[1] # Predicting, habitat suitability, occupancy, and census sizes of a rare

[2] plant species using iterative adaptive niche based sampling

[3] Reed Clark Benkendorf[1]*, Jeremie B. Fant[1,2], Sophie Taddeo[3]

[1] Chicago Botanic Garden, Glencoe, Illinois 60022, USA

[2] Plant Biology and Conservation, Northwestern University, Evanston, Illinois 60208, USA

[3] Department of Environmental and Ocean Sciences, University of San Diego, San Diego, California 92

[4] **Abstract**

[5] 1)

[6] 2)

[7] 3)

[8] 4)

[9] # 1 | INTRODUCTION

[10] The effects of anthropogenic stressors, e.g. land use and climate change, have lead to a global extinction

[11] crisis with estimates of the number of plant species facing extinction ranging from 20-40% (Brummitt *et al.*

[12] (2015), Pimm & Joppa (2015), Nic Lughadha *et al.* (2020)). Determining which plant species to focus our

[13] conservation efforts (e.g. active restoration, preserve creation, *ex situ* collections) on requires an array of data

[14] which seldom exist for decision makers (Heywood (2017)). These data generally outline simple biological

[15] and ecological parameters of species useful for detailing there rarity and how it's distribution relates to

[16] current and future anthropogenic stressors. Chief amongst these parameters, are the geographic extent of

[17] occurrence (range), the distribution of suitable habitat - the occupancy of this habitat as well as the spacing of

[18] occupied patches, and the census size of individual (sub-)populations (Commission (2001), Faber-Langendoen

[19] *et al.* (2012), *USFWS species status assessment framework* (2016)). While these parameters are relatively

[20] simple, characterizing them can be time consuming, generally requiring extensive field work and travel to

*Author for Correspondence: rbenkendorf@chicagobotanic.org

field sites - hence they oftentimes require proxies or heuristics for estimation in conservation assessments (Juffe-Bignoli *et al.* (2016), Bland *et al.* (2015), Pelletier *et al.* (2018)). Environmental niche models (ENMs or Species Distribution Models SDMs) have made enormous and respectable headway, respectively, in resolving the former two problems (extent, and occupancy) (Syfert *et al.* (2014), Anderson (2023), Kass *et al.* (2021)), however the historic mismatch between the resolution of variables governing species distributions and the data available to serve as predictors of environmental conditions have restricted the interpretation and implementation of these models in highly heterogeneous environments (Guisan *et al.* (2013)). Recent advances in remote-sensing technologies have allowed for the generation of useful models in these generally biodiverse systems, further these data offer promise to model additional population parameters (e.g. census size, population extents), however the utility and usage of high resolution have rarely been ground verified or ostensibly reported upon (Chiffard *et al.* (2020)).

Recently, considerable headway has been made in generating statistically robust environmental niche models (ENM's) spirited by: recent advances in collecting high-resolution environmental data, compute power, digitization of natural history museum records and the acquisition of citizen science records, and statistical methods especially downsampling (Feldman *et al.* 2021; Markham *et al.* 2023). However, ENM's are rarely ground verified, and even more seldom at landscape scales (A. Lee-Yaw *et al.* 2022). Hence, most of our knowledge about producing ENM's rely on simulated species and data, especially at spatial resolutions considered relatively coarse to practitioners.

An historic complication with the implementation and interpretation of ENM's is a mismatch between the spatial resolution of the independent variables available to model the species fundamental niche and the factors governing the true distribution of populations - the realized niche (Lembrechts *et al.* 2019; Carscadden *et al.* 2020; Chauvier *et al.* 2022). Recent papers have had mixed results regarding the effects of imprecisely mapped occurrence data on ENM predictions, with indications that models generated in more heterogenous environments, and at finer resolutions (e.g. ca. 3 arc-second relate to 10 arc-minutes (~14.5 km) at 38*)) suffer minor decreases in model performance (Graham *et al.* 2008; Smith *et al.* 2023) with real species, while increasing error in mapping has drastic effects on model predictions with virtually simulated species (Gábor *et al.* 2022).

A further mismatch of resolution is the year in which data on geographic localities were obtained and current conditions which allow for positive population growth (Bracken *et al.* 2022). Historic occurrence data may now represent conditions which are inhospitable to the maintenance of populations, or may even represent populations which even then were simply sinks from more robust populations (Bracken *et al.* 2022); using these records may decrease model performance. Collecting data on whether areas are favorable to contin-

ued recruitment of individuals from the soil seed bank, are perhaps more astute than whether long-lived individuals persist.

ENM models generally suffer from having few, generally spatially biased, occurrence records to serve as dependent variables, which generally fail to characterize the ecological breadth of the species (Feeley & Silman 2011; Stolar & Nielsen 2015). While many ENM's have high-performance metrics while tested on small subsets of hold-out data, they are unlikely to detect many new populations during ground verification (A. Lee-Yaw *et al.* (2022)). To increase the number of presences, and absences in sites which are relatively similar to those harboring presences, which can be used for training models iterative adaptive-niche based sampling (ANBS) has become increasingly employed (Guisan *et al.* 2006). In ANBS cells predicted to have high probabilities of occurrence, or where models have higher prediction uncertainty, are preferentially visited, and after each bout of field visits, a new model is fit incorporating the original data, plus the recently collected data (Mondain-Monval *et al.* (2024)). Using this process not only allows for the acquisition of a larger number of presences and absences, but also allows for verifying coordinate placement, that historic points are still extent, reducing model uncertainty, and for the acquisition of additional data such as census estimates and life stages (Stockwell & Peterson 2002; Wisz *et al.* 2008; Jansen *et al.* 2022). However, few studies have reported field verified ANBS sampling efforts of plant species [Johnson *et al.* (2023);; Borokini *et al.* (2023); Zimmer *et al.* (2023)].

An ENM predicts a single variable; the probability of suitable habitat for the taxon (taddeo2024grimes). Although, in most conservation applications, the insight desired from an ENM is the species realized distribution (cite). The bridge between an ENM and a plant populations presence is related to the dispersal of propagules and the establishment of the population, rather than distribution of the fundamental niche alone (Engler & Guisan 2009). Calls - and significant progress, have been made for better integration of SDM predictions with models of dispersal since the rise in popularity of ENM's (Guisan & Thuiller (2005), Franklin (2010)). Most of this progress focuses on the goal of 'connecting' currently suitable to habitat predicted to be suitable under climate change scenarios (Engler & Guisan (2009), Bocedi *et al.* (2014)). Accordingly, ground verification has not been made for them, nor has as much effort seemingly been made in attempting to simply improve ground-verification for rare species... To assist practitioners in detecting new populations, or extending the range of currently known populations, we propose modelling plant occupancy as a random variable dependent on distance from sites known to be occupied, the size and shape of the target site, and it's relation to other suitable habitats.

In a simple sense these perspectives link habitat suitability with the tenets of island biogeography (@). 'Distance' may be defined as euclidean (or Haversine for large distances), or as a least-cost distance reflecting

a generalized surface which conveys the difficulty for seeds to travel between the nearest occupied sites and the site of interest (Etherington (2016)).

Landscape metrics postulated to relate to the occupancy at a site include the patch metrics Core Area Index, roughly reflecting the probability of a propagule arriving to a larger patch, and the class metrics proximity and contagion both of which reflect the aggregation between occupied sites.

Obtaining reliable estimates of plant population census sizes can be time intensive process because it requires two major pieces of information 1) measurements of plant density and 2) population extent. In lieu of directly determining either the populations boundaries, or density, observers estimate population sizes, however these estimates are often unreliable (Reisch *et al.* (2018)). Several promising field methods for acquiring density measurements have recently been developed (Schorr 2013; Alfaro-Saiz *et al.* 2019; Rominger & Meyer 2019; Ermakova *et al.* 2021; Krening *et al.* 2021) to better estimate effective population census size (N), as well as progress in the application of genomic methods to estimate effective population sizes (N_e) (e.g. linkage disequilibrium) (Waples 2024). However these methods are focused on descriptive, rather than predictive processes, where they focus on estimating the uncertainty in measurements *within* a population, rather than across the species range. We propose that estimates of plant density are best generated via methods which allow for the fitting of statistical models, which can incorporate spatial covariates and predict estimates and measurements of uncertainty across gridded surfaces (Oliver *et al.* 2012; A. Lee-Yaw *et al.* 2022; Doser *et al.* 2024).

Relationships between facets of ENMs

Suitable Habitat
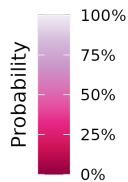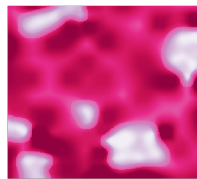
The probability of suitable habitat indicates areas which range from favorable (100%) to inhospitable (0%) for the species, based on the training data. Different techniques are used to interpret this probability, and to apply thresholding to it... This probability can then be used as a covariate in further modelling (e.g. density), or derived for other applications such as generating patch shapes and sizes, noting occupancy, and the distance between occupied and unoccupied patches.
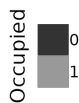
Suitable Patches
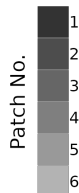
Species are limited to areas of habitat which are large enough to support positive population growth, or which serve as sinks of propagules. Given the density of populations, and size of invididual plants, not all patches have enough area to support an effective population. Further, species vary in their ability to grow on peripheral or marginal habitat around optimal habitat, decreasing the true patch size. Finally smaller patches create smaller "targets" for propagules to disperse to, decrease the probability of plant establishment.

Patch Occupancy

Training data explicitly denote the occupation status of certain areas of suitable habitat. Using these data we can mark patches of suitable habitat which contain known individuals or populations. These occupied patches can be combined with measures of distance to set up an adaptive sampling schema to find new populations.

Distances
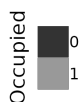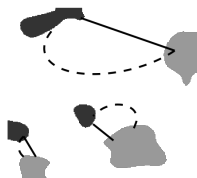
Patches are separated by both geographic (measured as euclidean for short distances or Haversine to accomodate the curvature of the earth) and environmental distances. These distances decrease the likelihood of propagules dispersing between patches. Environmental distances can be estimated by first creating generalized least-cost surfaces and then using various cost algorithms (e.g. electrical conductivity) to link patches known to be occupied with unvisited patches of suitable habitat.

Boundaries
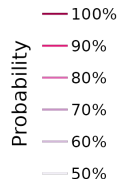
Different thresholds of habitat probability can be used with field verification to most accurately determine where populations end. These boundaries can then be used to refine patch shapes, the distances between them, and to estimate census sizes of the populations in a final modelling process.

Plant Density
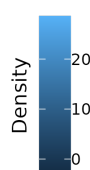
The number (abundance) of plants per specific area can be modelled and projected onto gridded surfaces. This modelling process can involve using the ENM probability surface as a covariate and clipping estimates to predefined boundaries. Summing values from all raster cells across cells can produce an estimate of the number of plants per (sub-)population.

103

5

104 Mapping the boundaries of a population is another time consuming, albeit essential task for realistic estimates

105 of population census size.

106 Currently, population boundaries are implicitly delineated by practitioners walking distances beyond which

107 gene flow between individuals should be minimal (e.g. 1km) searching for additional individuals, or via aerial

108 imagery (Elzinga *et al.* (1998)).

109 Many rare plant species are small and inconspicuous, and many taxa grow in small clustered groups in

110 specific habitats, making detection via either method difficult (Chen *et al.* (2013), Condit *et al.* (2000)).

111 High resolution spatial data eventually offer an ability to determine the extent of individual populations via

112 detection of the edges of suitable habitat.

2 | METHODS

Study Species & System

115 *Eriogonum coloradense* Small (Polygonaceae) is a synoecious mat forming perennial herb endemic to the

116 Central Southern Rocky Mountains in Colorado, U.S.A. It's known geographic range covers XX km^2, with

117 26 formally described populations, and it is thought to occur across a range of elevation, slopes, aspects, soil

118 types and habitats. It's elevation range spans xx - xx m, and the major habitat types it's known from include:

119 high elevation sagebrush steppe, sub-alpine grasslands, and alpine slopes. It is considered an S3/G3 species

120 by NatureServe, and a Tier 2 species by State Wildlife Action Plan by the Colorado Parks and Wildlife

121 Service.

Data Acqusition

123 Dependent data were gathered from iNaturalist (80 records) and the Consortium of Southern Rockies herbaria

124 (131 records) ('Southern rocky mountain herbaria portal', INATURALIST). These data were manually re-

125 viewed and 16 herbarium records which had low geolocation quality, or which were georeferenced to localities

126 which did not match their herbarium labels were removed.

127 Digital Elevation Models at 3arc (type), and 1arc (type), and Digital Elevation Products at 1/3 arc, and 1m,

128 resolution were acquired from the United States Geological Survey and clipped to the domain of analysis (a

129 rectangle buffered 16km (10 mi). beyond any known population). A 3m resolution digital elevation model,

130 which is not available at a native resolution from the USGS for this area, was created by bilinear resampling

131 of the 1m and 10m data. 1m resolution data were only available for roughly XX% of the domain (which

132 contained %% of known populations), the remainder of the species domains 3m data was created by linear

resampling of the 10m DEP, a comparison of these results in the overlapping region are present in appendix xx. These elevation products were used to create all geomorphology data sets using whiteboxTools (Wu & Brown (2022)). Vegetation cover data were made by combining the raster data into continuous covers: Forested, Shrub, and Herbaceous vegetation Tuanmu & Jetz (2014).

ClimateNA was used to create a data set at 3 arc-second resolution which then underwent simple bilinear interpolation to generate products at the finer resolutions (Wang *et al.* (2016)). Gray co-occurrence level matrices were produced using the glcm r package using 2023 NAIP aerial imagery, which underwent bilinear resampling to each resolution, using default setting but with windows of 5 in both directions (Zvoleff (2020)).

Ground Verification

The first round of ground verification was carried out from June-September 2024. All pre-existing occurrence points were considered candidates for revisits and all 24 trails leading to them were marked as SAMPLE UNITS. Each trail was manually mapped, and buffered 45m in each direction and XXX random points were drawn, thinned to distances > XXXm, leaving XXX random plots for assessment. XX trails were visited, allowing for the assessment of XX random points and XX occurrences. When conducting field work, all presences of *E. coloradense* were opportunistically noted, and to better describe the spread of the population points were subjectively placed, with consdirations of local abudance, ca. 30-50m from the previous one until traversing out of the population (n =). Additionally subjectively placed absences were also collected in areas which seemed favorable, or were in close proximity, to *E. coloradense* individuals; this occurred both in field (n =) and through use of aerial imagery on a computer afterwards (n =).

Adaptive Niche-Based Sampling was carried out in July of 2025. To determine whether adaptive sampling performed better than alternative sampling regimes (e.g. stratified, random), and if it could be improved by occupancy modelling 30 points were selected for each of the aforementioned sampling schema plus 30 occupancy points.

Comparision of Different Spatial Resolutions

Environmental Niche Models were generated at four spatial resolutions, 3 (~72m), 1 (~24m), and 1/3 (~8m) arc-seconds, and 3m.

The details of modelling were similar for each resolution.

Records were thinned to the distance of an hypotenuse of a cell to avoid replicates (Aiello-Lammens *et al.* (2015)). 1.1 times as many Absence, as presence, records were generated using the background function with method environmental distance from sdm (Naimi & Araujo (2016)), these records were then manually

reviewed and six records which were deemed in areas which may be possible presences were removed, after this the records were randomly sampled to reduce the data set size to XX records.

After the first iteration of modelling all additional presences and absences were thinned via a similar manner and combined with the original absence records. 'Presences' which had greater coordinate uncertainty than the resolution of modelling were removed.

All random forest models were fit using the ranger package, 95% confidence intervals were generated using the infinitesimal jackknife (Wager *et al.* (2014), Wright & Ziegler (2017)). The generation of CI's for the 3m data set were not considered feasible and their prediction was omitted.

Adaptive and Occupancy Based Field Sampling

500 stratified points ranging from 1-100% probability of suitable habitat were generated using sample (terra). Occupancy scores for each of these points were then calculated using (whiteboxtools). 200 points which maximized the spread of values along both dimensions (habitat, occupancy) were then chosen for ground verification.

Plant Density

Plant density was modelled using five methods, a generalized linear model (GLM) with a poisson error distribution and spatial autocorrelation structures, spAbundance, Random Forest and XgBoost regression.

Species Occupancy

Comparision of Juvenile and Mature Plant Models

Using the top performing model resolution () models were refit using only either juvenile or mature plant presences, the number of mature presences were limited to the number of juvenile occurrences. Models were fit at 6 sample sizes (n = 15, 30, 50, 75, 125, 200) 15 times each using a randomly sampled 60%-40% train-test split of data.

Simulations of Sample Size

The effect of sample size on model performance was simulated at 8 sample sizes (n = 15, 30, 50, 75, 125, 200, 300, 400) each 25 times, using a randomly sampled 60%-40% train-test split of data.

Simulations of Coordinate Errors

The effect of sample size on model performance was simulated at 6 sample sizes (n = 15, 30, 50, 75, 125, 200), with three proportions of records in error (0%, 5%, 10%, 20%), and 3 levels of coordinate uncertainty (0m, 10m, 100m, 1000m), at each of the 5 resolutions 25 times.

3 | RESULTS

Comparision of Different Spatial Resolutions

The area under the receiver operating curve (AUC-ROC) and precision-recall curve (AUC-PR), the True Skill Statistic, and Sensitivity and Specificity (Sofaer *et al.* (2019), Allouche *et al.* (2006)).

Ground Verification

Plant Density

Species Occupancy

Comparision of Juvenile and Mature Plant Models

Simulations of Sample Size

Simulations of Coordinate Errors

4 | DISCUSSION

5 | CONCLUSIONS

6 | ACKNOWLEDGMENTS

REFERENCES

A. Lee-Yaw, J., L. McCune, J., Pironon, S. & N. Sheth, S. (2022). Species distribution models rarely predict the biology of real populations. *Ecography*, **2022**, e05877.

Aiello-Lammens, M.E., Boria, R.A., Radosavljevic, A. & Vilela, B. (2015). spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, **38**, 541–545.

Alfaro-Saiz, E., Granda, V., Rodrìguez, A., Alonso-Redondo, R. & Garcìa-González, M.E. (2019). Optimal census method to estimate population sizes of species growing on rock walls: The case of mature primula pedemontana. *Global Ecology and Conservation*, **17**, e00563.

Allouche, O., Tsoar, A. & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of applied ecology*, **43**, 1223–1232.

Anderson, R.P. (2023). Integrating habitat-masked range maps with quantifications of prevalence to estimate area of occupancy in IUCN assessments. *Conservation Biology*, **37**, e14019.

Bland, L.M., Orme, C.D.L., Bielby, J., Collen, B., Nicholson, E. & McCarthy, M.A. (2015). Cost-effective assessment of extinction risk with limited information. *Journal of Applied Ecology*, **52**, 861–870.

Bocedi, G., Palmer, S.C., Pe'er, G., Heikkinen, R.K., Matsinos, Y.G., Watts, K. & Travis, J.M. (2014). Range s hifter: A platform for modelling spatial eco-evolutionary dynamics and species' responses to environmental changes. *Methods in Ecology and Evolution*, **5**, 388–396.

Borokini, I.T., Nussear, K., Petitpierre, B., Dilts, T.E. & Weisberg, P.J. (2023). Iterative species distribution modeling results in the discovery of novel populations of a rare cold desert perennial. *Endangered Species Research*, **50**, 47–62.

Bracken, J.T., Davis, A.Y., O'Donnell, K.M., Barichivich, W.J., Walls, S.C. & Jezkova, T. (2022). Maximizing species distribution model performance when using historical occurrences and variables of varying persistency. *Ecosphere*, **13**, e3951.

Brummitt, N.A., Bachman, S.P., Griffiths-Lee, J., Lutz, M., Moat, J.F., Farjon, A., Donaldson, J.S., Hilton-Taylor, C., Meagher, T.R., Albuquerque, S. & others. (2015). Green plants in the red: A baseline global assessment for the IUCN sampled red list index for plants. *PloS one*, **10**, e0135152.

Carscadden, K.A., Emery, N.C., Arnillas, C.A., Cadotte, M.W., Afkhami, M.E., Gravel, D., Livingstone, S.W. & Wiens, J.J. (2020). Niche breadth: Causes and consequences for ecology, evolution, and conservation. *The Quarterly Review of Biology*, **95**, 179–214.

Chauvier, Y., Descombes, P., Guéguen, M., Boulangeat, L., Thuiller, W. & Zimmermann, N.E. (2022). Resolution in species distribution models shapes spatial patterns of plant multifaceted diversity. *Ecography*, **2022**, e05973.

Chen, G., Kéry, M., Plattner, M., Ma, K. & Gardner, B. (2013). Imperfect detection is the rule rather than the exception in plant distribution studies. *Journal of Ecology*, **101**, 183–191.

Chiffard, J., Marciau, C., Yoccoz, N.G., Mouillot, F., Duchateau, S., Nadeau, I., Fontanilles, P. & Besnard, A. (2020). Adaptive niche-based sampling to improve ability to find rare and elusive species: Simulations and field tests. *Methods in Ecology and Evolution*, **11**, 899–909.

Commission, N.Resources.S.S. (2001). *IUCN red list categories and criteria*. IUCN.

Condit, R., Ashton, P.S., Baker, P., Bunyavejchewin, S., Gunatilleke, S., Gunatilleke, N., Hubbell, S.P., Foster, R.B., Itoh, A., LaFrankie, J.V. & others. (2000). Spatial patterns in the distribution of tropical tree species. *Science*, **288**, 1414–1418.

246 Doser, J.W., Finley, A.O., Kéry, M. & Zipkin, E.F. (2024). spAbundance: An r package for single-species

247   and multi-species spatially explicit abundance models. *Methods in Ecology and Evolution*, **15**, 1024–1033.

248 Elzinga, C.L., Salzer, D.W. & Willoughby, J.W. (1998). *Measuring & monitoring plant populations.* U.S.

249   Dept. of the Interior, Bureau of Land Management, Denver, Colo.

250 Engler, R. & Guisan, A. (2009). MigClim: Predicting plant distribution and dispersal in a changing climate.

251   *Diversity and distributions*, **15**, 590–601.

252 Ermakova, A., Trout, K., Terry, M.K., Whiting, C.V., Clubbe, C. & Fowler, N. (2021). Densities, plant

253   sizes, and spatial distributions of six wild populations of lophophora williamsii (cactaceae) in texas, USA.

254   *Journal of the Botanical Research Institute of Texas*, **15**, 149–160.

255 Etherington, T.R. (2016). Least-cost modelling and landscape ecology: Concepts, applications, and oppor-

256   tunities. *Current Landscape Ecology Reports*, **1**, 40–53.

257 Faber-Langendoen, D., Nichols, J., Master, L., Snow, K., Tomaino, A., Bittman, R., Hammerson, G., Heidel,

258   B., Ramsay, L., Teucher, A. & others. (2012). NatureServe conservation status assessments: Methodology

259   for assigning ranks. *NatureServe, Arlington, VA*.

260 Feeley, K.J. & Silman, M.R. (2011). Keep collecting: Accurate species distribution modelling requires more

261   collections than previously thought. *Diversity and distributions*, **17**, 1132–1140.

262 Feldman, M.J., Imbeau, L., Marchand, P., Mazerolle, M.J., Darveau, M. & Fenton, N.J. (2021). Trends and

263   gaps in the use of citizen science derived data as input for species distribution models: A quantitative

264   review. *PloS one*, **16**, e0234587.

265 Franklin, J. (2010). Moving beyond static species distribution models in support of conservation biogeogra-

266   phy. *Diversity and Distributions*, **16**, 321–330.

267 Gábor, L., Jetz, W., Lu, M., Rocchini, D., Cord, A., Malavasi, M., Zarzo-Arias, A., Barták, V. & Moudrỳ,

268   V. (2022). Positional errors in species distribution modelling are not overcome by the coarser grains of

269   analysis. *Methods in Ecology and Evolution*, **13**, 2289–2302.

270 Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Townsend Peterson, A., Loiselle, B.A. & Group,

271   N.P.S.D.W. (2008). The influence of spatial errors in species occurrence data used in distribution models.

272   *Journal of Applied Ecology*, **45**, 239–247.

273 Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N.G., Lehmann, A. & Zimmermann, N.E.

274   (2006). Using niche-based models to improve the sampling of rare species. *Conservation biology*, **20**,

275   501–511.

276 Guisan, A. & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models.

277   *Ecology letters*, **8**, 993–1009.

278 Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I., Regan,

T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C. & others. (2013). Predicting species distributions for conservation decisions. *Ecology letters*, **16**, 1424–1435.

Heywood, V.H. (2017). Plant conservation in the anthropocene–challenges and future prospects. *Plant diversity*, **39**, 314–330.

Jansen, J., Woolley, S.N., Dunstan, P.K., Foster, S.D., Hill, N.A., Haward, M. & Johnson, C.R. (2022). Stop ignoring map uncertainty in biodiversity science and conservation policy. *Nature Ecology & Evolution*, **6**, 828–829.

Johnson, S., Molano-Flores, B. & Zaya, D. (2023). Field validation as a tool for mitigating uncertainty in species distribution modeling for conservation planning. *Conservation Science and Practice*, **5**, e12978.

Juffe-Bignoli, D., Brooks, T.M., Butchart, S.H., Jenkins, R.B., Boe, K., Hoffmann, M., Angulo, A., Bachman, S., Böhm, M., Brummitt, N. & others. (2016). Assessing the cost of global biodiversity and conservation knowledge. *PLoS One*, **11**, e0160640.

Kass, J.M., Meenan, S.I., Tinoco, N., Burneo, S.F. & Anderson, R.P. (2021). Improving area of occupancy estimates for parapatric species using distribution models and support vector machines. *Ecological Applications*, **31**, e02228.

Krening, P.P., Dawson, C.A., Holsinger, K.W. & Willoughby, J.W. (2021). A sampling-based approach to estimating the minimum population size of the federally threatened colorado hookless cactus (sclerocactus glaucus). *Natural Areas Journal*, **41**, 4–10.

Lembrechts, J.J., Nijs, I. & Lenoir, J. (2019). Incorporating microclimate into species distribution models. *Ecography*, **42**, 1267–1279.

Markham, K., Frazier, A.E., Singh, K.K. & Madden, M. (2023). A review of methods for scaling remotely sensed data for spatial pattern analysis. *Landscape Ecology*, **38**, 619–635.

Mondain-Monval, T., Pocock, M., Rolph, S., August, T., Wright, E. & Jarvis, S. (2024). Adaptive sampling by citizen scientists improves species distribution model performance: A simulation study. *Methods in Ecology and Evolution*.

Naimi, B. & Araujo, M.B. (2016). Sdm: A reproducible and extensible r platform for species distribution modelling. *Ecography*, **39**, 368–375.

Nic Lughadha, E., Bachman, S.P., Leão, T.C., Forest, F., Halley, J.M., Moat, J., Acedo, C., Bacon, K.L., Brewer, R.F., Gâteblé, G. & others. (2020). Extinction risk and threats to plants and fungi. *Plants, People, Planet*, **2**, 389–408.

Oliver, T.H., Gillings, S., Girardello, M., Rapacciuolo, G., Brereton, T.M., Siriwardena, G.M., Roy, D.B., Pywell, R. & Fuller, R.J. (2012). Population density but not stability can be predicted from species distribution models. *Journal of Applied Ecology*, **49**, 581–590.

Pelletier, T.A., Carstens, B.C., Tank, D.C., Sullivan, J. & Espíndola, A. (2018). Predicting plant conservation priorities on a global scale. *Proceedings of the National Academy of Sciences*, **115**, 13027–13032.

Pimm, S.L. & Joppa, L.N. (2015). How many plant species are there, where are they, and at what rate are they going extinct? *Annals of the Missouri Botanical Garden*, **100**, 170–176.

Reisch, C., Schmid, C. & Hartig, F. (2018). A comparison of methods for estimating plant population size. *Biodiversity and conservation*, **27**, 2021–2028.

Rominger, K. & Meyer, S.E. (2019). Application of UAV-based methodology for census of an endangered plant species in a fragile habitat. *Remote Sensing*, **11**, 719.

Schorr, R.A. (2013). Using distance sampling to estimate density and abundance of saussurea weberi hultén (weber's saw-wort). *The Southwestern Naturalist*, **58**, 378–383.

Smith, A.B., Murphy, S.J., Henderson, D. & Erickson, K.D. (2023). Including imprecisely georeferenced specimens improves accuracy of species distribution models and estimates of niche breadth. *Global Ecology and Biogeography*, **32**, 342–355.

Sofaer, H.R., Hoeting, J.A. & Jarnevich, C.S. (2019). The area under the precision-recall curve as a performance metric for rare binary events. *Methods in Ecology and Evolution*, **10**, 565–577.

Southern rocky mountain herbaria portal.

Stockwell, D.R. & Peterson, A.T. (2002). Effects of sample size on accuracy of species distribution models. *Ecological modelling*, **148**, 1–13.

Stolar, J. & Nielsen, S.E. (2015). Accounting for spatially biased sampling effort in presence-only species distribution modelling. *Diversity and Distributions*, **21**, 595–608.

Syfert, M.M., Joppa, L., Smith, M.J., Coomes, D.A., Bachman, S.P. & Brummitt, N.A. (2014). Using species distribution models to inform IUCN red list assessments. *Biological Conservation*, **177**, 174–184.

Tuanmu, M.-N. & Jetz, W. (2014). A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. *Global Ecology and Biogeography*, **23**, 1031–1045.

USFWS species status assessment framework: An integrated analytical framework for conservation 3.4. (2016). U.S. Fish; Wildlife Service, 5275 VA-7, Falls Church, VA 22041.

Wager, S., Hastie, T. & Efron, B. (2014). Confidence intervals for random forests: The jackknife and the infinitesimal jackknife. *The journal of machine learning research*, **15**, 1625–1651.

Wang, T., Hamann, A., Spittlehouse, D. & Carroll, C. (2016). Locally downscaled and spatially customizable climate data for historical and future periods for north america. *PloS one*, **11**, e0156720.

Waples, R.S. (2024). Practical application of the linkage disequilibrium method for estimating contemporary effective population size: A review. *Molecular Ecology Resources*, **24**, e13879.

Wisz, M.S., Hijmans, R., Li, J., Peterson, A.T., Graham, C., Guisan, A. & Group, N.P.S.D.W. (2008).

345   Effects of sample size on the performance of species distribution models. *Diversity and distributions*, **14**,

346   763–773.

347 Wright, M.N. & Ziegler, A. (2017). ranger: A fast implementation of random forests for high dimensional

348   data in C++ and R. *Journal of Statistical Software*, **77**, 1–17.

349 Wu, Q. & Brown, A. (2022). *'Whitebox': 'WhiteboxTools' r frontend.*

350 Zimmer, S.N., Holsinger, K.W. & Dawson, C.A. (2023). A field-validated ensemble species distribution

351   model of eriogonum pelinophilum, an endangered subshrub in colorado, USA. *Ecology and Evolution*, **13**,

352   e10816.

353 Zvoleff, A. (2020). *Glcm: Calculate textures from grey-level co-occurrence matrices (GLCMs).*