Iterative species distribution modeling results in the discovery of novel populations of a rare cold desert perennial

Running title: Iterative species distribution modeling

Israel T. Borokini[1,2,*], Kenneth Nussear[3], Blaise Petitpierre[4], Thomas E. Dilts[5], Peter J. Weisberg[5]

[1]Ecology, Evolution and Conservation Biology graduate program, Department of Biology, University of Nevada, Reno, Reno NV 89557

[2]Department of Integrative Biology, University of California, Berkeley, Berkeley CA 94720

[3]Department of Geography, University of Nevada, Reno, Reno NV 89557

[4]Institute of Earth Surface Dynamics, University of Lausanne, Lausanne CH-1015, Switzerland

[5]Department of Natural Resources and Environmental Science, University of Nevada, Reno, Reno NV 89557

*Corresponding author: iborokini@berkeley.edu; ORCID: 0000-0002-1258-7932

ABSTRACT

Niche modeling for rare and range-restricted species can generate inaccurate predictions leading to an overestimation of a species geographic distribution. We used an iterative ensemble modeling approach and model-stratified field surveys to improve niche model formulation and better understand the ecological drivers of *Ivesia webberi* distribution. *I. webberi* is a U.S. federally threatened herbaceous species, narrowly distributed in the Western Great Basin Desert. Niche models for *I. webberi* were fitted using 10 replicates each of six modeling algorithms, while geographical projections of habitat suitability were generated using weighted ensembles of models with optimal performance. The resulting model projections were used to guide field surveys for five years, generating additional spatial data, which were added to the existing dataset for subsequent modeling. Model performance across iterations was investigated, while niche differences in the spatial dataset were explored. Model-guided field surveys resulted in the discovery of several new locations of *I. webberi* and an expansion of the species' known range by 63 km. Model performance was higher in the earlier but overfitted niche models; overfitting was corrected in the final models, while predicted habitat suitability was reduced by 50%. Findings show that *I. webberi* niche is associated with biotic, topographic and bioclimatic variables. Further, a partial overlap was observed between environmental conditions of the initial and new locations (Schoener's D=0.47), which can be decomposed into 93% of niche stability. This indicates that the majority of the newly discovered locations are within the environmental niche of the initial data.

KEYWORDS: habitat suitability, iterative ensemble modeling, niche overlap, field validation surveys, niche stability, *Ivesia webberi*, Great Basin Desert

1. INTRODUCTION

Limited empirical information on the geographical distributions of taxa (Wallacean shortfall; Whittaker et al. 2005) can impact the assessment of species rarity resulting in misguided conservation prioritizations (Coddington et al. 2009). Field surveys, especially those conducted using random sampling strategies, can generate additional biodiversity data to mitigate this; however, such surveys are costly, time-consuming, and ineffective for rare species, while human resources are limited (Hirzel & Guisan 2002, Guisan et al. 2006). Therefore, scientists and conservation managers have considered other cost-effective methods to stratify and prioritize field surveys, usiing, for example, expert opinion and quantitative niche modeling. Niche models can relate species' occurrences to their environmental conditions to quantify the realized niche, that is, species' known locations due to environmental tolerance observed in the field (Hutchinson 1957). These niche models generate geographic predictions of species habitat suitability that can be used to stratify and optimize sampling efficiency (Chiffard et al. 2020). Moreover, integrating the new spatial data from model-guided sampling can reduce spatial bias in subsequent modeling iterations, improve the predictive accuracy of niche models for rare species, and reliably identify biologically-relevant environmental factors (Singh et al. 2009, Rinnhofer et al. 2012).

Understanding the distribution of rare species is critical for effective conservation planning, but with few, incomplete, and biased spatial data, it can be challenging to model the niches of rare species with high predictive accuracy (Hernandez et al. 2006, Wisz et al. 2008), a condition referred to as the rare species modeling paradox (Lomba et al. 2010). This is because fewer occurrence points in spatial dataset indicates low prevalence, which weakens the

analytical power of the models and inflates bias in niche models (Vaughan & Ormerod 2003). Furthermore, correlative species distribution models include the underlying assumption that species are in equilibrium with their environment (i.e., temporal and spatial stationarity), and that all important and biologically-relevant variables have been included in the niche model (Elith & Leathwick 2009). This presents challenges to modeling rare species because the inclusion of many predictors when occurrences are few can lead to model overfitting (Wisz et al. 2008, Jarnevich et al. 2015). Moreover, limited natural history knowledge makes predictor variable selection challenging and potentially subjective for rare species (Aranda & Lobo 2011). Consequently, poorly fit models and misjudgments of model predictions can lead to over- or underestimation of the species' niche, resulting in poorly informed management decisions (Ramesh et al. 2017, Burns et al. 2020). Despite the development of several statistical methods to reduce prediction errors in niche modeling, the most practical way is to increase spatial data for rare species, which is inevitably linked with data collection during field surveys. Therefore, predictions of species distribution modeling (SDM) for rare species should not be treated as truth, but can be used as hypotheses for further ecological or biogeographical investigations (Stockwell & Peterson 2002, Jarnevich et al. 2015, Sofaer et al. 2019).

The discovery of new locations of targeted species from SDM-guided field surveys is well documented in the literature (e.g., Williams et al. 2009, de Siqueira et al. 2009, Särkinen et al. 2013, Burns et al. 2020). These novel discoveries underscore the importance of SDMs as an important conservation tool. SDMs have been used to evaluate the degree of species rarity (Broennimann et al. 2006), and identify areas that may serve as future climatic refugia (Sousa-Silva et al. 2014). Furthermore, SDMs are also used to advance scientific knowledge of species-

environment relationships (Jiménez-Valverde et al. 2011), and identify niche constraining environmental factors (Gorban et al. 2011). SDM predictions are often integrated into models of population and landscape genetics (e.g., Ikeda et al. 2016, Banerjee et al. 2019), and spatial phylogenetics (e.g., Thornhill et al. 2017). Beyond conservation uses, newly discovered occurrences may have significant ecological contributions to the understanding of the overall species' niche. For example, additional occurrence points may be found either within the existing realized niche space or in areas with different ecological conditions, thus expanding the species environmental niche. The COUE (that is, centroid shift, overlap, unfilling, and expansion) framework can be used to quantify realized niches of species from different ranges and categorize the niche position of newly discovered occurrences (Broennimann et al. 2012). This framework has been used to investigate niche dynamics between the native and invaded ranges of invasive species (Broennimann et al. 2012, Strubbe et al. 2013), as well as niche evolution vs conservatism between sister taxa (Villegas et al. 2021).

The aim of this study was to assess relative improvement of using iterative sampling approach alternating between niche modeling and model-guided field surveys relative to a presence/absence modeling approach using only data available at the onset of the study to predict the distribution of a rare plant (*Ivesia webberi* A. Gray). Therefore, we asked the following questions: (1) Which environmental variables determine the distribution of *I. webberi*, and how does the species-environment relationship change with each iteration of the SDMs given additional spatial data? (2) Do additional distribution data alter habitat suitability map projections across modeling iterations? (3) Is the environmental niche conserved

throughout the modeling iterations? (4) Do modeling iterations improve the predictive accuracy of species distribution models for *Ivesia webberi*?

2. MATERIALS AND METHODS

2.1. Study species and study area

Ivesia webberi is a U.S. federally listed threatened perennial forb restricted to the eastern foothills of the Sierra Nevada and the adjacent western edge of the Great Basin Desert. *I. webberi* was estimated to have originated between 1.3 and 3.8 million years ago (Töpel et al. 2012), and may be one of the many Great Basin Desert neoendemic and phylogenetically young taxa that have not had enough time to fully colonize their range (Kraft et al. 2010, Thornhill et al. 2017). At the outset of our study, it was known from 23 spatially-aggregated locations, occurring in or near ephemeral washes and dry forest meadow gaps in mostly gently sloped areas (Witham 2000). These presence locations were visited multiple times between 2015 and 2020, and therefore are not prone to positional error. These locations have experienced varying degrees of biological invasion pressures from *Bromus tectorum*, *Taeniatherum caput-medusae*, and *Poa bulbosa*, as well as disturbances from wildfires, cattle grazing and off-highway vehicle use.

The study extent was defined by a 60 km buffer from marginal ranges of known populations as of 2015. The species produces achenes which are not adapted for long-range dispersal; therefore, the study area was restricted to mask out expansive adjacent unsuitable areas of playas in the central Great Basin Desert. This modeling decision was guided by natural history indicating that the species is located in sparsely vegetated low sagebrush

124 (*Artemisia arbuscula*) communities in mid-elevation areas of the western Great Basin Desert and
125 the adjacent northern Sierra Nevada eastern foothills (US Fish and Wildlife Service [USFWS]
126 2014). Climatic conditions in these sites are characterized by relatively mild winters and hot
127 summers (Svejcar et al. 2017). Thus, temperatures range from an average of -5.8 °C in the
128 winter to an average of 28 °C in the summer, while annual precipitation varies between 25 and
129 33 cm, most of which falls as snow or rain during the winter months.

130 **2.2. Distribution data**

131 We began species distribution modeling in 2015 with 23 occurrence points and 758
132 absence points obtained from the Nevada Natural Heritage Program (NNHP). The absence
133 points represent areas where *I. webberi* was not detected during historical surveys by NNHP
134 botanists and citizen scientists. Additional spatial points were added following iterative
135 modeling and field validation cycles in predicted suitable habitats. In all modeling iterations,
136 the absence points were thinned using a distance of 7.5 km in *spThin* R package version 0.2.0
137 (Aiello-Lammens et al. 2015) to reduce the effects of spatial aggregation, and mitigate low
138 prevalence in the spatial dataset. Additionally, absence points within 5 km of an occurrence
139 point were removed to avoid false negatives. The remaining absence points were merged with
140 the presence points for niche modeling (Table 1).

141 **2.3. Predictor variables**

142 A total of 72 predictor variables describing edaphic, topographic, land cover, vegetative
143 cover, and climatic factors were assembled for fitting SDMs for *Ivesia webberi* (see Table S1). To
144 avoid overfitting and maintain a 1:10 ratio of predictor variables to occurrence points (Harrell
145 et al. 1996), the full set of predictor variables was reduced to six uncorrelated predictors (Table

2) using a combination of the Kendall *r* correlation coefficient (r>0.6), feature selection runs in *Boruta* R package version 4.0.0 (Kursa & Rudnicki 2010), and recursive feature elimination algorithm in *caret* R package version 6.0-78 (Kuhn 2008).

The climatic variables (actual evapotranspiration, minimum monthly temperature and summer seasonal precipitation) were downsampled from the Parameter-elevation Relationships on Independent Slopes Model (PRISM) climatic data (1971-2000) normals (Daly et al. 2008), from 800-m to 30-m spatial resolution using the Climatic Water Deficit Toolbox (Dilts et al. 2015) and ordinary kriging. The cosine-transformed aspect, ranging from -1 (south-facing slope) to +1 (north-facing slope), was derived from slope using the formula: $\theta \times \cos(\alpha)$, where θ is slope (in percentage), and α is aspect (in radians), while slope was calculated from the 1 arc-second digital elevation models (DEM; United States Geological Survey [USGS] 2017). Perennial herbaceous vegetative cover, a vegetation type raster layer, was obtained from the Multi-Resolution Land Characteristics (MRLC) development of the 2016 U.S. National Land Cover Database (NLCD; Xian et al. 2013). Topographic Position Index (TPI) was calculated from the DEM using the formula described by Weiss (2001).

2.4. Iterative ensemble niche modeling and model-based sampling

The SDMs were fitted at 30 m resolution to capture the landscape and ecological heterogeneity in the study area, particularly in the *Ivesia webberi* locations that occur within forest gaps. An ensemble modeling approach was used in all niche modeling iterations. The use of multi-algorithm ensemble models renders predictions less susceptible to biases, assumptions, or limitations of any individual algorithm, while broadening the types of environmental response functions that can be identified (Araújo & New 2006). SDMs have

been developed from a wide range of modeling techniques including regression, classification, and machine learning algorithms (Lauzeral et al. 2012). Because these algorithms have different predictive performances under different contingencies (Li et al. 2013), fitting of niche models using different algorithms and combining their model parameters to build a consensus or ensemble model is often recommended (Marmion et al. 2009). Ten replicates of six algorithms (Boosted Regression Trees, Random Forests, Maximum Entropy, Artificial Neural Networks, Generalized Additive Models and Generalized Linear Models) were fitted using the *biomod2* R package (Thuiller et al. 2009). All statistical packages were implemented in R statistical software version 4.0.2 (R Core Team 2020). See Table S2 for modeling details.

Model performance was evaluated using four metrics, including true skill statistic (TSS; Allouche et al. 2006), area under the curve (AUC) of the receiver operating characteristics (ROC) plot (Hanley & McNeil 1982), TSS-based specificity, and Boyce Index (Boyce et al. 2002). In each modeling iteration, three predictors, selected from the six uncorrelated variables, were used to fit the niche models. Models were fitted with 80% of the data with 20% used for k-fold cross-validation (Araújo et al. 2005, Thuiller et al. 2009). Model replicates with TSS ≥ 0.7 were averaged into ensemble models, which were used to produce geographic projections of habitat suitability (Marmion et al. 2009, Thuiller et al. 2009). On the habitat suitability maps, cells with ≥0.5 occurrence probability were considered suitable to delineate areas with higher habitat suitability values for field validation surveys. Uncertainty in habitat suitability projections was visualized on maps of coefficients of variation from the iterative niche ensemble models (Hortal 2008).

Habitat suitability maps produced by the SDMs were used to guide field validation surveys to areas of high predicted probability of occurrence. The non-thinned absence points were overlaid on the predicted habitat map and predicted suitable and unsuitable areas that had not been previously surveyed were selected for field validation. Field validation surveys were done alone by the first author (I.T.B.), between May and June of each year when the plants were in flower, to increase chances of detection. Additional spatial data from field surveys were used in the subsequent modeling iteration and site selection for post-modeling field validation. The iterative modeling and field surveys were repeated for five years. For each newly discovered population, we calculated the distance to the nearest previously known occurrence with the *FNN* R package version 1.1.3 (Beygelzimer et al. 2019).

The relative importance of the predictor variables in all iterative SDMs was evaluated using the jackknife test (Phillips et al. 2006), while species-environment relationships were described with partial response curves using the evaluation strip method (Elith et al. 2005) as implemented in the *biomod2* R package. We assessed the trends and statistical significance of the model performance across the years of iterative niche modeling, to investigate if additional spatial data improved the overall predictive accuracy of the iterative ensemble SDMs. Mean scores of the four model performance metrics for each of the six algorithms (10 replicates each) were regressed against the years of iterative SDMs using multivariate multiple linear regression (MMLR), while the statistical significance of the MMLR models were corrected using the Tukey post-hoc test.

We also assessed the reliability of these iterative SDM predictions by checking for model overfitting with a spatial cross-validation approach using block partitioning. Spatial

211 block partitioning is a nonrandom allocation of spatial data to reduce the effect of spatial bias
212 and autocorrelation in ecological models (Valavi et al. 2019). The entire study area was divided
213 into six equal latitudinal and longitudinal bins, which were then clustered into three spatial
214 blocks. Two spatial blocks were used for model training while the third block was used for
215 testing. Spatial block partitioning was done in the *blockCV* R package version 2.1.4 (Valavi et al.
216 2019), while the niche models were conducted in *biomod2* R package version 3.5.1, using
217 similar model tuning as used for the iterative SDMs. Partitioning our relatively small spatial
218 dataset could only meet the requirements for modeling with random forest, maximum
219 entropy, and artificial neural networks, which were used for the spatial block-based ensemble
220 niche modeling. Overfitting was assessed as the difference between the block (training) and
221 test AUC values (Warren & Siefert 2011).

222 **2.5. Assessment of the change in *I. webberi* niche across modeling iterations**

223 We used the COUE framework to investigate the position of the new locations relative
224 to the initial niche of *I. webberi*. The COUE framework, based on the principal component
225 analysis (PCA), allows for direct comparison of species-environment relationships
226 (Broennimann et al. 2012). A kernel density function is applied to smooth the varying
227 sampling sizes in the two sets of occurrence points in a PCA gridded environmental space to
228 calculate the niche metrics (Petitpierre et al. 2012, Broennimann et al. 2012). We calculated
229 niche overlap (Schoener's *D*), stability, expansion, and unfilling between the initial (2015) and
230 a combination of all new (2018-2020) *I. webberi* locations, based on the environmental space of
231 the six uncorrelated predictor variables. Schoener's *D* is calculated between the environmental
232 occupancy of the two niches, and it ranges between 0 (no overlap) and 1 (total overlap) which

233 represent niche divergence and similarity, respectively (Brown & Carnaval 2019). In this study,
234 niche stability represents the proportion of the environmental space in the newly discovered
235 locations available in the initial occurrences' environmental space, while niche expansion
236 represents the proportion of the environmental space in the new locations that are not
237 available in the initial locations. A niche unfilling estimate was used to investigate whether the
238 new occurrences only colonized a limited portion of the environmental space of the initial
239 occurrences (Petitpierre et al. 2012, Guisan et al. 2014). The COUE framework was
240 implemented using a niche similarity test, which assumes that the environmental niches in the
241 new occurrences are similar to the initial occurrences, given that habitat suitability map
242 projections of the initial occurrences were used for field validation surveys (Liu et al. 2020, Pili
243 et al. 2020). The niche similarity test generated random estimates for each of Schoener's *D*,
244 niche stability, expansion and unfilling, using 1000 randomizations of the niche positions of
245 the initial and newly discovered occurrences. These randomizations were used to check if the
246 observed niche overlap and stability were higher and if observed niche expansion and
247 unfilling were lower than expected by chance. Furthermore, we extracted values of the initial
248 niche density at the new locations to quantify the degree of niche stability or expansion in the
249 new locations. Niche density values range from 0 to 1; 0 represents new locations outside the
250 initial niche (that is, niche expansion), while 1 represents new locations in the core of the initial
251 niche. The niche similarity test was run with the development version 3.2.1 of the *ecospat* R
252 package available on *github* (Broennimann et al. 2016).
253 Additionally, we quantified the number of predicted suitable raster cells (≥0.5
254 probability of occurrence) in the habitat suitability maps. We also performed a niche overlap

255 analysis on the geographic projections of habitat suitability between the 2015 (initial) and 2020
256 (final) model iterations for each iterative niche model, using the *I* similarity metric, which is
257 based on Hellinger distance (Warren et al. 2008). The *I* similarity metric ranges from 0 to 1
258 representing the degree of pairwise similarity in niche model projections (Warren et al. 2008).
259 This map-based niche overlap test is a cell-by-cell comparison with a randomization test of
260 geographical predictions of the four iterative SDMs (Warren et al. 2008); it was performed in
261 the *dismo* R package (Hijmans et al. 2017).
262

263 **3. RESULTS**

264 **3.1. Which environmental variables are associated with the ecological niche of *I. webberi*,**
265 **and how does our understanding of these species-environment relationships change across**
266 **the SDMs?**

267 Throughout the iterative SDMs from 2015 to 2020, the perennial herbaceous vegetative
268 cover consistently contributed the most to the fitted distribution of *I. webberi* (Fig. 1a-d). In the
269 2020 model iteration, *I. webberi* showed an asymmetric and threshold response curve for
270 perennial vegetative cover, with suitable sites occurring in areas with moderate (>20%) to high
271 native perennial forb cover (Fig. 2a). Topographic Position Index (TPI) was the second most
272 important predictor across all model iterations. The response curve for TPI is bimodal and
273 asymmetric, illustrating that *I. webberi* occurs on sites that are either gentle lateral valleys or
274 ridges (Fig. 2c). AET was the third most important predictor in the 2015 niche model iteration
275 (Fig. 1a). The cosine-transformed slope aspect, a proxy for exposure to sunlight, came third in
276 the 2018 and 2019 iterations (Fig. 1 c and d), while summer seasonal precipitation was the third

277 most important predictor for the 2020 iteration (Fig. 1d). The response curve for summer
278 seasonal precipitation shows a threshold response, where the probability of *I. webberi*
279 occurrence was maximized at >25 mm summer precipitation, beyond which the curve
280 flattened (Fig. 2b).

281 **3.2. Does the additional spatial data impact *I. webberi* niche dynamics and habitat**
282 **suitability map projections across modeling iterations?**

283 The iterative ensemble SDMs and model-guided field surveys resulted in the discovery
284 of seven new locations of *I. webberi* (30.4% of the initial dataset), while two additional new
285 locations (8.7% of the initial dataset) were discovered opportunistically by local botanists. The
286 distance from the new locations to the closest known locations ranged from 30 m to 63 km
287 (Table 3). As a result, the northern distribution range of the species was expanded by 63 km
288 (Table 3). However, the percentage of the suitable raster cells in the ensemble habitat
289 projections decreased from 5.98% in 2015 to 3.34% in 2020 (Fig. 3 a-d). Despite the decrease in
290 the percentage of suitable grid cells, niche overlap between the geographical projections of the
291 2015 and 2020 model iterations were high (Hellinger's *I*=0.89). The model projections also
292 predicted higher probability of *I. webberi* occurrence in locations near the center of the study
293 area (Fig. 3 a-d). Prediction uncertainties (coefficients of variation) were relatively low across
294 all four projections (Fig. 4 a-d).

295 **3.3. Is *I. webberi* environmental niche conserved throughout the modeling iterations?**

296 The first two PCA axes explained 49% of the variation in the data, both of which
297 represent topo-climatic gradients (Figure 5, Table S3), while the third axis, representing the

perennial vegetation cover, explained an additional 17.4% of the variance (Table S3). The PCA niche similarity test shows that the environmental niche of the new occurrences overlaps that of the initial occurrences with marginal significance (Schoener's D=0.47; p=0.05). This finding was corroborated by the niche stability result showing that the environmental niche of the new occurrences is similar to the initial occurrences (niche stability=0.93; Fig. 5), although this high value was marginally significant (p=0.09; Fig. 6). Furthermore, the values of the new locations in the niche density of the initial occurrences ranged from 0.21 to 0.87 (Table 3), which shows that all of these new points are found within the initial niche, thus, niche stability. However, niche changes between the initial and new occurrences are due to unfilling (estimate=0.47; p=0.11) rather than expansion (estimate=0.07; p=0.09; Fig. 6). The majority of the randomized niche overlap and stability estimates were lower than the observed values (Fig. 6a & b, respectively), while the majority of the randomized expansion and unfilling estimates were higher than the observed values (Fig. 6c & d, respectively).

3.4. Do modeling iterations improve the predictive accuracy and reliability of the SDMs for *Ivesia webberi*?

Figures 7a-d shows the mean performance metrics for the iterative ensemble SDMs between 2015 and 2020. The TSS-based model performance scores significantly decreased from 0.81 in 2015 to 0.78 in the 2020 model iterations (Tukey post-hoc: p=0.01). Similarly, the AUC scores significantly decreased from 0.89 to 0.82 between the 2015 and 2020 model iterations respectively (Tukey post-hoc: p=0.02). However, both Boyce Index and specificity showed nonsignificant (p>0.05) changes between 2015 and 2020 model iterations (0.44 to 0.41, and 92.04 to 95.58, respectively). The predictive performance of the spatial block niche modeling for the

2015 spatial data indicates model overfitting (AUC_{BLOCK}=0.81, AUC_{TEST}=0.47), in contrast to the 2020 spatial data which did not exhibit overfitting (AUC_{BLOCK}=0.47, AUC_{TEST}=0.52).

4. DISCUSSION

Within a five-year period, our iterative modeling approach resulted in the discovery of nine novel locations (representing 39% of the initial known distribution) and a 63 km expansion of the predicted geographical range of a federally threatened perennial forb. The discovery of new locations from model-guided field surveys is frequently reported for rare species in the literature, and highlight the importance of SDMs and model-guided field surveys in conservation. As a result of enlarged occurrence datasets and known ranges, many threatened species have subsequently been delisted from the Endangered Species Act (Keinath et al. 2014, Sofaer et al. 2019). Moreover, with sufficient spatial data, models can reliably identify biologically relevant ecological factors that support species persistence and predict their potential distributions. In this study, the number of *I. webberi* occurrences increased by 39% (from n=23 to n=32) and *I. webberi* patch occupancy in many of the new locations compares well with those of the original locations. Therefore, findings from this study can guide decisions on future *I. webberi* management. Moreover, previous studies have also reported major revisions to conservation management and reserve designs due to the additional biodiversity data from model-guided field surveys (Platts et al. 2010), including decisions regarding translocation of species of conservation concern (Draper et al. 2019). Findings of multiple analyses show that the majority of the new locations are found within the environmental niche of the initial occurrences. We observed high niche stability (93%) and low

342 niche expansion (7%) between the environmental conditions in the initial and new
343 occurrences, while both the initial niche density values of the new locations and the niche
344 dynamics plot (Fig. 5) illustrate the position of the new locations within the realized niche
345 space of the initial occurrences. This is not surprising considering that the field validation
346 surveys that resulted in the discovery of these novel locations were based on initial models.
347 The observed niche overlap and stability estimates are higher than the majority of the
348 randomly generated niches, whereas niche unfilling and expansion are lower than most of the
349 random niches generated in the similarity test (Fig. 6). In spite of the nonsignificant
350 randomization results, these findings provide partial support for niche similarity between the
351 initial and novel occurrences. The marginally significant randomizations ($0.05<p<0.15$) could
352 be attributed to a limited statistical power due to the low number of occurrences and high
353 degree of geographical similarity in both the initial and new datasets (Brown & Carnaval
354 2019). The unfilled portion of the niche (Fig. 5) suggests that there may be more *I. webberi*
355 locations yet to be discovered or suitable habitat yet to be colonized due to the species' limited
356 dispersal capacity.

357 Additional spatial data can significantly impact the predictive performance of iterative
358 niche models, due to their effect on model parameters (Guisan et al. 2006). In this study, we
359 observed changes in model performance and geographical projections, despite the minimal
360 changes in the three predictors used across all model iterations. Specificity is based on
361 omission error rates, which represent the percentage of false negatives in the spatial data.
362 Therefore, slight increases in specificity across the model iterations suggest that the additional
363 spatial data slightly reduced presence-absence ratio in the overall spatial data and also

364 reduced the model omission errors (Lauzeral et al. 2012, Chiffard et al. 2020). However, the
365 reduction of AUC, TSS, and Boyce Index in all but the final model iteration may be attributed
366 to overfitting due to insufficient occurrences in the dataset. Additional spatial datasets from
367 multi-year sampling may have corrected model overfitting, but they also resulted in reduced
368 SDM performance. This is consistent with previous studies that also reported reduced niche
369 performance when correcting overfitting in niche models (Guisan et al. 2006, Peterson et al.
370 2007). Therefore, a fair performance assessment for iterative niche modeling should focus on
371 model generalizability as the primary measure of performance as opposed to model fit for any
372 given year. A rigorous approach to assessing model generalizability (or lack of over-fitting) is
373 to use spatially independent data for model validation, as in the spatial block niche modeling
374 approach employed in this study. Secondly, some of the additional absence points were
375 sampled from areas that were predicted to be suitable. This can introduce noise into spatial
376 data used for iterative niche modeling because the absence of *I. webberi* in these predicted
377 suitable sites may be due to dispersal limitation (Lobo et al. 2010, Lauzeral et al. 2012). Field
378 observations support the suitability of some of these surveyed sites because they have similar
379 edaphic and topographic features and the occurrence of common associates like *Balsamorhiza*
380 *hookeri, Artemisia arbuscula, Antennaria dimorpha*, and *Phlox longifolia*. McCune (2016) reported
381 similar circumstances where common floristic associates of several studied plants were found
382 in sites predicted to be suitable. Therefore, the inclusion of such absence points in iterative
383 niche models can result in the underprediction of the potential niche and a reduction in model
384 performance (Araújo & Peterson 2012).

385 The biology of a species may also affect the predictive performance of niche models
386 (Marmion et al. 2009, Regos et al. 2019), and particularly the performance of iterative SDMs
387 following the addition of new spatial data (Guisan et al. 2006, Lauzeral et al. 2012). Despite its
388 relatively restricted geographical range, *I. webberi* is locally abundant in occurrence locations
389 and it exhibits mixed mating system (Borokini et al. 2021). These traits suggest high
390 colonization potential and wider niche breadth (Grant & Kalisz 2019), which fits the
391 description of satellite-type species (Hanski 1982, Collins et al. 1993). For satellite-type species,
392 low dispersal capacity limits the full colonization of suitable habitat and may reduce
393 predictive performance of SDMs (Edwards et al. 2004). Araújo & Peterson (2012) cautioned
394 that areas of commission errors should be interpreted carefully for species with fewer
395 occurrences because they may represent suitable habitats that are yet to be colonized (that is,
396 potential niche). This may be true for the neo-endemic *I. webberi* which may not yet be in
397 equilibrium with its suitable environment (Araújo & Pearson 2005) because it has not yet fully
398 colonized its range (Kraft et al. 2010, Thornhill et al. 2017). To reduce spatial bias in iterative
399 SDMs, additional spatial data must be collected using stratified sampling from both sites with
400 predicted higher and low probabilities of species occurrence (Edwards et al. 2004, Guisan et al.
401 2006). Additionally, absence points too close to presence points in ordination space (thus
402 sharing similar environmental conditions) should be excluded from subsequent modeling.
403 A combination of biotic and topo-climatic variables contributes to the niche of *I. webberi*.
404 Throughout modeling iterations, perennial herbaceous cover and Topographic Position Index
405 consistently contributed the most to *I. webberi* distribution, while cumulative actual
406 evapotranspiration (AET), Cosine aspect, and summer seasonal precipitation also contributed

to the species niche in model iterations. The perennial herbaceous cover may have constrained *I. webberi* niche to areas of suitable vegetative community, thus representing a biotic component of the species niche. Vegetative land cover is reported in literature as an important predictor of habitat suitability for rare plants (Gogol-Prokurat 2011, McCune 2016). Topographic Position Index illustrates topographic heterogeneity which impacts microclimatic conditions and influences plant distribution and diversity in high-altitude and heterogeneous landscapes (Chardon et al. 2014, Thornhill et al. 2017). The greater probability of *I. webberi* occurrence in areas with higher cosine aspect in the 2018 and 2019 model iterations shows that the species prefers cooler north-facing slopes which receive less sunlight. Though topographic variables are not proximal (Austin 2002), they have been used successfully as spatial delineators, and to represent missing climatic variables especially in high-altitude areas, map species habitat suitability, reduce niche model overprediction, and increase model performance (Lassueur et al. 2006, Fois et al. 2018).

Summer seasonal precipitation and AET, the bioclimatic variables, represent the availability of water and energy which governs the timing of spring regeneration and seed germination in *I. webberi*. Summer seasonal precipitation may play an important role in *I. webberi* seed dispersal, as has been observed for spring-germinating plants in other cold deserts of the world (Chen et al. 2019). Field observations show that *I. webberi* seeds are dispersed by gravity-assisted surface run-off due to summer precipitation, resulting in the colonization of interspace microsites and decolonized roads and trails. This localized seed movement due to summer precipitation was also reported for *I. tweedyi* and *I. lycopodioides* var. *scandularis* (Moseley 1993, Pollak 1997). Taken together, the SDM predictions are congruent

429 with field observations that *I. webberi* suitable habitats are found on gentle slopes and ridges
430 dominated by native perennial forbs, herbs, annual grasses, and fewer stands of native shrubs,
431 interspersed with bare ground or gravel-covered microsites. Unfortunately, these sites are
432 vulnerable to anthropogenic disturbances and colonization by invasive species which have
433 altered wildfire regimes in the Great Basin Desert (Chambers et al. 2014, Morris & Rowe 2014).
434 Species with small population size and restricted geographical distributions are more
435 vulnerable to future environmental changes and are frequently targets of conservation priority
436 (Lomba et al. 2010, Sousa-Silva et al. 2014). In this study, we explored the efficacy of two
437 complementary approaches for addressing the challenges associated with SDMs for rare
438 species: iterative ensemble modeling and model-guided field sampling. These two
439 complementary approaches can reduce spatial bias, allow for model fine tuning that can
440 improve model performance, and increase the chances of detecting novel locations that can
441 either fill the realized niche space or expand the species niche breadth and hence the known
442 geographical distribution. Improved model performance will enhance reliable assessment of
443 species-environment relationships. Iterative SDMs are particularly important for guiding
444 future efforts to improve species distribution datasets and allow for a tighter integration of
445 models with data, leading ultimately to more accurate and ecologically meaningful SDMs.
446

447 **ACKNOWLEDGEMENTS:** This research was funded through a grant from the U.S. Fish and
448 Wildlife Service CFDA program (award number F15AC00784), with additional funding
449 support from the Northern California Botanists (NCB) Botany research scholarship program,
450 and the Society for Conservation Biology (SCB) graduate student research fellowship. The

authors are grateful to the Nevada Natural Heritage Program for supplying the species occurrence data and absence points for modeling. Additional data from Sarah Kulpa of the U.S. Fish and Wildlife Service, Reno Field Office, Valda Lockie of the Bureau of Land Management, Susanville CA Office, and the California Native Plant Society is also appreciated. This work also benefited from expert advice from Dr. Town Peterson, Dr. Olivier Broennimann, August (Tianxiao) Hao, and Damien Georges.

REFERENCES

Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP (2015) spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography 38: 541-545. https://doi.org/10.1111/ecog.01132

Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). J Appl Ecol 43: 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Aranda SC, Lobo JM (2011) How well does presence-only-based species distribution modelling predict assemblage diversity? A case study of the Tenerife flora. Ecography 34: 31–38. https://doi.org/10.1111/j.1600-0587.2010.06134.x

Araújo MB, New M (2007) Ensemble forecasting of species distributions. Trends Ecol Evol 22: 42-47. https://doi.org/10.1016/j.tree.2006.09.010Araújo MB, Pearson RG (2005) Equilibrium of species' distributions with climate. Ecography 28: 693-695. https://doi.org/10.1111/j.2005.0906-7590.04253.x

Araújo MB, Peterson AT (2012) Uses and misuses of bioclimatic envelope modeling. Ecology 93: 1527-1539. https://doi.org/10.1890/11-1930.1

Araújo MB, Rozenfeld A (2014) The geographic scaling of biotic interactions. Ecography 37: 406-415. https://doi.org/10.1111/j.1600-0587.2013.00643.x

Araújo MB, Pearson RG, Thuiller W, Erhard M (2005) Validation of species–climate impact models under climate change. Global Change Biol 11: 1054-1513. https://doi.org/10.1111/j.1365-2486.2005.01000.x

Banerjee AK, Mukherjee A, Guo W, Ng WL, Huang Y (2019) Combining ecological niche modeling with genetic lineage information to predict potential distribution of *Mikania micranth*a Kunth in South and Southeast Asia under predicted climate change. Glob Ecol Conserv 20: e00800. https://doi.org/10.1016/j.gecco.2019.e00800

Beygelzimer A, Kakadet S, Langford J, Arya S, Mount D, Li S (2019) FNN: Fast Nearest Neighbor Search Algorithms and Applications. R package version 1.1.3. https://CRAN.R-project.org/package=FNN

Borokini IT, Klingler KB, Peacock MM (2021) Life in the desert: The impact of geographic and environmental gradients on genetic diversity and population structure of *Ivesia webberi*. Ecol Evol 11: 17537-17556. https://doi.org/10.1002/ece3.8389

Boyce MS, Vernier PR, Nielsen SE, Schmiegelow FKA (2002) Evaluating resource selection functions. Ecol Modell 157(2-3): 281-300. https://doi.org/10.1016/S0304-3800(02)00200-4

Broennimann O, Fitzpatrick MC, Pearman PB, Petitpierre B, Pellissier L, Yoccoz NG, Thuiller W, Fortin M-J, Randin C, Zimmermann NE, Graham CH, Guisan A. (2012) Measuring ecological niche overlap from occurrence and spatial environmental data. Glob Ecol Biogeog 21: 481– 497. https://doi.org/10.1371/journal.pone.0049508

Broennimann O, Thuiller W, Hughes G, Midgley GF, Alkemade JMR, Guisan A (2006) Do geographic distribution, niche property and life form explain plants' vulnerability to global change? Glob Change Biol 12: 1079-1093. https://doi.org/10.1111/j.1365-2486.2006.01157.x

Brown JL, Carnaval AC (2019) A tale of two niches: methods, concepts, and evolution. Front Biogeog 11: e44158. https://doi.org/10.21425/F5FBG44158

Burns PA, Clemann N, White M (2020) Testing the utility of species distribution modelling using Random Forests for a species in decline. Austral Ecol 45: 706-716. https://doi.org/10.1111/aec.12884

Chambers JC, Bradley BA, Brown CS, D'Antonio C, Germino MJ, Grace JB, Hardegree SP, Miller RF, Pyke DA (2014) Resilience to stress and disturbance, and resistance to *Bromus tectorum* L. invasion in cold desert

shrublands of western North America. Ecosystems 17: 360–375. https://doi.org/10.1007/s10021-013-9725-5

Chardon NI, Cornwell WK, Flint LE, Flint AL, Ackerly DD (2014) Topographic, latitudinal and climatic distribution of *Pinus coulteri*: geographic range limits are not at the edge of the climate envelope. Ecography 37: 1–12. https://doi.org/10.1111/ecog.00780

Chen Y, Shi X, Zhang L, Baskin JM, Baskin CC, Liu H, Zhang D (2019) Effects of increased precipitation on the life history of spring- and autumn-germinated plants of the cold desert annual *Erodium oxyrhynchum* (Geraniaceae). AoB PLANTS 11: plz004. https://doi.org/10.1093/aobpla/plz004

Coddington JA, Agnarsson I, Miller JA, Kuntner M, Hormiga G (2009). Undersampling bias: the null hypothesis for singleton species in tropical arthropod surveys. J. Animal Ecol. 78: 573-584. https://doi.org/10.1111/j.1365-2656.2009.01525.x

Collins SL, Glenn SM, Roberts DW (1993) The hierarchical continuum concept. J Veg Sci 4(2): 149-156. https://doi.org/10.2307/3236099

Daly C, Halbleib M, Smith JI, Gibson WP, Doggett MK, Taylor GH, Curtis J, Pasteris PA (2008) Physiographically-sensitive mapping of temperature and precipitation across the conterminous United States. Int J Climat 28: 2031-2064. https://doi.org/10.1002/joc.1688

de Siqueira MF, Durigan G, de Marco Jr. P, Peterson AT (2009) Something from nothing: Using landscape similarity and ecological niche modeling to find rare plant species. J Nat Conserv 17: 25-32. https://doi.org/10.1016/j.jnc.2008.11.001

Dilts TE, Weisberg PJ, Dencker CM, Chambers JC (2015) Functionally relevant climate variables for arid lands: a climatic water deficit approach for modelling desert shrub distributions. J Biogeog 42: 1986-1997. https://doi.org/10.1111/jbi.12561

Draper D, Marques I, Iriondo JM (2019) Species distribution models with field validation, a key approach for successful selection of receptor sites in conservation translocations. Glob Ecol Conserv 19: e00653. https://doi.org/10.1016/j.gecco.2019.e00653

Edwards Jr. TC, Cutler DR, Zimmermann NE, Geiser L, Alegria J (2004) Model-based stratifications for enhancing the detection of rare ecological events. Ecology 86: 1081-1090. https://doi.org/10.1890/04-0608

Elith J, Leathwick JR (2009) Species Distribution Models: ecological explanation and prediction across space and time. Ann Rev Ecol Evol Syst 40: 677-697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Elith J, Ferrier S, Huettmann F, Leathwick JR (2005) The evaluation strip: A new and robust method for plotting predicted responses from species distribution models. Ecol Modell 186: 280-289. https://doi.org/10.1016/j.ecolmodel.2004.12.007

Fois M, Cuena-Lombraña A, Fenu G, Bacchetta G (2018) Using species distribution models at local scale to guide the search of poorly known species: Review, methodological issues and future directions. Ecol Modell 385: 124–132. https://doi.org/10.1016/j.ecolmodel.2018.07.018

Gogol-Prokurat M (2011). Predicting habitat suitability for rare plants at local spatial scales using a species distribution model. Ecol App 21: 33–47. https://doi.org/10.1890/09-1190.1

Gorban A, Pokidysheva L, Smirnova E, Tyukina T (2011) Law of the minimum paradoxes. Bull Math Biol 73: 2013–2044. 10.1007/s11538-010-9597-1

Grant A-G, Kalisz S (2019) Do selfing species have greater niche breadth? Support from ecological niche modeling. Evolution 74: 73-88. https://doi.org/10.1111/evo.13870

Guisan A, Broennimann O, Engler R, Vust M, Yoccoz NG, Lehmann A, Zimmermann NE (2006) Using niche-based models to improve the sampling of rare species. Conserv Biol 20: 501-511. https://doi.org/10.1111/j.1523-1739.2006.00354.x

Guisan A, Petitpierre B, Broennimann O, Daehler C, Kueffer C (2014) Unifying niche shift studies: Insights from biological invasions. Trends Ecol Evol 29: 260–269. https://doi.org/10.1016/j.tree.2014.02.009

Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143: 29–36. https://doi.org/10.1148/radiology.143.1.7063747

Hanski I (1982) Dynamics of regional distribution: the core and satellite species hypothesis. Oikos 38: 210-221. https://doi.org/10.2307/3544021

Harrell FE, Lee KL, Mark DB (1996) Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 15: 361–387. https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4

Hernandez PA, Graham CH, Master LL, Albert DL (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29: 773-785. https://doi.org/10.1111/j.0906-7590.2006.04700.x

Hijmans RJ, Phillips S, Leathwick J, Elith J (2017) Dismo: Species Distribution Modeling. R Package Version 1.0-12. http://CRAN.R-project.org/package=dismo

Hirzel A, Guisan A (2002) Which is the optimal sampling strategy for habitat suitability modelling. Ecol Modell 157(2-3): 331-341. https://doi.org/10.1016/S0304-3800(02)00203-X

Hortal J (2008) Uncertainty and the measurement of terrestrial biodiversity gradients. J Biogeog 35: 1335–1336. https://doi.org/10.1111/j.1365-2699.2008.01955.x

Hutchinson GE (1957) Concluding remarks. Cold Spring Harb Symp Quant Biol 22: 415–427.

Ikeda DH, Max TL, Allan GJ, Lau MK, Shuster SM, Whitham TG (2016) Genetically informed ecological niche models improve climate change predictions. Glob Change Biol 23: 164-176. https://doi.org/10.1111/gcb.13470

Jarnevich CS, Stohlgren TJ, Kumar S, Morisette JT, Holcombe TR (2015) Caveats for correlative species distribution modeling. Ecol Inform 29: 6-15. https://doi.org/10.1016/j.ecoinf.2015.06.007

Jiménez-Valverde A, Peterson AT, Soberón J, Overton JM, Aragón P, Lobo JM (2011) Use of niche models in invasive species risk assessments. Biol Invasions 13: 2785-2797. https://doi.org/10.1007/s10530-011-9963-4

Keinath DA Griscom HR, Andersen MD (2014) Habitat and distribution of the Wyoming pocket gopher (*Thomomys clusius*). J Mammal 95: 803–813. https://doi.org/10.1644/13-MAMM-A-226

Kraft NJ, Baldwin BG, Ackerly DD (2010) Range size, taxon age and hotspots of neoendemism in the California flora. Divers Distrib 16: 403–413. https://doi.org/10.1111/j.1472-4642.2010.00640.x

Kuhn M (2008) Building Predictive models in R using the *caret* package. J Stat Softw 28(5): 1-26. 10.18637/jss.v028.i05

Kursa MB, Rudnicki WR (2010) Feature selection with the *Boruta* package. J Stat Softw 36(11): 1-13. 10.18637/jss.v036.i11

Lassueur T, Joost S, Randin CF (2006) Very high-resolution digital elevation models: Do they improve models of plant species distribution? Ecol Modell 198(1–2): 139-153. https://doi.org/10.1016/j.ecolmodel.2006.04.004

582 Lauzeral C, Grenouillet G, Brosse S (2012) Dealing with noisy absences to optimize species distribution models:
583 An iterative ensemble modelling approach. PLOS ONE 7: e49508.
584 https://doi.org/10.1371/journal.pone.0049508

585 Liu C, Wolter C, Xian W, Jeschke JM (2020) Most invasive species largely conserve their climatic niche. Proc Nat
586 Acad Sci 117: 23643-23651. https://doi.org/10.1073/pnas.2004289117

587 Lobo JM, Jiménez-Valverde A, Hortal J (2010) The uncertain nature of absences and their importance in species
588 distribution modelling. Ecography 33: 103–114. https://doi.org/10.1111/j.1600-0587.2009.06039.x

589 Lomba A, Pellissier L, Randin C, Vicente J, Moreira F, Honrado J, Guisan A (2010) Overcoming the rare species
590 modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. Biol Conserv
591 143: 2647-2657. https://doi.org/10.1016/j.biocon.2010.07.007

592 Marmion M, Parviainen M, Luoto M, Heikkinen RK, Thuiller W (2009) Evaluation of consensus methods in
593 predictive species distribution modelling. Divers Distrib 15: 59-69. https://doi.org/10.1111/j.1472-
594 4642.2008.00491.x

595 McCune JL (2016) Species distribution models predict rare species occurrences despite significant effects of
596 landscape context. J Appl Ecol 53: 1871-1879. https://doi.org/10.1111/1365-2664.12702

597 Morris LR, Rowe RJ (2014) Historical land use and altered habitats in the Great Basin. J Mammal 95: 1144–1156.
598 https://doi.org/10.1644/13-MAMM-S-169

599 Moseley RK (1993) Floristic inventory of subalpine parks in the Coeur d'Alene river drainage, northern Idaho.
600 Report to Cooperative Challenge Cost-share Project Idaho Panhandle National Forests and Idaho
601 Department of Fish and Game; 62 pp.

602 Peterson AT, Papeş M, Eaton M (2007) Transferability and model evaluation in ecological niche modeling: A
603 comparison of GARP and Maxent. Ecography 30: 550–560. https://doi.org/10.1111/j.0906-
604 7590.2007.05102.x

605 Petitpierre B, Kueffer C, Broennimann O, Randin C, Daehler C, Guisan A (2012) Climatic niche shifts are rare
606 among terrestrial plant invaders. Science 335: 1344–1348. 10.1126/science.1215933

607 Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions.
608 Ecol Modell 190: 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

609 Pili AN, Tingley R, Sy EY, Diesmos MLL, Diesmos AC (2020) Niche shifts and environmental non-equilibrium
610 undermine the usefulness of ecological niche models for invasion risk assessments. Sci Rep 10: 7972.
611 https://doi.org/10.1038/s41598-020-64568-2

612 Platts PJ, Ahrends A, Gereau RE, McClean CJ, Lovett JC, Marshall AR, Pellikka PKE, Mulligan M, Fanning E,
613 Marchant R (2010) Can distribution models help refine inventory-based estimates of conservation
614 priority? A case study in the Eastern Arc forests of Tanzania and Kenya. Divers Distrib 16: 628-642.
615 https://doi.org/10.1111/j.1472-4642.2010.00668.x

616 Pollak O (1997) Morphology and dynamics of alpine populations of *Ivesia lycopodioides* ssp. *scandularis* from the
617 White Mountains of California. Univ Cal White Mount Res Station Symp 1: 97– 116.

618 R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical
619 Computing, Vienna, Austria. [online]. Available from URL: http://www.R-project.org/

620 Ramesh V, Gopalakrishna T, Barve S, Melnick DJ (2017) IUCN greatly underestimates threat levels of endemic
621 birds in the Western Ghats. Biol Conserv 210: 205– 221. https://doi.org/10.1016/j.biocon.2017.03.019

622 Regos A, Gagne L, Alcaraz-Segura D, Honrado JP, Domínguez J (2019) Effects of species traits and environmental
623 predictors on performance and transferability of ecological niche models. Sci Rep 9: 4221.
624 https://doi.org/10.1038/s41598-019-40766-5

Rinnhofer LJ, Roura-Pascual N, Arthofer W, Dejaco T, Thaler-Knoflach B, Wachter GA, Christian E, Steiner FM, Schlick-Steiner BC (2012) Iterative species distribution modelling and ground validation in endemism research: An Alpine jumping bristletail example. Biodiv Conserv 21: 2845–2863. https://doi.org/10.1007/s10531-012-0341-z

Särkinen T, Gonzáles P, Knapp S (2013) Distribution models and species discovery: the story of a new *Solanum* species from the Peruvian Andes. PhytoKeys 31: 1–20. https://doi.org/10.3897/phytokeys.31.6312

Singh NJ, Yoccoz NG, Bhatnagar YV, Fox JL (2009) Using habitat suitability models to sample rare species in high-altitude ecosystems: a case study with Tibetan argali. Biodiv Conserv 18: 2893–2908. https://doi.org/10.1007/s10531-009-9615-5

Sofaer HR, Jarnevich CS, Pearse IS, Smyth RL, Auer S, Cook GL, Edwards Jr. TC, Guala GF, Howard TG, Morisette JT, Hamilton H (2019) Development and delivery of species distribution models to inform decision-making. BioScience 69: 544–557. https://doi.org/10.1093/biosci/biz045

Sousa-Silva R, Alves P, Honrado J, Lomba A (2014) Improving the assessment and reporting on rare and endangered species through species distribution models. Global Ecol Conserv 2: 226-237. http://dx.doi.org/10.1016/j.gecco.2014.09.011

Stockwell DRB, Peterson AT (2002) Effects of sample size on accuracy of species distribution models. Ecol Modell 148: 1-13. https://doi.org/10.1016/S0304-3800(01)00388-X

Strubbe D, Broennimann O, Chiron F, Matthysen E (2013) Niche conservatism in non-native birds in Europe: niche unfilling rather than niche expansion. Global Ecol Biogeog 22: 962-970. https://doi.org/10.1111/geb.12050

Svejcar T, Boyd C, Davies K, Hamerlynck E, Svejcar L (2017) Challenges and limitations to native species restoration in the Great Basin, USA. Plant Ecol 218: 81–94. https://doi.org/10.1007/s11258-016-0648-z

Thornhill AH, Baldwin BG, Freyman WA, Nosratinia S, Kling MM, Morueta-Holme N, Madsen TP, Ackerly DD, Mishler BD (2017) Spatial phylogenetics of the native California flora. BMC Biol 15: 96. https://doi.org/10.1186/s12915-017-0435-x

Thuiller W, Lafourcade B, Engler R, Araújo MB (2009) BIOMOD – a platform for ensemble forecasting of species distributions. Ecography 32: 1-5. https://doi.org/10.1111/j.1600-0587.2008.05742.x

Töpel M, Antonelli A, Yesson C, Eriksen B (2012) Past climate change and plant evolution in Western North America: a case study in Rosaceae. PLOS ONE 7: e50358. https://doi.org/10.1371/journal.pone.0050358

U.S. Fish and Wildlife Service [USFWS] (2014) Species Report for *Ivesia webberi* (Webber's ivesia). Nevada Fish and Wildlife Office, Reno. United States Fish and Wildlife Service. Accessed at https://www.fws.gov/nevada/nv_species/documents/webber_ivesia/ivwe_species_report1_8_14_final.pdf

U.S. Geological Survey [USGS] (2017) 1 Arc-second Digital Elevation Models (DEMs) – USGS National Map 3DEP Downloadable Data Collection. U.S. Geological Survey. Accessed at https://www.usgs.gov/core-science-systems/ngp/tnm-delivery/

Valavi R, Elith J, Lahoz-Monfort JJ, Guillera-Arroita G (2019) blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Meth Ecol Evol 10: 225–232. https://doi.org/10.1111/2041-210X.13107

Vaughan IP, Ormerod SJ (2003) Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. Conserv Biol 17: 1601-1611. https://doi.org/10.1111/j.1523-1739.2003.00359.x

Warren DL, Seifert SN (2011) Ecological niche modeling in Maxent: The importance of model complexity and the performance of model selection criteria. Ecol Appl 21(2): 335– 342. https://doi.org/10.1890/10-1171.1

Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution*, 62(11), 2868-2883. https://doi.org/10.1111/j.1558-5646.2008.00482.x

Weiss A (2001) Topographic position and landform analysis. Poster presentation, ESRI User Conference, San Diego, California, USA.

Whittaker RJ, Araújo MB, Paul J, Ladle RJ, Watson JEM, Willis KJ (2005) Conservation biogeography: assessment and prospect. Divers Distrib 11: 3–23. https://doi.org/10.1111/j.1366-9516.2005.00143.x

Williams JN, Seo C, Thorne J, Nelson JK, Erwin S, O'Brien J, Schwartz MW (2009) Using species distribution models to predict new occurrences for rare plants. Divers Distrib 15: 565-576. https://doi.org/10.1111/j.1472-4642.2009.00567.x

Wisz MS, Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A, NCEAS Predicting Species Distributions Working Group (2008) Effects of sample size on the performance of species distribution models. Divers Distrib 14: 763-773. https://doi.org/10.1111/j.1472-4642.2008.00482.x

Witham CW (2000) Current Knowledge and Conservation Status of *Ivesia webberi* A. Gray (Rosaceae), the Webber ivesia, in Nevada. Status Report prepared for the Nevada Natural Heritage Program, Dept. of Conservation and Natural Resources, Carson City, NV. 52pp.

Xian G, Homer C, Meyer D, Granneman B (2013) An approach for characterizing the distribution of shrubland ecosystem components as continuous fields as part of NLCD. *ISPRS* J Photogramm Remote Sens 86: 136-149. https://doi.org/10.1016/j.isprsjprs.2013.09.009

TABLES

Table 1. Iterative niche modeling with increasing number of presence and absence points for *Ivesia webberi*.

Year	Presence points	Raw absence points	Thinned absence	Predictor variables used for final modeling
2015	23	758	53	Perennial herbaceous cover, Topographic Position Index (TPI) and annual evapotranspiration
2018	26	1652	90	Perennial herbaceous cover, TPI and cosine aspect
2019	27	1881	75	Perennial herbaceous cover, TPI and cosine aspect
2020	32	2289	102	Perennial herbaceous cover, TPI and summer mean precipitation

693 **Table 2. Descriptions of six uncorrelated predictor variables used to fit preliminary niche**

694 **models for *Ivesia webberi*. The three predictor variables used for the iterative niche models**

695 **were selected from this pool. All predictors were resampled to 30 m resolution**

Predictor variable	Relationship with species
Actual evapotranspiration (AET)	An estimate of the amount of water removed from an area by both evaporation and transpiration. AET, a direct predictor, is a proxy estimate of plant productivity
Cosine aspect	Higher values indicate north-facing slopes, which receive less sunlight
Perennial herbaceous vegetative cover	A spatial vegetative cover delineation representing native grasses, perennial forbs, and cacti, which includes areas of *I. webberi* distribution. It is considered a representation of biotic interactions and accounting for community assemblage in sites harboring *I. webberi*
Minimum monthly temperature	A direct predictor that potentially influences plant distribution (Araújo & Rozenfeld, 2014). Vegetative and seed regeneration of *I. webberi* are dependent on cold stratification that characterizes late winter and early spring seasons
Summer seasonal precipitation	A direct predictor that potentially influences plant distribution. Summer precipitation causes surface runoffs which facilitate localized gravity-enhanced seed dispersal and colonization of empty niches. Precipitation and temperature in winter and spring seasons influence the phenology of *I. webberi*
Topographic position index (TPI)	A scale-dependent variable describing the elevation of a cell in relation to the mean elevation of the neighboring cells. At the scale of 333 m, TPI distinguishes between mountains and valleys in the study area. The study area is characterized by topographic heterogeneity which can limit dispersal and distribution, and also act as proxy for microclimatic conditions

696

697 **Table 3. Niche density, predicted habitat suitability, and distance of new locations to the nearest neighbor in the**
698 **initial points**

Location name	Finding	Year	Niche density	Predicted habitat suitability	Distance from known location (km)
Wildcat Hill	Opportunistic: discovered by Bureau of Land Management (BLM) staff during land surveys	2018	0.27	0.73	8.07
Unit 6 extension	Model: predicted suitable sites near known location	2018	0.87	0.59	0.38
Smoke Creek Road	Opportunistic: discovered during California Native Plant Society vegetative surveys	2019	0.86	0.37	62.98
Unit 4 extension	Model: high predicted suitability	2020	0.38	0.62	0.03
South end of HJWA	Model: high predicted suitability	2020	0.58	0.64	2.99
HJWA south end #2	Model: suitable sites near known location	2020	0.71	0.30	2.26
Private land discovery #1	Model: suitable sites near known location	2020	0.21	0.21	1.39
Private land discovery #2	Model: suitable sites near known location	2020	0.36	0.28	2.23
New Smoke Creek Road	Model: suitable sites near known location	2020	0.83	0.15	1.22

699

FIGURES

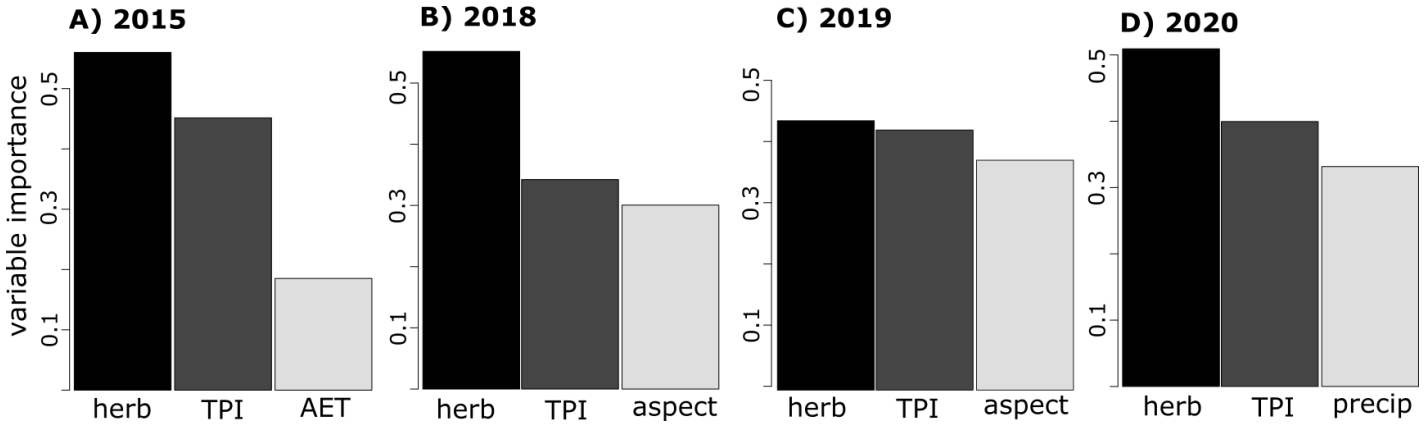

Figure 1. Variable contributions to the iterative niche modeling for *Ivesia webberi* from (a) 2015 to (d) 2020. The three predictors used for each year of iterative modeling were selected from the preliminary modeling. Herb represents the perennial herbaceous vegetative cover, TPI stands for Topographic Position Index at 333 m, AET stands for cumulative actual evapotranspiration, aspect represents cosine-transformed aspect, while precip stands for summer mean precipitation

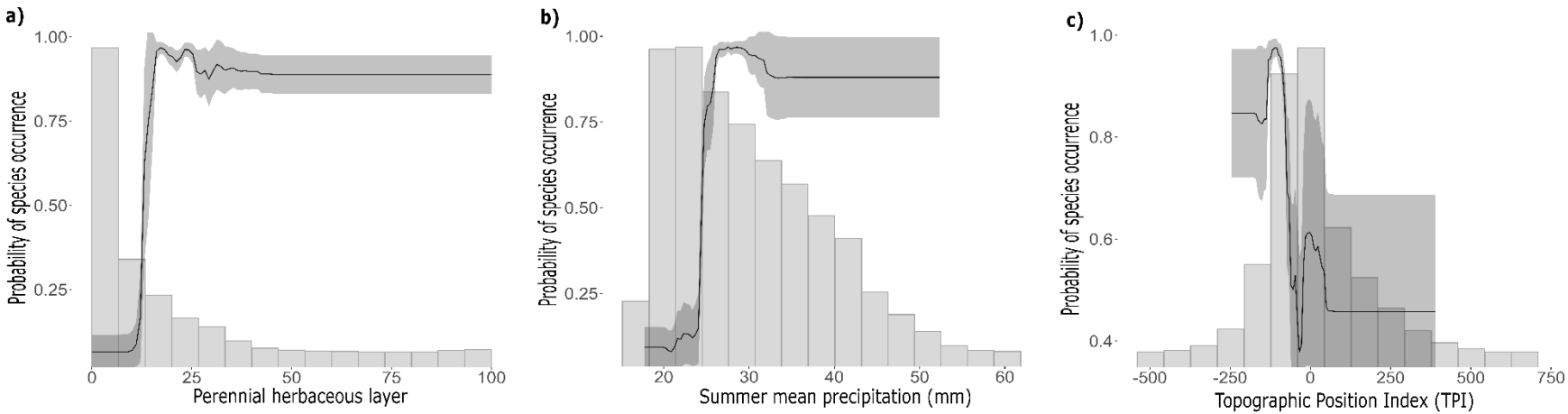

Figure 2. Partial response plots showing the predicted probability of *Ivesia webberi* occurrence in a) Perennial herbaceous vegetative cover, b) summer mean precipitation, and c) Topographic Position Index. The partial response plots were generated using the Boosted Regression Trees, while the histogram represent the predicted values from 10,000 randomly sampled background points from the three variables used for the niche modeling. The partial response plots for each of the 10 model replicates of the six SDM algorithms are included in Figure S1 Supplemental Information.

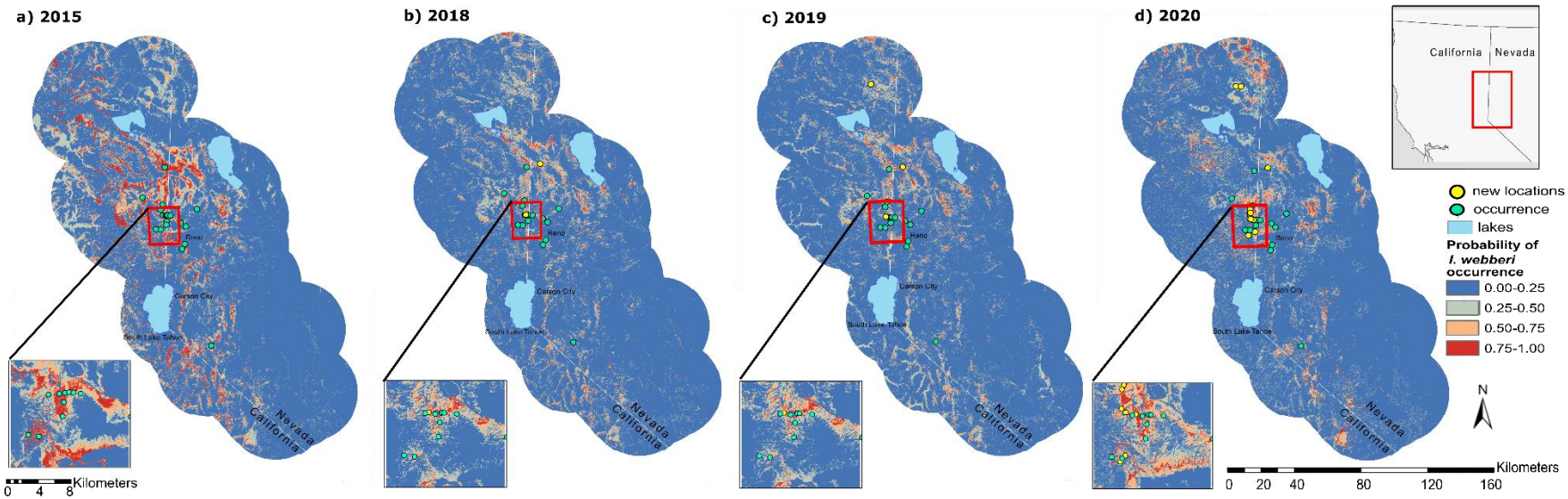

Figure 3. Maps of the predicted geographical distribution of *Ivesia webberi* in the western Great Basin Desert, with both the original and new occurrence points overlay. Red colored pixels represent areas of predicted high probability of *I. webberi* occurrence, orange pixels represent intermediate probability of species occurrence, while blue pixels are predicted areas of zero to low probability of occurrence. The occurrence points in green are the original *I. webberi* occurrence points used for niche modeling, while yellow colored occurrence points represent the novel populations

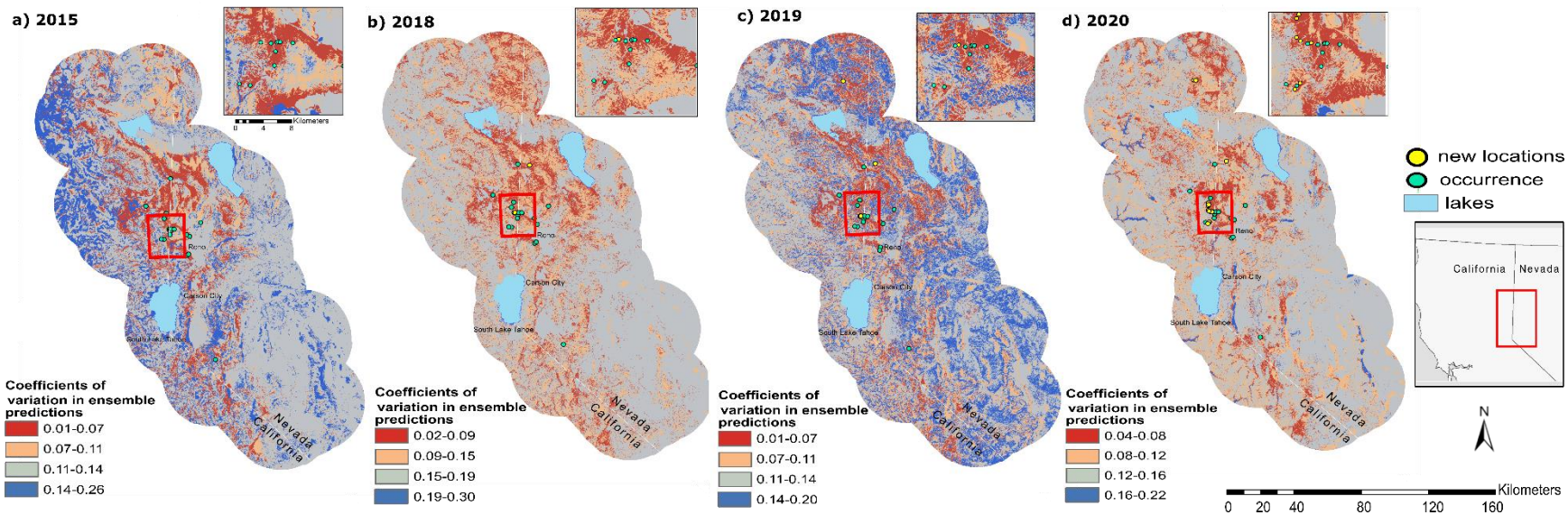

Figure 4. Maps of the coefficients of variation within the ensemble predictions of *Ivesia webberi* in the western Great Basin Desert, with both the original and new occurrence points overlay. Red colored pixels represent areas of low prediction uncertainty, orange pixels represent intermediate prediction uncertainty, while blue pixels are predicted areas of high uncertainty in model predictions.

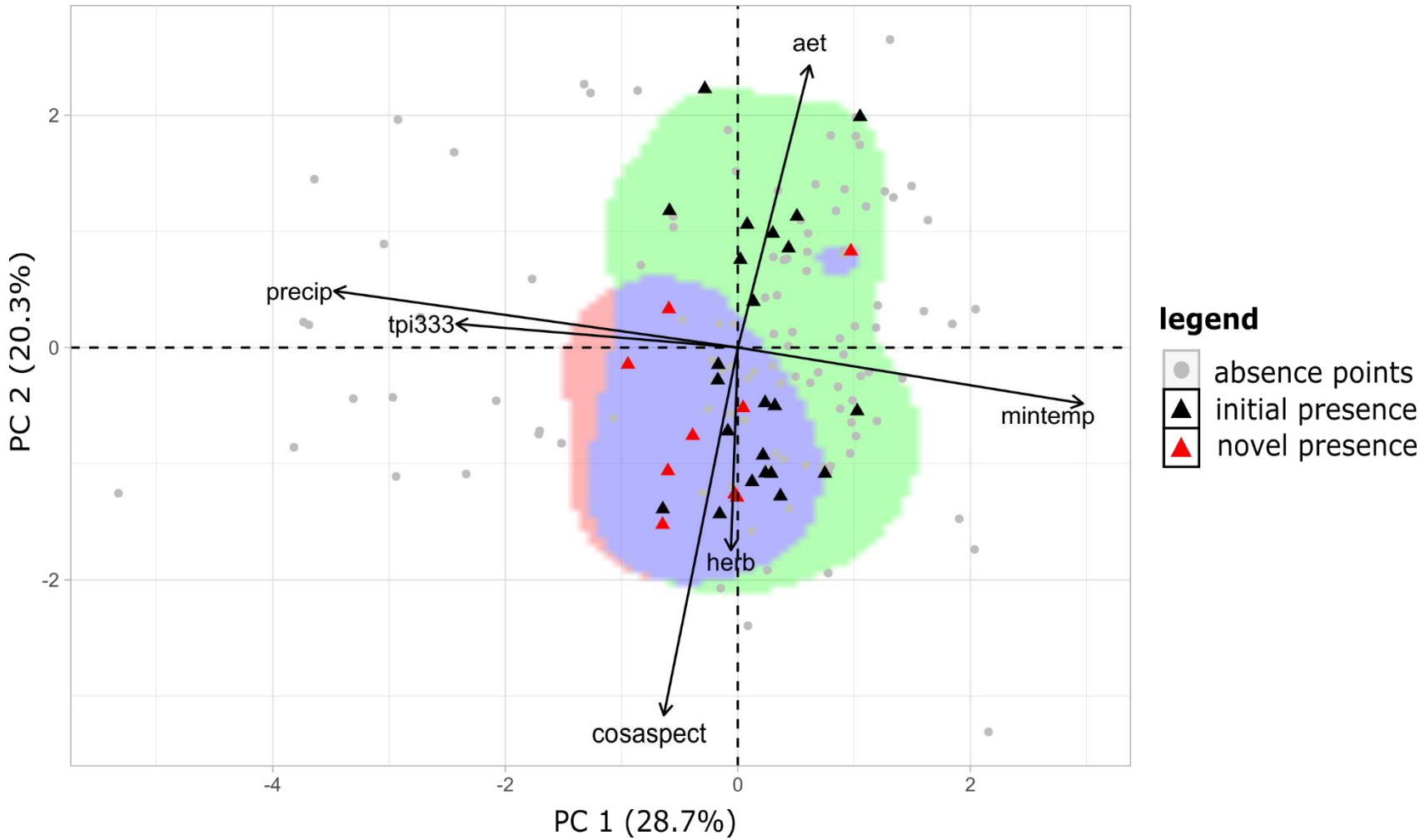

Figure 5. PCA biplot of the environmental predictors that influence *Ivesia webberi* niche in the western Great Basin Desert. These absence and presence data are combinations of initial (black) and new locations (red). The green area on the plot represents the environmental niche occupied only by the initial occurrences (unfilling), while the blue area represents a subset of the initial niche occupied by both initial and novel occurrences (stability) and pink colored areas represent a portion of the niche occupied only by the new locations (expansion).

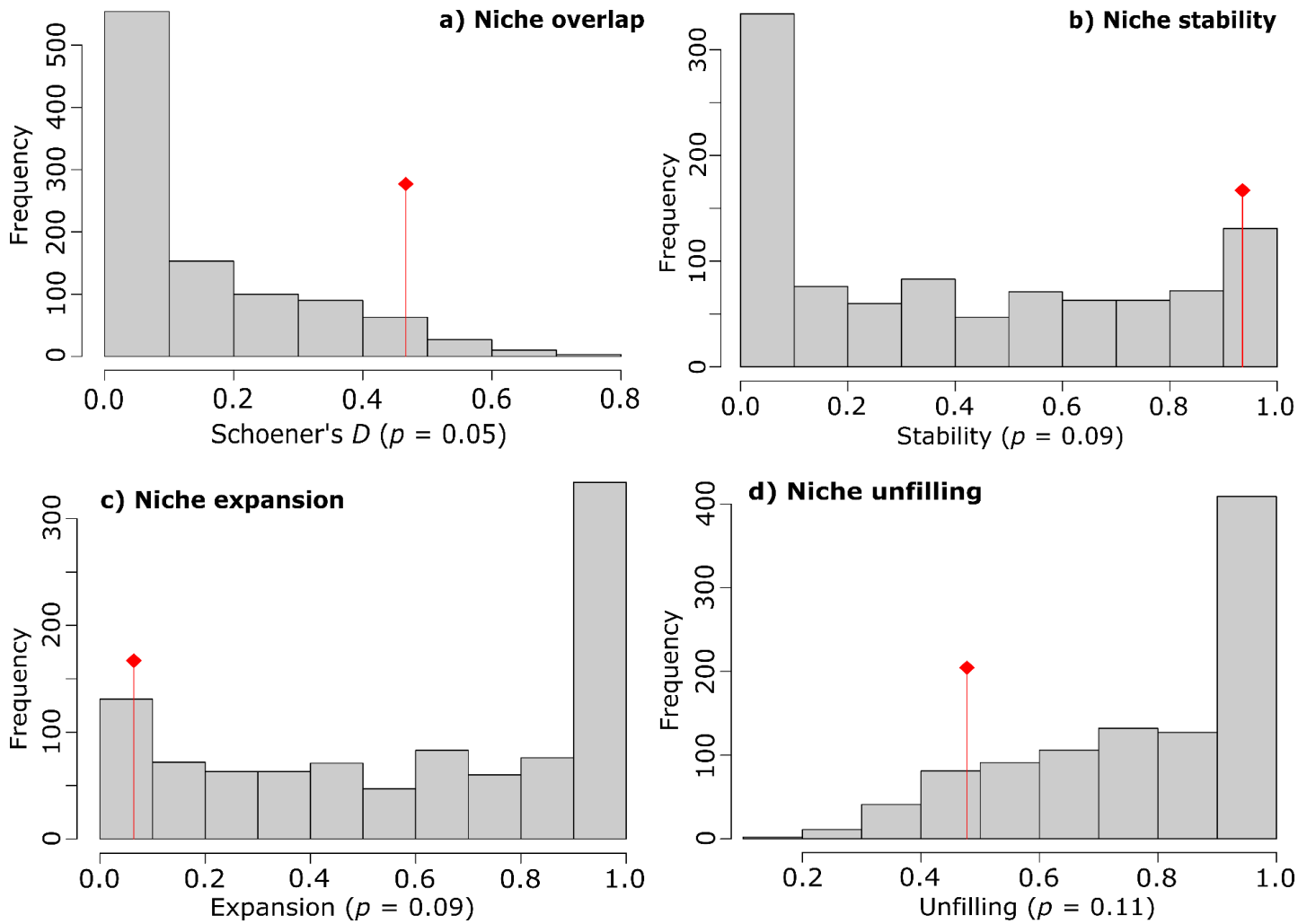

Figure 6. Histogram plots of the randomized values for (a) niche overlap, measured as the Schoener's *D*, (b) niche stability, (c) niche expansion, and (d) niche unfilling between the initial and novel occurrence locations for *Ivesia webberi*. The red bar on each plot represents the actual niche metric. For each niche estimate, 1000 randomizations were done using a niche similarity test that randomly shifts the centroids of the initial and novel realized niches.

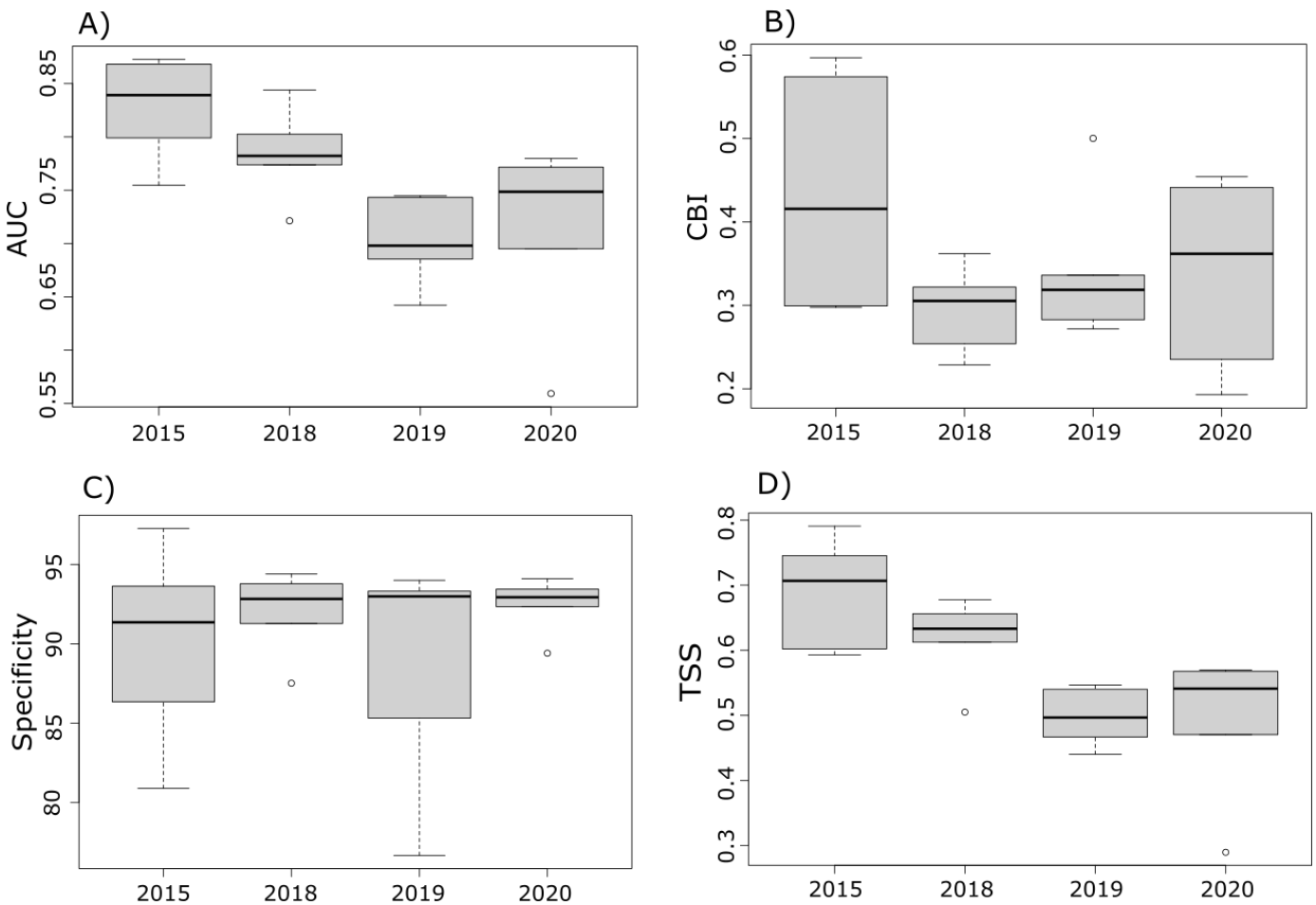

Figure 7. Boxplots of the mean values (n=10 replicates each for six algorithms) of model performance in (a) area under curve (AUC) of the receiver operating characteristic (ROC) plot, (b) Boyce index (BI), (c) specificity, and (d) true skill statistic (TSS) across the years of iterative niche modeling (shown in x axes).