Balancing transferability and complexity of species distribution models for rare species conservation

Nolan A. Helmstetter | Courtney J. Conway | Bryan S. Stevens | Amanda R. Goldberg

Abstract

Aim: Species distribution models (SDMs) are valuable for rare species conservation and are commonly used to extrapolate predictions of habitat suitability geographically to regions where species occurrence is unknown (i.e., transferability). Spatially structured cross-validation can be used to infer transferability, yet, few studies have evaluated how delineation of cross-validation folds affects model complexity and predictions. We developed SDMs using multiple cross-validation approaches to understand the implications for predicting habitat suitability for northern Idaho ground squirrels, a rare, federally threatened species that has been extensively surveyed in regions where known populations occur, resulting in >8000 presence locations.

Location: Idaho, USA.

Methods: We delineated cross-validation folds by mimicking the manner in which predictions would be geographically extrapolated or by using existing dispersal barriers. We varied the distance between, number, and directionality of folds. We conducted a grid search on statistical regularization parameters to optimize model complexity, covering a range of values exceeding that typically implemented. For each cross-validation approach, we selected optimal regularization and model complexity based on out-of-sample predictive ability.

Results: Delineation of cross-validation folds substantially affected resulting model complexity and extrapolated predictions. All cross-validation approaches resulted in models with apparently high out-of-sample predictive ability, yet optimal model complexity varied substantially among the approaches. Regularization demonstrated a noisy relationship between model complexity and prediction, where local optima in predictive performance were common at small values.

Main conclusion: Subtle modelling decisions can have large consequences for predictions of habitat suitability and transferability of SDMs. When transferability is the goal, cross-validation approaches should be considered carefully and mimic the manner in which spatial extrapolation will occur, else overly complex models with inflated assessments of predictive accuracy may result. Further, spatially structured cross-validation may not guard against over-parameterization, and assessing a broader
Species distribution models (SDMs) are increasingly popular and powerful tools for applied conservation (Elith & Leathwick, 2009; Guisan & Thuiller, 2005; Rushton et al., 2004). SDMs relate presence-background, presence-absence, or detection-non-detection data (Guiller-Arroita et al., 2015) to environmental attributes that are used to predict the distributions of organisms in space and time (Araújo et al., 2005; Elith & Franklin, 2013; Elith et al., 2006; Ferrier et al., 2002), and these predictions are used to map habitat suitability and inform conservation of rare species in locations where presence/absence information is unknown or imperfect (Araújo et al., 2005; Ferrier et al., 2002; Guisan et al., 2013). Evaluating the ability of SDMs to predict to new areas or time periods (hereafter transferability) is therefore critical if model predictions are to be extrapolated with confidence (Wenger & Olden, 2012; Yates et al., 2018). Model transferability is often limited by over-parameterization, where complex models that over-fit data from one region result in poor prediction in other regions (Heikkinen et al., 2012). Consequently, limiting model complexity is a necessary component of building SDMs when predictive transferability is the primary goal (Araújo et al., 2005; Radosavljevic & Anderson, 2014; Randin et al., 2006; Wenger & Olden, 2012).

Empirical evaluation of SDM predictive performance is best done with out-of-sample data that are not used to calibrate models and estimate parameters (Hooten & Hobbs, 2015; Olden et al., 2002). Alternatively, computationally intensive cross-validation techniques can be used to approximate a model’s predictive ability with new data when true out-of-sample data are not available (Fielding & Bell, 1997; Hijmans, 2012; Roberts et al., 2017). This is commonly done using k-fold cross-validation, where data are randomly partitioned into k folds and data from k - 1 fold(s) are used to fit the model, whereas data from the remaining fold are used to test predictive ability. This process is then iterated over all data folds, leaving each fold out to assess predictive ability in turn, where the overall predictive score is averaged over the k iterations. This approach is common, yet randomly selecting data points to include in cross-validation folds does not force the model to extrapolate into new environmental space or new spatio-temporal dependence structures and therefore may not accurately approximate a model’s transferability in space and time (Roberts et al., 2017). This can produce SDMs that are overfit to attributes of the locations where field data were collected, with excessive numbers of environmental predictors, overly complex relationships between occurrence data and those variables, and artificially inflated assessments of predictive accuracy (Olden & Jackson, 2000; Peterson et al., 2011; Roberts et al., 2017).

Structured cross-validation methods that partition data non-randomly into spatially or temporally distinct folds can more accurately assess transferability of SDMs (Araújo et al., 2005; Radosavljevic & Anderson, 2014; Roberts et al., 2017). Structuring cross-validation folds in space allows investigators to optimize model complexity for transferability while avoiding the overfitting of species–environment relationships (Anderson & Gonzalez, 2011; Gerber & Northrup, 2019). Roberts et al. (2017) discussed numerous aspects of cross-validation design that affect reliability of predictive accuracy estimates, yet the delineation of spatial folds is not always intuitive and there are no one-size-fits-all approaches for spatially structuring cross-validation when the goal is to extrapolate predictions to other portions of the species’ range. The number of cross-validation folds, their geographic boundaries, the directionality of partitions and the environmental variation within each fold may affect assessments of model transferability (Roberts et al., 2017; Wenger & Olden, 2012), but the best options may not be obvious to investigators a priori. Moreover, while Roberts et al. (2017) discussed cross-validation designs to reliably estimate predictive accuracy of SDMs, they left unaddressed the question of how different cross-validation approaches may affect model selection. Hence, the sensitivities of SDM structures (e.g., covariates included), predicted habitat maps and conservation implications to different applications of non-random cross-validation are poorly understood for most applied conservation problems.

While structured cross-validation approaches can be used to assess transferability of SDMs for conservation purposes, statistical regularization can also be used to limit model complexity and achieve better out-of-sample prediction (Elith et al., 2011; Hooten & Hobbs, 2015). Common approaches to statistical regularization (e.g., lasso and ridge regression; Hooten & Hobbs, 2015) include one or more parameter(s) that reduce (i.e., penalize) model complexity by shrinking individual regression coefficients for uninformative variables towards zero, thereby reducing their contributions towards prediction and resulting in a simpler model (Bickel et al., 2006). Regularization parameters are often selected arbitrarily or fixed in advance by investigators (e.g., using default settings of Maxent; Phillips et al., 2006), yet statistical regularization can also be implemented concurrently with optimizing the predictive performance of SDMs. In this case, regularization parameter values are selected to optimize predictive performance of resulting models, given the observed data (Gerber & Northrup, 2019; Hastie et al., 2009). This
requires a grid search over a range of plausible values for the regularization parameter(s) (Hooten & Hobbs, 2015), where the predictive performance is assessed for each value and the optimal amount of regularization (and therefore strength of covariate shrinkage) is selected in order to achieve optimal out-of-sample prediction. Thus, selection of a regularization penalty becomes an integral part of the process of predictive model selection (Gerber et al., 2015; Stevens & Conway, 2019). This approach is intuitive and powerful, yet the range and resolution of regularization values must be defined by the investigator. In practice, the range of regularization values considered is often limited (Anderson & Gonzalez, 2011; Radosavljevic & Anderson, 2014), whereas the implications of selecting a sub-optimal regularization parameter (and therefore a sub-optimally predictive SDM) for rare species conservation are typically unknown.

We address the challenge of developing SDMs that are optimized for spatial transferability to predict habitat suitability and aid conservation and recovery planning for species of conservation concern. We developed SDMs using a variety of spatially structured cross-validation approaches that varied in the distance between, number and directionality of spatial folds in order to understand the sensitivity of the resulting SDM and the implications for predicting habitat suitability to aid conservation decision-making. We also used statistical regularization to optimize predictive ability under each cross-validation approach, considered a broad range of values for regularization parameters and assessed the implications of selecting a sub-optimal regularization parameter (i.e., a sub-optimal level of model complexity) for predicting habitat suitability of a rare animal. We conducted this analysis with the goal of developing a predictive, spatially transferable SDM to map range-wide habitat suitability and aid ongoing recovery efforts for northern Idaho ground squirrels (*Urocitellus brunneus*), a small mammal that is federally threatened within the United States and whose current distribution is patchy and irregular within its historical range (Goldberg, Conway, Evans Mack, et al., 2020). Large portions of the northern Idaho ground squirrel range are labeled as unoccupied, but small and previously undiscovered populations are frequently found. Thus, a SDM that accurately predicts habitat suitability in unsampled regions would benefit recovery efforts for this species.

## 2 | METHODS

### 2.1 | Study species

Northern Idaho ground squirrels are only known to occur in two counties (Adams and Valley counties) in west-central Idaho (USFWS, 2003) and are believed to persist only in a small proportion of their historical range (Figure 1). This rare ground squirrel lives in isolated mountain meadows surrounded by coniferous forest at elevations of 1150-m to 1550-m (Yensen & Sherman, 1997). A century of anthropogenic fire suppression has resulted in forest encroachment, which is hypothesized to eliminate habitat and be the primary threat to the species (Burak et al., 2018; Yensen & Sherman, 1997). Northern Idaho ground squirrels use burrows extensively during all seasons and soil properties are thought to play an important role in habitat selection (Goldberg, Conway, Evans Mack, et al., 2020; Yensen, 1991; Yensen et al., 1991). Individuals can hibernate for up to 8 months (Yensen, 1991; Goldberg and Conway, 2021) and south-facing slopes that receive more solar radiation have a higher probability of presence (Goldberg, Conway, Evans Mack, et al., 2020). South-facing slopes have more solar radiation that increases the amount of available foraging time during the summer (i.e., the squirrel’s active season) by reducing snowpack earlier in the year (USFWS, 2003). Conservation efforts to restore northern Idaho ground squirrel habitat, demographic studies at known colonies and surveys to locate new populations are currently being implemented (Allison et al., 2019; Burak et al., 2018; Goldberg, Conway, Evans Mack, et al., 2020; Goldberg, Conway, Tank, et al., 2020; Wagner & Evans Mack, 2020).

### 2.2 | Presence-background data

We obtained northern Idaho ground squirrel presence locations (GPS locations) from surveys and live-trapping conducted from 2001–2018 by Idaho Department of Fish and Game (Wagner & Evans Mack, 2020) and University of Idaho (Allison et al., 2019; Goldberg, 2018). Only active season presence locations were used because suitable hibernacula habitat can be substantially different (Goldberg, Conway, Evans Mack, et al., 2020). When >1 presence point was located within a given 30-m raster pixel (the spatial resolution of environmental predictor variables), we retained that pixel as a...
single occurrence to avoid pseudo-replication. This resulted in 8878 presence locations that we used to build SDMs, and these points are representative of all known extant populations of northern Idaho ground squirrels.

The extent from which background locations is selected and the number of background locations can influence modelling results (Barbet-Massin et al., 2012; VanDerWal et al., 2009) and defines the range of environmental conditions that are considered available to a species. We selected background locations from within 1.2-km buffers surrounding presence locations, where 1.2-km is the longest recorded dispersal distance for northern Idaho ground squirrels (Sherman & Runge, 2002). We used a 1:10 ratio (Barbet-Massin et al., 2012) of presence to background locations to characterize environmental conditions at presence and available locations.

2.3 Environmental variables

We considered 11 environmental variables believed a priori to be related to the distribution of northern Idaho ground squirrels (Table 1). We included topographic predictor variables (i.e., elevation and topographic solar radiation index) because these variables are important predictors of northern Idaho ground squirrel habitat (see above). We included slope as a predictor because shallow slopes are associated with active season habitat (Goldberg, Conway, Evans Mack, et al., 2020). Further, while active season habitat is associated with open meadows (i.e., reduced canopy cover) and frequent low-severity fires that reduced conifer encroachment (Yensen & Sherman, 1997), areas with increased canopy cover are associated with hibernacula locations, and therefore, we hypothesized a negative relationship would exist with distance from >25% canopy cover (Goldberg, Conway, Evans Mack, et al., 2020). We included soil composition (e.g., per cent clay) and depth to bedrock in our analyses because northern Idaho ground squirrels utilize burrows extensively throughout the year and particular soil textures are associated with burrowing mammals (Laundré & Reynolds, 1993). Lastly, we included existing vegetation (LANDFIRE, 2016) to differentiate current suitable habitat (e.g., grassland) from regions where the species could not or are unlikely to persist (e.g., developed roads and agriculture or conifer and hardwood forests). All environmental variables were measured as raster data (or converted to raster data; e.g., soil properties) with a 30-m spatial resolution. Because the biological effects of environmental variables can be sensitive to the scales of observation for covariate data (and can change nonlinearly among scales; McGarigal et al., 2016; Wiens, 1989), we considered four spatial extents (30-m, 90-m, 150-m and 300-m) for each variable (excluding the distance-based and ordinal variable: distance to >25% canopy cover and fire interval). To calculate covariate values for each variable at 90-m, 150-m and 300-m extents, we averaged the cell values of neighbouring pixels contained within the buffered area for each spatial extent. For the categorical variable (e.g., LANDFIRE, 2016), we reduced the number of categories by grouping similar features (e.g., developed and developed roads) and calculated the proportion of area made up by each category at each spatial extent, resulting in 7 predictor variables that were either binary (where the binary is assumed to be 0% or 100% cover for each cell) at a 30-m spatial extent or continuous at an increased spatial extent (i.e., 90-m, 150-m and 300-m). To select the optimal scale for each covariate, we ran univariate linear models with the regularization parameters set to zero (i.e., no penalty) and selected the best scale for each variable based on its out-of-sample predictive ability (using average test AUC as described below). Once individual variables were scale-optimized, we combined the 11 variables into multi-scale SDMs and optimized their complexity for out-of-sample prediction. We replicated this process under each cross-validation approach (described below), resulting in multi-scale predictive SDMs (McGarigal et al., 2016) that could vary in their structure (both covariates and scales) among the different cross-validation approaches used to assess spatial transferability.

### Table 1: Environmental variables used to develop species distribution models for northern Idaho ground squirrels and the data source

| Categories           | Variables                                      | Source                                             |
|----------------------|------------------------------------------------|----------------------------------------------------|
| Soil properties      | Percent clay                                   | Jorgensen (1995), Soil Survey Staff (2015a,2015b,2015c,2015d,2015e, 2018) |
|                      | Percent sand                                   |                                                   |
|                      | Percent silt                                   |                                                   |
|                      | Depth to bedrock (cm)                          |                                                   |
| Topography           | Elevation (m)                                  | USGS (1999)                                        |
|                      | Slope                                           |                                                   |
|                      | Topographic solar radiation index              |                                                   |
| Existing vegetation  | Agriculture, developed, developed roads        | LANDFIRE (2016)                                   |
|                      | Barren, sparsely vegetated, snow-ice           |                                                   |
|                      | Conifer, conifer-hardwood, hardwood            |                                                   |
|                      | Exotic herbaceous                              |                                                   |
|                      | Grassland                                      |                                                   |
|                      | Riparian                                       |                                                   |
|                      | Shrubland                                      |                                                   |
| Fire interval        | Mean fire return interval                      | LANDFIRE (2016b)                                  |
| Canopy cover         | Percent canopy cover                           | NLCD (2016)                                       |
|                      | Distance to >25% canopy cover                  |                                                   |
2.4 | Species distribution models

We used r (R Core Team, 2017) packages “maxnet” (Phillips et al., 2017) and “ENMeval” (Muscarella et al., 2014) to develop SDMs. Package “maxnet” is the open source version of Maxent, a popular program used to develop SDMs from presence-background data. Package “maxnet” compares the distribution of environmental variables measured at background locations to the distribution of environmental variables at known presence locations (Phillips et al., 2017; Phillips & Dudík, 2008), resulting in a SDM that models relative probability of use analogously to a resource selection function (Lele et al., 2013). Model fitting is subject to user-specified constraints that determine the shape of the possible relationships between hypothesized covariates and the relative probability of use (termed feature classes), which are selected a priori by the investigator. Moreover, “maxnet” uses a built-in lasso regression approach to regulate model complexity (Elith et al., 2011), optimizing a penalized likelihood that includes a regularization parameter to shrink regression coefficients towards zero for uninformative variables. The larger the regularization parameter, the greater the coefficient shrinkage, resulting in simpler models with more regression coefficients whose values are shrunk towards zero (thus reducing their contribution towards prediction). Each feature class has a default regularization parameter value (Phillips & Dudík, 2008) that specifies the amount of shrinkage. However, the magnitude of the penalty term can be adjusted manually via a regularization multiplier that is multiplied by the default parameter, where the product determines the overall magnitude of shrinkage (Anderson & Gonzalez, 2011). The default setting for the regularization multiplier is one, effectively maintaining the default regularization parameters for each feature class. Despite evidence that optimizing regularization parameters and feature class selection can aid in identifying a model with an appropriate level of complexity (Anderson & Gonzalez, 2011; Radosavljevic & Anderson, 2014; Shcheglovitova & Anderson, 2013), default settings are often maintained by most users (Morales et al., 2017).

We evaluated model transferability by using three non-random, spatially structured cross-validation approaches (hereafter 2-fold, 3-fold and 5-fold cross-validation). We delineated spatially structured cross-validation folds to either resemble the manner in which we intended to spatially extrapolate predictions (i.e., latitudinally north or south of known presence locations) or by using natural biological units (i.e., dispersal barriers; Barbosa et al., 2021). The 2-fold and 5-fold cross-validation approaches (Figure 2a,c) delineated presence-background data latitudinally and thus mimicked the manner in which the predictions were intended to be extrapolated to map range-wide habitat suitability. However, the 2-fold cross-validation approach provided a more conservative evaluation of transferability (i.e., only two geographically distinct folds; Wenger & Olden, 2012) whereas the 5-fold cross-validation approach was less conservative and used more folds with less geographic separation and an equal number of presence and background locations within each of the northern four folds. The 3-fold cross-validation approach (Figure 2b) consisted of natural biological units divided into folds based on likely barriers (i.e., distance or geographic) to dispersal. Specifically, the two northern folds were delineated based on a high-elevation mountain range located between them that is believed to limit dispersal, and dispersal between the two northern folds and the southern fold is limited because they are 70-km apart (Barbosa et al., 2021). We also developed a model using a traditional random 5-fold cross-validation approach for comparison. We randomly selected a 1:10 ratio of presence to background locations from the buffered presence locations (described above) in a manner that allowed us to maintain a 1:10 ratio of presence to background locations within each fold for all cross-validation approaches and use the same background locations for all model comparisons (see Figure S3 and S3.1 for a detailed explanation).

We ran a grid search on the regularization multiplier for each cross-validation approach to optimize model complexity for out-of-sample prediction, from 0.1 to 50.0 using 0.1 increments. We extended the grid search for the 2- and 3-fold cross-validations to 80.0 and 70.0, respectively, because no clear reduction in out-of-sample prediction occurred (Figure 3). We constrained models to include combinations of only three possible covariate relationships (linear, quadratic and hinge). We selected an optimally predictive model for each cross-validation based on each models’ out-of-sample predictive ability, using average test AUC as the predictive score. Our objective was to develop a model to predict relative habitat suitability for northern Idaho ground squirrels, which is the primary capability of presence-background data (i.e., such data cannot identify absolute probabilities of occurrence; Guillera-Arroita et al., 2015). AUC is a direct measure of a model’s ability to
discriminate the relative probability of a point being a presence (1) or background location (0) (i.e., relative probability of use or relative occurrence rate), and thus we believe AUC was useful as a measure of predictive ability because it directly measured what our SDM was intended to do in practice.

2.5 | Spatially explicit predictions

We mapped optimal model predictions within a buffered area surrounding the historical range of northern Idaho ground squirrels to assess sensitivity of modelling outputs most directly relevant for conservation. We expanded the extent of predictions using U.S. Geological Survey 10-digit hydrologic unit watershed boundaries (USGS, 2013) because the existing historical range boundaries are approximations and northern Idaho ground squirrels may occur outside the published range. We projected our model within all watersheds whose boundaries overlapped the historical range, those that contained any northern Idaho ground squirrel detections and those whose boundaries were <700 m from the historical range or <1400 m from known squirrel presence locations. Within this spatial extent, we also performed a multivariate environmental similarity surface (MESS) analysis (Elith et al., 2010) to identify regions in geographic space where our predictions were extrapolating in environmental space (i.e., the values of environmental variables were outside the range of values used to build the model), and removed those areas from our maps. We retained the "raw" (i.e., exponential) predicted values of relative occurrence rate over our study region rather than using the "maxnet" logistic or complementary-log-log ("cloglog") link functions because they require assumptions about prevalence (and therefore absolute probability of occurrence) that are typically unwarranted when developing SDMs using presence-background data (i.e., presence-background data cannot estimate prevalence and therefore cannot scale predictions to absolute probabilities; Guillera-Arroita et al., 2015). However, to make comparisons and interpretations easier among models, we rescaled predicted values of relative occurrence rate to range from 0 to 1 by dividing all predicted values by the largest observed predicted value for a given model. Lastly, we compared MESS maps and ranges of spatial autocorrelation of predictor variables among the various spatial cross-validation folds (Roberts et al., 2017) to better understand how and when models fit to data from individual folds were extrapolating in environmental and geographic space.

3 | RESULTS

The optimal covariate relationships (feature classes, variables included and scales of effect) and regularization multiplier varied among top models for each cross-validation approach (Figure 3), demonstrating the sensitivity of predictive model selection to the design of cross-validation assessments. All of the optimally predictive models suggested an apparently high out-of-sample predictive ability (>0.80 average test AUC), and most spatially structured cross-validation approaches resulted in test AUC response curves (plotted as a function of the regularization multiplier) with noisy surfaces that included local optima (i.e., peaks not representative of true optimal out-of-sample predictive ability) interspersed with troughs (Figure 3).
The optimal model for 2-fold cross-validation included six non-zero model coefficients and had a regularization multiplier of 58.2, far exceeding the default value of 1.0 in Maxent that is commonly used by investigators. In contrast, the 3-fold, 5-fold and random 5-fold cross-validation approaches resulted in much more complex models, with 31, 104 and 255 non-zero coefficients, and regularization multipliers of 0.1, 1 and 0.1, respectively. Linear, quadratic and hinge covariate transformations (i.e., feature classes) were included in the optimal 5-fold and the random 5-fold cross-validation approaches (Table 2), resulting in models with complex nonlinear relationships between predictors and relative occurrence rate (Figure 4; Figures S1.1 and S1.2). Only linear and quadratic relationships were included in the optimally predictive 3-fold and 2-fold cross-validation approaches (Table 2), and ranged from the finest resolution (30-m) to the coarsest resolution (300-m) measured.

Summary of data attributes showed that the range of spatial autocorrelation among some habitat covariates was larger than fold-specific block sizes (i.e., the spatial extent of the folds) for the northern and southern portions of the species’ range, and that the spatial dependence structure for covariates changed among these regions (Figure S2.1). In addition, MESS maps generated using data from individual cross-validation folds demonstrated that extrapolation in environmental space occurred when predicting between folds in the north and south (2-fold cross-validation; Figure S2.2), but was limited when predicting across folds within the northern region (3-fold and 5-fold cross-validation; Figure S2.2). Thus, using data primarily from the north to predict directionally towards the south (and vice versa) resulted in extrapolation in environmental space (i.e., new combinations of covariate values) and into a new spatial dependence structure for the covariate data.

**TABLE 2** Environmental variables and their associated spatial extents (30-m, 90-m, 150-m or 300-m) and feature classes (superscripts) included in the optimally predictive models for each spatially structured cross-validation approach (2-fold, 3-fold and 5-fold) and the random cross-validation approach (random 5-fold).

| Variables                              | 2-fold | 3-fold | 5-fold | Random 5-fold |
|----------------------------------------|--------|--------|--------|---------------|
| Percent clay                           | –      | 30LQ   | 30H    | 300H          |
| Percent sand                           | –      | 30LQ   | 30H    | 300H          |
| Percent silt                           | –      | 30LQ   | 30H    | 300H          |
| Depth to bedrock (cm)                  | –      | 30LQ   | 30H    | 300H          |
| Elevation                              | –      | 300LQ  | 300H   | 300H          |
| Slope                                  | 300L   | 300LQ  | 300H   | 300H          |
| Topographic solar radiation index      | 300Q   | 300L   | 300H   | 300H          |
| Agriculture, developed, developed roads| –      | 300LQ  | 300H   | 300H          |
| Barren, sparsely vegetated, snow-ice   | –      | 300LQ  | 300H   | 300H          |
| Conifer, conifer-hardwood, hardwood    | 300LQ  | 300LQ  | 150L   | 150H          |
| Exotic herbaceous                      | –      | 30C    | 90Q    | 300H          |
| Grassland                              | –      | 30C    | –      | 30C           |
| Riparian                               | –      | 300L   | 300H   | 300LQ         |
| Shrubland                              | 300Q   | 300LQ  | 300H   | 150H          |
| Fire interval                          | –      | 30LQ   | 30H    | 30H           |
| Percent canopy cover                   | 150L   | 150LQ  | 90H    | 90H           |
| Distance to >25% canopy cover          | –      | 30LQ   | 30H    | 30H           |

Note: Possible combinations of feature classes include linear (L), quadratic (Q) and hinge (H). Categorical (C) variables (LANDFIRE, 2016) included in the optimal models with a spatial extent of 30-m (i.e., binary variables, where the binary is assumed to be 0% or 100%) are not subject to additional covariate transformations. Dashes indicate a variable that was not included in the optimal model.

**FIGURE 4** Partial-effect plots for a predictor variable (shrubland) used in the optimally predictive model from the (a) 2-fold, (b) 3-fold, (c) 5-fold and (d) random cross-validation approaches. Values of relative occurrence rate were rescaled between 0 and 1 by dividing all values by the largest value. Examining partial-effect plots can give insight into whether species distribution models may be overfit.
Per cent canopy cover, topographic solar radiation index (TRASP), slope, shrubland, and conifer, conifer-hardwood and hardwood were environmental predictor variables included in all optimally predictive models across all cross-validation approaches (Figures S1.1–S1.4). However, these were the only environmental variables included in the optimal model for the least complex (2-fold) cross-validation approach. Directional relationships (i.e., negative or positive) between these variables and relative occurrence rate were also generally similar across all models. In contrast, the more complex models (i.e., generated via 3-fold, 5-fold and random 5-fold cross-validation) included all or nearly all of the environmental predictor variables considered (Figures S1.2–S1.4).

When optimally predictive SDMs were projected across the historical range of northern Idaho ground squirrels, there were strong differences in the distribution of predicted values among models and hence differences in predictions of suitable habitat. For the least complex model (resulting from 2-fold cross-validation), the distribution of predicted values of relative occurrence rate was shifted towards larger values and was also more variable (Figures 5 and 6). Comparatively, most of the landscape was predicted as having a low relative occurrence rate using the 3-fold and 5-fold cross-validation approaches that fit more complex models based primarily on data from the northern region (Figures 5 and 6). Values of relative occurrence rate at known presence locations ranged from 0.04 to 0.99 for the 2-fold cross-validation approach but were constrained to lower values for the 3-fold and 5-fold cross-validation approaches (Figure 6).

4 | DISCUSSION

We built SDMs to extrapolate predictions of habitat suitability for a threatened ground squirrel to portions of the species’ range where there was limited sampling, while also addressing technical aspects of SDM development for the purpose of spatial transferability,
including: (a) assessment of multiple ways to delineate cross-validation folds for assessing predictive ability, and (b) assessment of a much broader range of statistical regularization parameters than is typically considered by investigators. While previous research has shown that design of cross-validation folds affect the accuracy of estimates of predictive error (Roberts et al., 2017), we demonstrate that these technical details are also critical for predictive model selection and inferring the spatial distribution of high-quality habitat in unsampled portions of a species range. Importantly, optimally predictive models generated via different cross-validation approaches produced vastly different maps of habitat suitability and resulted in differing interpretations of the spatial distributions of high-quality habitat across the historical range of northern Idaho ground squirrels. Moreover, complexity of optimally predictive models was considerably different among the approaches, where in our example, spatially structured cross-validation and fewer data folds favoured less complex models. Greater geographical separation among folds appeared necessary to reliably assess predictive ability across the species range, as the spatial dependence structures of the environmental covariates changed geographically. These results demonstrate that when spatial transferability is the goal, careful consideration of appropriate cross-validation techniques relative to the data structure and intended model use is necessary to induce extrapolations realistically (i.e., prediction into new geographic and environmental space, and new covariate dependence structures). We also demonstrate that optimal values of regularization parameters can be outside the range considered in many applications, resulting in the selection of sub-optimal parameters for regulating model complexity (and hence sub-optimal SDMs and predictions) if larger parameter values are not considered.

Cross-validation is a statistical technique designed to approximate out-of-sample predictive ability of a model when true out-of-sample data are not available (i.e., when a traditional training-testing approach to model building is not possible; Hooten & Hobbs, 2015). Consequently, when one uses random cross-validation to estimate predictive ability of SDMs, they are effectively assuming the stochastic processes governing distributions of species in the region of prediction are identical to those in the region of data collection. In contrast, we used non-random, spatially structured cross-validation because we were attempting to estimate what each model’s predictive ability would be when extrapolated to a new or unsampled portion of a species’ range with potentially different environmental conditions and processes governing distributions (Yates et al., 2018). And our results suggested that there indeed were differences among subregions of the range, even for a rare species whose distribution only includes two counties. The reliability of estimates of predictive ability must therefore depend on how well the spatial partitions of cross-validation folds reflect how and where the model will be extrapolated in practice to inform conservation (Roberts et al., 2017).

Our results corroborate recent studies (Roberts et al., 2017) demonstrating that one can get an inflated assessment of predictive ability by failing to carefully consider how the model will be extrapolated, even when using spatially structured cross-validation. Overly optimistic assessments of predictive error can result from a number of factors, including use of non-independent holdout data (i.e., data from different folds are not independent), closeness of training and testing data (i.e., folds are juxtaposed in space and time), failure of folds to adequately capture data dependence structures that change in space (i.e., folds do not align with changes in spatial autocorrelation structure) and failure of folds to mimic extrapolation in environmental space that will occur under the SDM’s use (i.e., between-fold differences in covariate combinations are optimistic relative to how the model will be extrapolated; Roberts et al., 2017). All of these factors affected our 3- and 5-fold cross-validation results (Appendix S2), and we believe these same factors are common with many SDMs when presence data are not available over portions of the species’ range. Consequently, if a SDM is to be used to predict habitat suitability into poorly sampled areas for conservation planning (e.g., for predicting habitat in other latitudinal subregions as in our example), then spatial cross-validation folds should be designed in such a way as to mimic that intended extrapolation (see also Wenger & Olden, 2012). Otherwise different optimal models with overly optimistic predictive assessments that do not reflect directional predictions can result.

We demonstrated that selection of optimally predictive models using different cross-validation approaches resulted in structurally different models and predictions of habitat suitability. Despite the fact that different cross-validation approaches all resulted in models with apparently high predictive ability, these models resulted in maps of habitat suitability that were quite different and therefore resulted in fundamentally different interpretations of the spatial distribution of high-quality habitat. For example, the 2-fold cross-validation approach suggests a larger proportion of the historical range consists of suitable habitat, whereas predictions of suitable habitat were limited to smaller proportions of the overall range for the 3-fold and 5-fold cross-validation approaches. These results highlight how subtle differences in transferability assessment (e.g., how spatially structured cross-validation folds are delineated) and the attributes of cross-validation folds (e.g., whether extrapolation occurs between folds or not) can lead to dramatic differences in resulting model complexity and predictions, with strong implications for on-the-ground habitat conservation.

Our results also demonstrate why the most appropriate cross-validation approach should be carefully considered based on the SDMs intended use (i.e., how the model will be extrapolated). Our goal was to predict habitat suitability across the historical range of northern Idaho ground squirrels, whereas the vast majority (96%) of presence data come from the northern portion of the range (i.e., where most known populations occur). Thus, the dominant spatial extrapolation occurs between populations located in the north and south, which also coincides with extrapolation of predictions into new environmental space and spatial dependence structures (Appendix S2). Therefore, we believe the model generated via the spatially structured 2-fold cross-validation approach is most appropriate for predicting suitability of northern Idaho ground squirrel habitat in unsampled (or poorly sampled) areas because the folds
mimic the model extrapolation most closely. Further, the 2-fold cross-validation model included environmental predictor variables (Figure S1.1) that appear important for northern Idaho ground squirrel habitat selection across the species’ entire range and is therefore useful for range-wide conservation decision-making. In contrast, the more complex models resulting from 3-fold and 5-fold (both random and spatial) cross-validation included most or all of the environmental predictor variables, which may indicate that some variables (i.e., those not included in the 2-fold model) are specific to populations located in the northern regions of the species’ range, and thus the more complex models may be useful for informing conservation at much smaller spatial extents in regions close to the northern populations. Lastly, irregular spatial distributions of presence locations are common for rare species that occupy small proportions of their range, leading to challenges when delineating spatially explicit cross-validation folds (Roberts et al., 2017). For example, sample size among folds can vary (as with our case study) and extrapolation in environmental space between folds with different sample sizes can artificially inflate apparent error rates (i.e., produce biased high estimates of prediction error; Roberts et al., 2017). The tradeoffs involved are highlighted in our example, wherein a relatively small number of presence locations were available in the southern portion of the historical range. However, the geographic distance needed to adequately capture the intended extrapolation of the model (Figure S2.2) was beyond that of the majority of presence locations clustered in the northern portion of the range. Therefore, incorporating the southern presence locations as an independent fold was necessary to both extrapolate to new dependence structures and avoid inflated estimates of prediction accuracy (i.e., apparent predictive accuracy estimates that are biased high). Ideally, one should strive to achieve an approximately equal number of presence locations among folds when using spatially explicit cross-validation to assess transferability. Yet our example illustrates that this can be difficult to achieve in practice when building predictive SDMs with clustered observations of rare species, particularly when directional extrapolation to new regions is the modelling goal.

Examining modelled species–environment relationships that result from multiple spatially explicit cross-validation approaches, in addition to evaluating fold-specific environmental spaces and autocorrelation patterns, can provide insight as to which models may be overfitting patterns in the data and thus predicting overly complex habitat relationships. In our example, less geographically distinct cross-validation folds (and more folds) resulted in increased model complexity with additional variables, more nonlinear relationships (e.g., increased number of quadratic relationships), and more complex nonlinear relationships (e.g., inclusion of the hinge feature class) between northern Idaho ground squirrel presence and environmental conditions. Based on our understanding of the species’ ecology and the landscape in which it resides, we believe the complicated and often nonlinear relationships resulting from several of the cross-validation approaches (i.e., 3-fold, 5-fold and random cross-validation; Figure 4, Figures S1.2–S1.4) may be biologically unrealistic. Wenger and Olden (2012) reported that predictive performance of SDMs decreased as cross-validation approaches became more conservative but that fewer folds may be necessary in certain contexts. In our example, we found that fewer, carefully thought-out folds resulted in more biologically realistic results with little loss of predictive ability, highlighting the benefit of investigating multiple spatially structured cross-validation approaches when developing SDMs for transferability (Wenger & Olden, 2012).

We also demonstrated that complexity of optimally predictive models was heterogeneous among cross-validation approaches, with fewer spatially structured cross-validation folds separated by increased geographic distance resulting in less complex models. Roberts et al. (2017) argued that fitting models that are too complex for optimal prediction in new areas will often escape detection of investigators without carefully designed, structured cross-validation approaches. We demonstrated that spatially structured cross-validation alone does not necessarily guard against selection of overfit models. All but one of our spatially explicit cross-validation approaches produced models with extensively choppy and questionable covariate relationships (e.g., 5-fold cross-validation) or superfluous covariates with little effect on predictions of habitat suitability (e.g., 3-fold cross-validation). This problem was likely caused by the same factors that contribute to inaccurate predictive error assessments, namely the closeness of cross-validation folds from the northern region combined with overfitting data from the north when the autocorrelation structure and environmental conditions changed towards the south. Complex models that overfit patterns in data from one region will often extrapolate poorly to new regions with different dependence structures and covariate combinations, making it desirable to intentionally induce such extrapolations among cross-validation folds as we did here (Roberts et al., 2017; Yates et al., 2018). In our example only the spatially structured 2-fold cross-validation intentionally induced the intended extrapolation (both in geographic and environmental space), and this approach produced a model with both biologically realistic relationships and no superfluous covariates, and thus little sign of overfitting. As such, explicit steps to limit complexity of SDMs are often warranted if the goal of modelling is spatial extrapolation, including the use of cross-validation folds that intentionally induce extrapolations and the optimization of regularization penalties.

Statistical regularization provides an effective framework for reducing model complexity and the effect of uninformative variables when optimizing SDMs for transferability. Yet our results demonstrate that relying on default values or arbitrarily limited ranges for grid searches over the regularization parameter space can produce models with sub-optimal prediction (even within a cross-validation scenario) that overfit the data. We showed that relationships between regularization and predictive ability can be noisy, with local optima and flat spots, indicating that a wider range of regularization parameter values may be necessary to empirically differentiate local from true optima. However, the magnitude of the optimal regularization parameter and the grid search can vary under different predictive assessments (e.g., different cross-validation approaches), model parameterizations (e.g., complexity of covariate transformations)
and data structures. In our case, the optimal regularization parameter for 2-fold cross-validation far exceeded “maxnet’s” default value of 1.0, as well as the upper limit that many investigators use when implementing grid searches to optimize regularization (typically 0–10). Moreover, for cross-validation approaches whose optimal models had regularization parameters ≤ 1 the models did not appear penalized enough for resulting predictions to achieve optimal transferability across the historical range of northern Idaho ground squirrels. Consequently, regularization parameters exceeding those typically used (i.e., default values or small grid searchers with arbitrary cut off values) may need to be considered if a goal of the SDM is to optimize model complexity and extrapolate predictions to areas where presence data are limited. Consideration of larger regularization parameter values may also be necessary when extrapolation in environmental space occurs between cross-validation folds, when a large number of occurrence data are available, or when species-habitat relationships are modelled at a fine resolution (e.g., 30-m).

Importantly, the upper bound of a grid search might be indicated by regularization parameter values that force all regression coefficients to zero, thus resulting in an intercept only model for predictions. Nonetheless, additional research is warranted on the factors that dictate optimal values of regularization parameters and conditions under which larger ranges of penalty values need consideration.

Most relationships among predictor variables and relative occurrence rate from the optimally predictive 2-fold model reinforce current knowledge of northern Idaho ground squirrel habitat selection. However, shrubland effects were surprising (Figure 4a), because shrub cover is not generally believed to benefit northern Idaho ground squirrels. We suspect the positive relationship of shrubland to relative occurrence rate could be an artefact of the observational scale used for this variable (300-m). Northern Idaho ground squirrel populations often occur in small meadow habitats surrounded by shrubby and/or forested habitat and larger scales of effect may associate habitat selection with vegetation surrounding open meadows (i.e., meadows surrounded by shrubs were better habitat). It is also possible that meadows surrounded by shrubby vegetation have simply not yet been negatively impacted by effects of forest expansion and succession on occupancy, as succession is currently at an earlier stage (and thus quality could decline further over time as shrubs are replaced by trees on the meadow margins). Moreover, further efforts to model habitat suitability at and around more isolated populations (e.g., the population in the southern part of the range) could be useful for local-scale conservation decision-making, particularly if factors that influence habitat selection differ among populations and dispersal from other populations is limited.

Lastly, the scale at which animals perceive and interact with the environment can influence habitat selection and vary among environmental variables that affect a species’ distribution (Mayor et al., 2009; McGarigal et al., 2016; Wiens, 1989). The spatial scales at which environmental variables affect species presence are often not consistent among covariates, and therefore, integrating scale-optimized predictor variables into a multi-scale SDM can improve predictions (McGarigal et al., 2016; Stevens & Conway, 2019). Our results corroborate several studies with other taxa where optimal covariate scales for predicting habitat suitability were not consistent among variables (McGarigal et al., 2016; Stevens & Conway, 2019). However, we also found that selection of optimally predictive scales was sensitive to the cross-validation approach used, which indicates that different scales likely predict better for different purposes (i.e., under different extrapolations) because different cross-validation approaches mimic different extrapolations (both in environmental and geographic space). To our knowledge, this is a novel result and it reinforces our suggestion that cross-validation folds be designed carefully to mimic the intended goals of extrapolation (else scales may be optimal for other purposes). This result further emphasizes that covariate observational scales should not be selected arbitrarily but instead should be optimized relative to specific study objectives (Stevens & Conway, 2019).

5 | CONCLUSION

SDMs are powerful tools that aid in conservation and recovery planning for rare species and continued development of best practices for SDM construction that are tailored to their intended use is imperative. Methods to assist with balancing SDM complexity and transferability can bolster the spatial predictive ability of SDMs and improve biological realism of results. Non-random, spatially structured cross-validation approaches provide investigators with important information about the ability of SDMs to extrapolate predictions geographically to portions of a species range where localized information on presence/absence is less well-known, and statistical regularization is a vital tool for optimizing model complexity while developing transferable models. Not all spatially structured cross-validation approaches will be appropriate for a given application, which can result in inflated estimates of predictive accuracy and unwarranted confidence in model predictions. Consequently, spatially explicit cross-validation fold should carefully mimic a models’ intended extrapolation, and different modelling objectives and prediction goals will likely necessitate different cross-validation approaches (Roberts et al., 2017). Thus, care needs to be taken to reduce the risks of overfit models and overly optimistic assessments of predictive ability in locations where presence data are limited or not available. Otherwise SDMs can generate sub-optimal predictions and thus sub-optimal characterizations of habitat suitability for rare species in need of conservation. Lastly, we assessed these challenges for a rare species with a very restricted distribution. We expect the challenges of balancing complexity and transferability for predictive SDMs will only be more severe for species with larger distributions, given that extrapolations into unsampled areas (in both environmental and geographic space) are likely to be even more pronounced than in our example.

ACKNOWLEDGEMENTS

Funding was provided by the Idaho Department of Fish and Game and the United States Fish and Wildlife Service. We thank D.
Evans-Mack, L. Nutt and G. Burak for their involvement in the modelling process and for contributing presence locations of northern Idaho ground squirrels. We also thank all additional staff and field technicians that assisted in collecting presence locations of northern Idaho ground squirrels.

**AUTHORS’ CONTRIBUTIONS**

N.H., C.C., and A.G. developed the project. N.H. and B.S. performed

**DATA AVAILABILITY STATEMENT**

Data subject to third party restrictions. The data used for this study are subject to third party restrictions and are not publicly available (i.e., the data include precise locations of a rare, federally threatened species).

**REFERENCES**

Allison, A. Z. T., Morris, A. E., & Conway, C. J. (2019). 
Effects of Forest Encroachment on Behavior and Demography of the Northern Idaho Ground Squirrel: Annual Progress Report 2019. Idaho Cooperative Fish and Wildlife Research Unit, Moscow, Idaho.

Anderson, R. P., & Gonzalez, I., Jr (2011). Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent. Ecological Modeling, 222(15), 2796–2811. https://doi.org/10.1016/j.ecolmodel.2011.04.011

Araújo, M. B., Pearson, R. G., Thuiller, W., & Erhard, M. (2005). Validation of species–climate impact models under climate change. Global Change Biology, 11(9), 1504–1513. https://doi.org/10.1111/j.1365-2486.2005.01000.x

Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many? Methods in Ecology and Evolution, 3(2), 327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Barbosa, S., Andrews, K., Goldberg, A., Singh-Gour, D., Hohenlohe, P. A., Conway, C. J., & Waits, L. (2021). The role of neutral and adaptive genomic variation in population diversification and speciation applied to two Idaho ground squirrel sister species of conservation concern. Molecular Ecology, in press.

Bickel, P. J., Li, B., Tsybakov, A. B., van de Geer, S. A., Yu, B., Valdés, T., Rivero, C., Fan, J., & van der Vaart, A. (2006). Regularization in statistics. Test, 15(2), 271–344. https://doi.org/10.1007/BF02607055

Burak, G. S., Goldberg, A. R., Galloway, J. M., Evans Mack, D., & Conway, C. J. (2018). Collaborating to save a tiny threatened species: What does the northern Idaho ground squirrel need to survive? The Wildlife Professional, 12(5), 39–42.

Elith, J., & Franklin, J. (2013). Species distribution modeling. In S. Levin (Ed.), Encyclopedia of biodiversity (2nd ed., pp. 692–705). Academic Press.

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, J. R., Huettmann, F., Leathwick, R. J., Lehmann, A., Li, J., Lohmann, G. L., Loiselle, A. B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC Overton, M. J., Townsend Peterson, A., ... Zimmermann, E. N. (2006). Novel methods improve prediction of species’ distributions from occurrence data. Ecography, 29(2), 129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shift ing species. Methods in Ecology and Evolution, 1(4), 330–342. https://doi.org/10.1111/j.2041-210X.2010.00036.x

Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics, 40, 677–697. https://doi.org/10.1146/annurev.eyecolsy.110308.120159

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. Diversity and Distributions, 17(1), 43–57. https://doi.org/10.1111/j.1472-4442.2010.10072.x

Ferrier, S., Watson, G.,Pearce, J., & Drielsma, M. (2002). Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level Modelling. Biodiversity & Conservation, 11(12), 2275–2307.

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation, 24(1), 38–49. https://doi.org/10.1017/S0376892997000088

Gerber, B. D., Kendall, W. L., Hooten, M. B., Dubovsky, J. A., & Drewien, R. C. (2015). Optimal population prediction of sandhill crane recruitment based on climate-mediated habitat limitations. Journal of Animal Ecology, 84(5), 1299–1310. https://doi.org/10.1111/1365-2656.12370

Gerber, B. D., & Northrup, J. M. (2019). Improving spatial predictions of animal resource selection to guide conservation decision making. Ecology, 101(3), e02953.

Goldberg, A. R. (2018). Diet, disease, and hibernation behavior of Northern Idaho Ground Squirrels. PhD thesis, University of Idaho.

Goldberg, A. R., & Conway, C. J. (2021). Hibernation behavior of a federally-threatened ground squirrel: Climate change and habitat selection implications. Journal of Mammmalogy, in press.

Goldberg, A. R., Conway, C. J., Evans Mack, D., & Burak, G. (2020). Winter versus summer habitat selection in a threatened ground squirrel. Journal of Wildlife Management, 84, 1548–1559.

Goldberg, A. R., Conway, C. J., Tank, D., Andrews, K. R., Gour, D. S., & Waits, L. (2020). Diet of a rare herbivore based on DNA metabarcoding of feces: Selection, seasonality, and survival. Ecology and Evolution, 2020(10), 7627–7643. https://doi.org/10.1002/ece3.6488

Guiller-Merio, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentiini, P. E., McCarthy, M. A., Tingley, R., & Wintle, B. A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. Global Ecology and Biogeography, 24(3), 276–292. https://doi.org/10.1111/geb.12268

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. Ecology Letters, 8(9), 993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, L., Sutcliffe, P. R., Tulloch, A. I. T., Regan, T. J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T. G., Rhodes, J. R., Maggini, R., Setterfield, S. A., Elith, J., Schwartz, M. W., Wintle, B. A., Broennimann, O., Austin, M., … Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. Ecology Letters, 16(12), 1424–1435. https://doi.org/10.1111/ele.12189

**ORCID**

Nolan A. Helmstetter
https://orcid.org/0000-0002-3173-5443

Courtney J. Conway
https://orcid.org/0000-0003-0492-2953

Bryan S. Stevens
https://orcid.org/0000-0001-9463-5426

Amanda R. Goldberg
https://orcid.org/0000-0003-3094-8241

**PEER REVIEW**

The peer review history for this article is available at https://publo

| ns.com/publon/10.1111/ddi.13174.

| The peer review history for this article is available at https://publo

| ns.com/publon/10.1111/ddi.13174.
far should you stray from what you know? Ecological Modelling, 220(4), 589–594. https://doi.org/10.1016/j.ecolmodel.2008.11.010

Wagner, B., & Evans Mack, D. (2020). Long-term Population Monitoring of Northern Idaho Ground Squirrel: 2019 Implementation and Population Estimates. Annual Report, Idaho Department of Fish and Game, Boise, ID.

Wenger, S. J., & Olden, J. D. (2012). Assessing transferability of ecological models: An underappreciated aspect of statistical validation. Methods in Ecology and Evolution, 3(2), 260–267. https://doi.org/10.1111/j.2041-210X.2011.00170.x

Wiens, J. A. (1989). Spatial scaling in ecology. Functional Ecology, 3(4), 385–397. https://doi.org/10.2307/2389612

Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., Fielding, A. H., Bamford, A. J., Ban, S., Barbosa, A. M., Dormann, C. F., Elith, J., Embling, C. B., Ervin, G. N., Fisher, R., Gould, S., Graf, R. F., Gregr, E. J., Halpin, P. N., ... Sequeira, A. M. M. (2018). Outstanding challenges in the transferability of ecological models. Trends in Ecology & Evolution, 33(10), 790–802. https://doi.org/10.1016/j.tree.2018.08.001

Yensen, E. (1991). Taxonomy and distribution of the Idaho Ground Squirrel, Spermophilus brunneus. Journal of Mammalogy, 72(3), 583–600. https://doi.org/10.2307/1382142

Yensen, E., Luscher, M. P., & Boyden, S. (1991). Structure of burrows used by the Idaho ground squirrel, Spermophilus brunneus. Northwest Science, 65(3), 93–100.

Yensen, E., & Sherman, P. W. (1997). Spermophilus brunneus. Mammalian Species, 560, 1–5. https://doi.org/10.2307/3504405

BIOSKETCH

The Conway Lab is associated with the Idaho Cooperative Fish and Wildlife Unit at the University of Idaho and is dedicated to answering questions related to behavioural ecology and management of bird and mammal populations. The laboratory conducts research on a variety of species of concern found across North America and supports both graduate and undergraduate research while fostering collaborative efforts with state and federal agencies and public stakeholders.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Helmstetter NA, Conway CJ, Stevens BS, Goldberg AR. Balancing transferability and complexity of species distribution models for rare species conservation. Divers Distrib. 2021;27:95–108. https://doi.org/10.1111/ddi.13174