated with both soil pH and LDI (i.e., mean CoC = −0.31 × pH − 0.40 × LDI + 6.58506, adj. R² = 0.33, P < 0.001), suggesting that calibration might improve the model’s transferability to riverine sites across the US. Our objective was to demonstrate the applicability of the method we developed, and how it can be used to diagnose confounding factors influencing model performance. If a user is interested in investigating the transferability of our wetland ecological condition model and understanding the ecological processes generating these differences in model performance, we strongly recommend calibrating the model for the chosen context.

When validation results are taken directly from the literature, it can be challenging to synthesize these results. With few exceptions (e.g., Landis and Koch 1977, Araújo et al. 2005), there have been no generally agreed-upon validation criteria in ecological modeling (Rykiel 1996, Gardner and Urban 2003, Stow et al. 2009). Although progress has been made on the classification of validation methods (e.g., Bennett et al. 2013), there is a growing need for consistency in the validation metrics, and transparent approaches for establishing criteria for what constitute acceptable results across the range of tests performed. For example, the popularity of visual validation methods employing pattern recognition (Stow et al. 2009, Bennett et al. 2013, Kubicek et al. 2015) complicates the goal of standardization. Publication bias can further reduce the completeness of a model application niche analysis by emphasizing only those contextual hyperspaces where the model has been transferred successfully.

**Future work**

Improving the transparency and consistency in ecological model transfers is critical for building confidence in the management decisions that are being made based on those transfers. Currently, our methodology is most practical for (1) simple models where predictor and response datasets can be gathered from databases and the literature, and (2) when models have already been validated at application sites across a contextual gradient, and where model users have maintained good documentation on model calibration and validation for both successful and unsuccessful transfers. We do not recommend using this method for models where generalizability is assumed to be weak at the offset. This is frequently found when modeling distributions of common or widespread species (e.g., butterflies, studied by Kharouba et al. 2009). It is difficult to distinguish between suitable and unsuitable habitats for these species due to their phenotypic plasticity, or ability to adapt (Randin et al. 2006, Zanini et al. 2009).

Despite the inherent challenges for some ecological study systems (e.g., widespread species, high plasticity), we see possibilities for broadening the usefulness of the model application niche analysis methodology by (1) creating databases of geospatial dataset for contextual variables that are known to affect the transferability of ecological models, (2) developing contextual dissimilarity metrics and tools that can account for complexities such as qualitative variables and data distributions, (3) increasing the number of validations performed when models are transferred, (4) developing generally agreed-upon...
validation criteria and/or standardizing methods, and (5) building databases for model application validation results (e.g., EcoService Models Library described in Bruins et al. 2016), which include both successful and unsuccessful transfers. This methodology sets the stage for further inquiries into the development of consistent and transparent practices for use during model selection.

ACKNOWLEDGMENTS

The Oak Ridge Institutes for Science and Education (ORISE) postdoctoral fellowship program, the U.S. EPA (USEPA) National Health and Environmental Effects Research Laboratory (NHEERL), and the University of Arkansas provided funding for this research. We are grateful to Gregg Lomnicky and John Wilson of Consolidated Safety Services, Inc. and Miranda Gray and Ming Sheng of the USEPA EcoService Models Library (ESML) Development Team for conversations on model transferability during the construction of the ESML database. We would like to thank Robert Brooks and Denice Wardrop of the Riparia Center at The Pennsylvania State University for access to the Riparia wetland datasets. We would also like to thank Teresa Magee and Karen Blocksom from the USEPA NHEERL-Western Ecology Division (WED) office and Sarah Lehmann and Gregg Serenbetz from the USEPA Office of Water for access to the 2011 NWCA datasets. We are also appreciative to Robert McKane, Susan Yee, and Lauri Green (at USEPA, NHEERL) for their reviews of this manuscript.

LITERATURE CITED

Aber, J. D., and C. A. Federer. 1992. A generalized, lumped-parameter model of photosynthesis, evapotranspiration and net primary production in temperate and boreal forest ecosystems. Oecologia 92:463–474.

Allen, R. W., O. Amram, A. J. Wheeler, and M. Brauer. 2011. The transferability of NO and NO2 land use regression models between cities and pollutants. Atmospheric Environment 45:369–378.

Amano, T. 2012. Unravelling the dynamics of organisms in a changing world using ecological modelling. Ecological Research 27:495–507.

Anderson, R. B., and I. Gonzalez Jr. 2011. Species-specific tuning increases robustness to sampling bias in models of species distributions: an implementation with Maxent. Ecological Modelling 222:2796–2811.

Anderson, R. P., and A. Raza. 2010. The effects of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus Nephelomys) in Venezuela. Journal of Biogeography 37:1378–1393.

Aratujo, M. B., R. G. Pearson, W. Thuiller, and M. Erhard. 2005. Validation of species–climate impact models under climate change. Global Change Biology 11:1504–1513.

Arhonditsis, G. B., and M. T. Brett. 2004. Evaluation of the current state of mechanistic aquatic biogeochemical modeling. Marine Ecology Progress Series 271:13–26.

Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams. 1998. Large-area hydrologic modeling and assessment: Part I. Model development. Journal of the American Water Resources Association 34:73–89.

Arnold, J. G., et al. 2012. SWAT: model use, calibration and validation. Transactions of the American Society of Agricultural and Biological Engineers 55:1491–1508.

Baird, M. E., S. J. Walker, B. B. Wallace, I. T. Webster, and J. S. Parslow. 2003. The use of mechanistic descriptions of algal growth and zooplankton grazing in an estuarine eutrophication model. Estuarine, Coastal and Shelf Science 56:685–695.

Bamford, A. J., A. Monadjem, M. D. Anderson, A. Anthony, W. D. Borello, M. Bridgeford, and I. C. W. Hardy. 2009. Trade-offs between specificity and regional generality in habitat association models: a case study of two species of African vulture. Journal of Applied Ecology 46:852–860.

Bennett, N. D., et al. 2013. Characterising performance of environmental models. Environmental Modelling and Software 40:1–20.

Bishel-Machung, L., R. P. Brooks, S. S. Yates, and K. L. Hoover. 1996. Soil properties of reference wetlands and wetland creation projects in Pennsylvania. Wetlands 16:532–541.

Blumberg, B., and R. L. Cunningham. 1982. An introduction to soils of Pennsylvania. Department of Agricultural and Extension Education, Teacher Education Series 23(4), Publication Distribution Center, The Pennsylvania State University, University Park, Pennsylvania, USA.

Boughton, E. H., P. F. Quintana-Ascencio, P. J. Bohlen, D. G. Jenkins, and R. Pickert. 2010. Land-use and isolation interact to affect wetland plant assemblages. Ecography 33:461–470.

Bried, J. T., S. K. Jog, and J. W. Matthews. 2013. Floristic quality assessment signals human disturbance over natural variability in a wetland system. Ecological Indicators 34:260–267.

Brinson, M. M. 1993. A hydrogeomorphic classification for wetlands. Technical Report WRP-DE-4, U.S. Army Corps of Engineers, Waterways Experiment Station, Vicksburg, Mississippi, USA.
Brown, M. T., and B. Vivas. 2005. Landscape development intensity index. Environmental Monitoring and Assessment 101:289–309.

Bruins, R. J. F., T. Canfield, C. Duke, L. Kapustka, A. Nahlik, and R. Schäfer. 2016. Using ecological production functions to link ecological processes to ecosystem services. Integrated Environmental Assessment and Management 13:52–61.

Burrows, S. N., S. T. Gower, M. K. Clayton, D. S. Mackay, D. E. Ahl, J. M. Norman, and G. Diak. 2002. Application of geostatistics to characterize Leaf Area Index (LAI) from flux tower to landscape scales using a cyclic sampling design. Ecosystems 5:667–679.

Canty, A. and B. Ripley. 2016. Boot: bootstrap R (S-Plus) functions. R package version 1.3-18. https://cran.r-project.org/web/packages/boot/index.html

Chamberlain, S. J., and H. M. Ingram. 2012. Developing coefficients of conservatism to advance floristic quality assessment in the Mid-Atlantic region. Journal of the Torrey Botanical Society 139:416–427.

Chamberlain, S. J., and R. P. Brooks. 2016. Testing a rapid Floristic Quality Index on headwater wetlands in central Pennsylvania, USA. Ecological Indicators 60:1142–1149.

Chamberlain, S. J., and R. P. Brooks. 2016. Testing a rapid Floristic Quality Index on headwater wetlands in central Pennsylvania, USA. Ecological Indicators 60:1142–1149.

Cleveland, M. R., E. A. H. Smithwick, R. P. Brooks, and D. H. Wardrop. 2011. The wetland disturbance index: links with soil and water nitrate concentrations. Wetlands 31:853–863.

Cole, C. A., P. W. Shaffer, and M. E. Kentula. 2002. Comparison of hydrology of wetlands in Pennsylvania and Oregon (USA) as an indicator of transferability of hydrogeomorphic (HGM) functional models between regions. Environmental Management 30:265–278.

Cowardin, L. M., V. Carter, F. C. Golet, and E. T. LaRoe. 1979. Classification of wetlands and deepwater habitats of the United States. U.S. Department of the Interior, Fish and Wildlife Service, Office of Biological Services, Washington, D.C., USA.

Cresser, M. S., N. Ahmed, R. P. Smart, T. Arowolo, L. J. Calver, and P. J. Chapman. 2006. Predicting Gran alkalinity and calcium concentrations in river waters over a national scale using a novel modification to the G-BASH model. Environmental Pollution 143:361–366.

Cretini, K. F., J. M. Visser, K. W. Krauss, and G. D. Steyer. 2012. Development and use of a floristic quality index for coastal Louisiana marshes. Environmental Monitoring and Assessment 184:2389–2403.

Davison, A. C., and D. V. Hinkley. 1997. Bootstrap methods and their applications. Cambridge University Press, Cambridge, UK.

De Marco, P., J. A. F. Diniz-Filho, and L. M. Bini. 2008. Spatial analysis improves species distribution modelling during range expansion. Biology Letters 4:577–580.

Dee, S. M., and C. Ahn. 2012. Soil properties predict plant community development of mitigation wetlands created in the Virginia Piedmont, USA. Environmental Management 49:1022–1036.

Drohan, P., and R. P. Brooks. 2013. Hydric soils across Pennsylvania reference, disturbed, and mitigated wetlands. Pages 129–157 in R. P. Brooks and D. H. Wardrop, editors. Mid-Atlantic freshwater wetlands: advances in wetlands science, management, policy, and practice. Springer, New York, New York, USA.

Duan, Q., et al. 2006. Model parameter estimation experiment (MOPEX): an overview of science strategy and major results from the second and third workshop. Journal of Hydrology 320:3–17.

Elith, J., M. Kearney, and S. Phillips. 2010. The art of modeling range-shifting species. Methods in Ecology and Evolution 1:330–342.

Elith, J., and J. R. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics 40:677–697.

Engel, B. A., R. Srinivasan, J. G. Arnold, C. Rewerts, and S. J. Brown. 1993. Nonpoint-source (NPS) pollution modeling using models integrated with geographic information systems (GIS). Water Science and Technology 28:685–690.

Estes, L. D., A. G. Mwangi, P. R. Reillo, and H. H. Shugart. 2011. Predictive distribution modeling with enhanced remote sensing and multiple validation techniques to support mountain bongo antelope recovery. Animal Conservation 14:521–532.

Falk, W., and K. H. Mellert. 2011. Species distribution models as a tool for forest management planning under climate change: risk evaluation of Abies alba in Bavaria. Journal of Vegetation Science 22: 621–634.

Fick, S. E., and R. J. Hijmans. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology. https://doi.org/10.1002/joc.5086; http://worldclim.org/

Fukuda, S., and K. Hiramatsu. 2008. Prediction ability and sensitivity of artificial intelligence-based habitat preference models for predicting spatial distribution of Japanese medaka (Oryzias latipes). Ecological Modelling 215:301–313.

Gan, T. W., and S. J. Burges. 2006. Assessment of soil-based and calibrated parameters of the Sacramento model and parameter transferability. Journal of Hydrology 320:117–131.

Gardner, R. H., and D. L. Urban. 2003. Model validation and testing: past lessons, present concerns, future prospects. Pages 184–203 in C. D. W. Canham, J. J. Cole, and W. K. Lauenroth, editors.
Models in ecosystem science. Princeton University Press, Princeton, New Jersey, USA.

Goodall, D. W. 1972. Building and testing ecosystem models. Pages 173–194 in J. N. R. Jeffers, editor. Mathematical models in ecology. Blackwell Scientific, Oxford, UK.

Gutrich, J. J., K. J. Taylor, and M. S. Fennessy. 2009. Restoration of vegetation communities of created depressional marshes in Ohio and Colorado (USA): the importance of initial effort for mitigation success. Ecological Engineering 35:351–368.

Heikkinen, R. K., M. Marmion, and M. Luoto. 2012. Does the interpolation accuracy of species distribution models come at the expense of transferability? Ecography 35:276–288.

Hengl, T., et al. 2017. SoilGrids250m: global gridded soil information based on machine learning. PLoS ONE 12:e0169748. ftp://ftp.soilgrids.org/

Homer, C. G., J. A. Dewitz, L. Yang, S. Jin, P. Danielson, G. Xian, J. Coulston, N. D. Herold, J. D. Wickham, and K. Megown. 2015. Completion of the 2011 National Land Cover Database for the conterminous United States: representing a decade of land cover change information. Photogrammetric Engineering and Remote Sensing 81:345–354. https://www.mrlc.gov/nlcd2011.php

Jenny, H. 1941. Factors of soil formation: a system of quantitative pedology. McGraw-Hill, New York, New York, USA.

Jørgensen, S. E. 2008. Overview of the model types available for development of ecological models. Ecological Modelling 215:3–9.

Jørgensen, S. E., and S. N. Nielsen. 2012. Tool boxes for an integrated ecological and environmental management. Ecological Indicators 21:104–109.

Kharouba, H. M., A. C. Algar, and J. T. Kerr. 2009. Historically calibrated predictions of butterfly species’ range shift using global change as a pseudo-experiment. Ecology 90:2213–2222.

Kubicek, A., F. Jopp, B. Breckling, C. Lange, and H. Reuter. 2015. Context-oriented validation of individual-based models in ecology: a hierarchically structured approach to validate quantitative, compositional and quantitative characteristics. Ecological Complexity 22:178–191.

Landis, J. R., and G. G. Koch. 1977. The measurement of observer agreement for categorical data. Biometrics 33:159–174.

Lane, C. R., E. D’Amico, and B. Autrey. 2012. Isolated wetlands of the southeastern United States: abundance and expected condition. Wetlands 32:753–767.

Lee, S.-G., and D.-K. Yun. 2003. Clustering categorical and numeric data: a new procedure using multidimensional scaling. International Journal of Information Technology and Decision Making 2:135–159.

Leftwich, K. N., P. L. Angermeier, and C. A. Dolloff. 2012. Factors influencing behavior and transferability of habitat models for a benthic stream fish. Transactions of the American Fisheries Society 126:725–734.

Legendre, P., and M. J. Fortin. 1989. Spatial pattern and ecological analysis. Vegetatio 80:107–138.

Li, C., S. Frolking, and T. A. Frolking. 1992a. A model of nitrous oxide evolution from soil driven by rainfall events: 1. model structure and sensitivity. Journal of Geophysical Research 97:9759–9776.

Li, C., S. Frolking, and T. A. Frolking. 1992b. A model of nitrous oxide evolution from soil driven by rainfall events: 2. model applications. Journal of Geophysical Research 97:9777–9783.

Liu, S., R. Portela, N. Rao, A. Ghermandi, and X. Wang. 2011. Environmental benefit transfers of ecosystem service valuation. Pages 55–77 in E. Wolanksi and D. S. McLusky, editors. Treatise on estuarine and coastal science. Volume 12. Elsevier, London, UK.

Lopez, R. D., and M. S. Fennessy. 2002. Testing the floristic quality assessment index as an indicator of wetland condition. Ecological Applications 12:487–497.

Matthews, J. W., G. Spyreas, and A. G. Endress. 2009. Trajectories of vegetation-based indicators used to assess wetland restoration. Ecological Applications 19:2093–2107.

McAlpine, C. A., J. R. Rhodes, M. E. Bowen, D. Lunney, J. G. Callaghan, D. L. Mitchell, and H. P. Possingham. 2008. Can multiscale models of species’ distribution be generalized from region to region? A case study of the koala. Journal of Applied Ecology 45:558–567.

McCune, B., and J. B. Grace. 2002. Analysis of ecological communities. MJM Software, Gleneden Beach, Oregon, USA.

McDonald, C. P., and N. R. Urban. 2010. Using a model selection criterion to identify appropriate complexity in aquatic biogeochemical models. Ecological Modelling 221:428–432.

McGarity, T. O., and W. E. Wagner. 2003. Legal aspects of the regulatory use of environmental modeling. Environmental Law Reporter: News and Analysis 33:10751–10774.

McKay, L., T. Bondelid, T. Dewald, J. Johnston, R. Moore, and A. Rea. 2012. NHDPlus Version 2: user guide. CM130105CT0027. Office of Water, U.S. Environmental Protection Agency, Washington, D.C., USA. https://www.epa.gov/waterdata/nhdplus-national-hydrography-dataset-plus

Mesgaran, M. B., R. D. Cousins, and B. L. Webber. 2014. Here be dragons: a tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. Diversity and Distributions 20:1147–1159.
Meynard, C. N., and J. F. Quinn. 2007. Predicting species distributions: a critical comparison of the most common statistical models using artificial species. Journal of Biogeography 34:1455–1469.

Miller, E. W. 1995. Chapter 5: soil resources. Pages 67–73 in E. Miller, editor. Geography of Pennsylvania. The Pennsylvania State University Press, University Park, Pennsylvania, USA.

Miller, J. R., M. G. Turner, E. A. H. Smithwick, C. L. Dent, and E. H. Stanley. 2004. Spatial extrapolation: the science of predicting ecological patterns and processes. BioScience 54:310–320.

Miller, S. J., and D. H. Wardrop. 2006. Adapting the floristic quality assessment index to indicate anthropogenic disturbance in central Pennsylvania wetlands. Ecological Indicators 6:313–326.

Miller, S. J., D. H. Wardrop, and M. M. Harlan. 2009. Riparian plant communities as predictors of instream condition: a case study in the Upper Penn Creek watershed, Pennsylvania. Bartonia 64:19–35.

Miller, S. J., D. H. Wardrop, W. M. Mahaney, and R. P. Brooks. 2006. A plant-based index of biological integrity (IBI) for headwater wetlands in central Pennsylvania. Ecological Indicators 6:290–312.

Murray, J. V., S. Low Choy, C. A. McAlpine, H. P. Possingham, and A. W. Goldizen. 2011. Evaluating model transferability for a threatened species to adjacent areas: implications for rock-wallaby conservation. Austral Ecology 36:76–89.

Nichols, S. A. 1999. Floristic quality assessment of Wisconsin lake plant communities with example applications. Journal of Lake and Reservoir Management 15:133–141.

NRC (National Research Council). 2007. Models in environmental regulatory decision making. The National Academies Press, Washington, D.C., USA.

Owens, H. L., et al. 2013. Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. Ecological Modeling 263:10–18.

Pabian, S. E., and M. C. Brittingham. 2007. Terrestrial liming benefits birds in an acidified forest in the northeast. Ecological Applications 17:2184–2194.

Park, R. A., T. V. Armentano, and C. L. Cloonan. 1986. Predicting the effects of sea level rise on coastal wetlands. Pages 129–152 in J. G. Titus, editor. Effects of changes in stratospheric ozone and global climate, Volume 4: sea level rise. U.S. Environmental Protection Agency, Washington, D.C., USA.

Parton, W. J., D. S. Schimel, C. V. Cole, and D. S. Ojima. 1987. Analysis of factors controlling soil organic matter levels in Great Plains grasslands. Soil Science Society of America Journal 51:1173–1179.

Pease, A. A., J. M. Taylor, K. O. Winemiller, and R. S. King. 2015. Ecoregional, catchment, and reach-scale environmental factors shape functional-trait structure of stream fish assemblages. Hydrobiologia 753:265–283.

Peterson-Smith, J., D. H. Wardrop, C. A. Cole, C. P. Cirmo, and R. P. Brooks. 2009. Hydrogeomorphology, environment, and vegetation associations across a latitudinal gradient in highland wetlands of the northeastern USA. Plant Ecology 203:155–172.

Pont, D., B. Hugueny, U. Beier, D. Goffaux, A. Melcher, R. Noble, C. Rogers, N. Roset, and S. Schmutz. 2006. Assessing river biotic condition at a continental scale: a European approach using functional metrics and fish assemblages. Journal of Applied Ecology 43:70–80.

Poplawski, K., et al. 2009. Intercity transferability of land use regression models for estimating ambient concentrations of nitrogen dioxide. Journal of Exposure Science and Environmental Epidemiology 19:107–117.

R Core Team. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Randin, C. F., T. Dinmböck, S. Dullinger, N. E. Zimmermann, M. Zappa, and A. Guisan. 2006. Are niche-based species distribution models transferable in space? Journal of Biogeography 33:1689–1703.

Rastetter, E. B. 1996. Validating models of ecosystem response to global change: How can we best assess models of long-term global change? BioScience 46:190–198.

Reiss, K. C. 2006. Florida wetland condition index for depressional forested wetlands. Ecological Indicators 6:337–352.

Ribeiro Jr., P. J., and P. J. Diggle. 2001. geoR: a package for geostatistical analysis. R-News 1:15–18. https://cran.r-project.org/web/packages/geoR/index.html

Rosenberger, R. S., and T. T. Phipps. 2007. Correspondence and convergence in benefit transfer accuracy: a meta-analytic review of the literature. Pages 22–43 in S. Navrud and R. Ready, editors. Environmental values transfer: issues and methods. Kluwer Academic, Dordrecht, The Netherlands.

Rosenberger, R. S., and T. D. Stanley. 2006. Measurement, generalization, and publication: sources of error in benefit transfers and their management. Ecological Economics 60:372–378.

Rykiel Jr., E. J. 1996. Testing ecological models: the meaning of validation. Ecological Modelling 90:229–244.

Scheller, R. M., and D. J. Mladenoff. 2004. A forest growth and biomass module for a landscape simulation model, LANDIS: design, validation, and application. Ecological Modelling 180:211–229.

Simon, S. A., T. K. Collins, G. L. Kauffman, W. H. McNab, and C. L. Ulrey. 2005. Ecological zones in...
the southern Appalachians: first approximation. Research Paper SRS-41, U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, North Carolina, USA.

Smith, D. B., W. F. Cannon, L. G. Woodruff, F. Solano, J. E. Kilburn, and D. L. Fey. 2013. Geochemical and mineralogical data for soils of the conterminous United States. U.S. Geological Survey Data Series 801. U.S. Department of the Interior, U.S. Geological Survey, Reston, Virginia, USA. https://pubs.usgs.gov/ds/801/

Spash, C. L., and A. Vatn. 2006. Transferring environmental value estimates: issues and alternatives. Ecological Economics 60:379–388.

Srinivasan, R., and J. G. Arnold. 1994. Integration of a basin-scale water quality model with GIS. Water Resource Bulletin 30:453–462.

Stow, C. A., J. Jolliff, D. J. McGillicuddy, S. C. Doney, J. I. Allen, M. A. Friedrichs, K. A. Rose, and P. Wallhead. 2009. Skill assessment for coupled biological/physical models of marine systems. Journal of Marine Systems 76:4–15.

Swink, F., and G. Wilhelm. 1994. Plants of the Chicago region. Fourth Edition. Indiana Academy of Science, Indianapolis, Indiana, USA.

Tobler, W. 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography 46:234–240.

USEPA (U.S. Environmental Protection Agency). 2011. Level III ecoregions of the continental United States (revision of Omernik, 1987). U.S. Environmental Protection Agency, National Health and Environmental Effects Laboratory-Western Ecology Division, Corvallis, Oregon, USA. https://pubs.usgs.gov/tm/04/c03/virtual_CD/useco.pdf

USEPA (U.S. Environmental Protection Agency). 2016. National Wetland Condition Assessment: 2011 Technical Report. EPA-843-R-15-006. U.S. Environmental Protection Agency, Washington, D.C., USA. https://www.epa.gov/national-aquatic-resource-surveys/national-wetland-condition-assessment-2011-results

USEPA ATFERM (U.S. Environmental Protection Agency, Agency Task Force on Environmental Regulatory Modeling). 1994. Agency guidance for conducting external peer review of environmental regulatory modeling. EPA/100B-94-001. U.S. Environmental Protection Agency, Washington, D.C., USA.

USGS (U.S. Geological Survey). 2012. 100-meter resolution elevation of the conterminous United States. Raster digital data. National Atlas of the United States, Rolla, Missouri, USA. https://nationalmap.gov/small_scale/ml/d/elev100.html

Vanreusel, W., D. Maes, and H. Van Dyck. 2007. Transferability of species distribution models: a functional habitat approach for two regionally threatened butterflies. Conservation Biology 21:201–212.

Venables, W. N., and B. D. Ripley. 2002. Modern applied statistics with S. Fourth Edition. Springer, New York, New York, USA.

Wenger, S. J., and J. D. Olden. 2012. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods in ecology and Evolution 3:260–267.

Whittingham, M. J., J. R. Krebs, R. D. Swetnam, J. A. Vickery, J. D. Wilson, and R. P. Freckleton. 2007. Should conservation strategies consider spatial generality? Farmbird birds show regional not national patterns of habitat association. Ecology Letters 10:25–35.

Wickham, H. 2009. ggplot2: elegant graphics for data analysis. Springer-Verlag, New York, New York, USA. https://cran.r-project.org/web/packages/ggplot2/index.html

Wiens, J. A. 1989. Spatial scaling in ecology. Functional Ecology 3:385–397.

Wilhelm, G., and D. Ladd. 1988. Natural area assessment in the Chicago region. Pages 361–375 in Transactions 53rd North American Wildlife and Natural Resources Conference, Louisville, Kentucky. Wildlife Management Institute, Washington, D.C., USA.

Wu, J., and H. Li. 2006. Chapter 1: concepts of scale and scaling. Pages 3–15 in J. Wu, K. B. Jones, H. Li, and O. Loucks, editors. Scaling and uncertainty analysis in ecology: methods and applications. Springer, Dordrecht, The Netherlands.

Zanini, F., J. Pellet, and B. R. Schmidt. 2009. The transferability of distribution models across regions: an amphibian case study. Diversity and Distributions 15:469–480.

Zurell, D., J. Elith, and B. Schröder. 2012. Predicting to new environments: tools for visualizing model behaviour and impacts on mapped distributions. Diversity and Distributions 18:628–634.

Supporting Information

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecs2.1974/full