Predicting, habitat suitability, habitat occupancy, and census sizes of a rare plant species using iterative adaptive niche based sampling

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 921

Abstract

1)

2)

3)

4)

1 | INTRODUCTION

The effects of anthropogenic stressors, e.g. land use and climate change, have lead to a global extinction crisis with estimates for the number of plant species facing extinction ranging from X-X%. Determining which plant species to focus our conservation efforts (e.g. . . .) towards requires an array of data which seldom exist for decision makers. These data generally outline simple parameters of the species, detailing the rarity of it, and how it's distribution relates to anthropogenic stressors. Chief amongst these parameters, are the geographic extent (range), the distribution of suitable habitat, and the occupation of suitable habitats by the species, as well as the geographic and census size of individual populations. While simple, estimating these parameters is expensive, generally requiring ground verified data - hence they often require proxies or heuristics for estimation. Environmental niche models have made enormous headway in resolving the former two problems, however the historic mismatch between the resolution of variables governing species distributions and the data available to serve as predictors of environmental have restricted the interpretation and implementation

*Author for Correspondence: rbenkendorf@chicagobotanic.org

of these models. Recent advances in remote-sensing technologies have allowed for well-designed ground verification campaigns, and efforts made towards better understanding parameters of individual populations e.g. both the geographic and census sizes of them. However, the acquisition of ground verified data sets to guide our modelling efforts are still ostensibly rare, likely given the expense and difficulty in obtaining them.

ENM's and mismatch between spatial and temporal resolutions 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 perhaps most importantly - statistical methods. *However, ENM's are rarely ground verified, and even more seldom at landscape scales - further many recent advances, especially statistically, have recently been made with individual populations of common species rather than rare species.* A complication with the implementation and interpretation of these models is a fundamental mismatch between the spatial resolution of the independent variables and the factors governing the distribution of populations. Recent advances in remote sensing and statistical downsampling (e.g. ClimateNA, 3DEP) are bridging this gap, however the resolution of these data are now oftentimes finer than the occurrence data used as dependent variables. Using low quality geo-located data can lead to spurious results, negating the effect of high resolution independent variables *MORE*. 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. Indeed, 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. Collecting data on whether areas are favorable to continued recruitment of individuals from the soil seed bank, are perhaps more astute than whether long-lived individuals persist.

Improving Model Performance - Iterative adaptive-nice based sampling after expert mapping and sampling. 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. 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 verified To increase the number of presences, and absences in ecologically 'similar' sites, which can be used for training models iterative adaptive-niche based sampling (IANBS) has become increasingly employed. In IANBS cells with high probabilities of occurrence are preferentially visited, and after each bout of field visits, a new model is fit incorporating the original data, plus the recently collected data.

However, we posit that evaluating the effects of IANBS is complicated as the species are oftentimes initially 'under surveyed'; in these instances the true distribution of the population is generally poorly defined - despite it being easy to do so by a naturalist. Here we showcase the usage of a robust first-stage expert subjective sampling effort to generate a large pool of random variables for elaboration after a second modelling bout. Using this process not only allows for the acquisition of a large number of presences, but also absences in the areas in geographic and ecological similarity to these presences. It further allows for verification of the placement of the dependent variables, which are likely to adversely affect each generation of models.

Prioritizing Survey Efforts - Occupied Sites An ENM predicts a single outcome; the probability of suitable habitat for the species. Although, in general, an analysts true feature of interest is the species realized distribution.

However, the bridge between an ENM and a populations presence is related to the dispersal of propagules and the establishment of the population, rather than factors intrinsic to the species ecology.

To further assist groups in detecting new populations, or extending the range of currently known populations, we propose modelling occupancy as a random variable dependent on distance from sites known to be occupied, and the landscape configuration of the target site. ‘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 typical seeds to travel between the nearest occupied sites and the site of interest.

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.

< p5 > **Estimating deme borders and census sizes** Obtaining reliable estimates of plant population census sizes can be a time consuming task, requiring two major pieces of information 1) measurements of density, and 2) population extent. Generally these values are obtained via the use of many long (50m) transects, which may require multiple reads and data recorders.

In addition to these personnel requirements, many plant species are endemic to steep, and often loose, slopes, prohibiting the use of transects. Given these complications, many now prefer to obtain these data via genomic methods (e.g. linkage disequilibrium), a similarly expensive task. Here we estimate plant density gleaned during preliminary surveys by a naturalist using simple statistical techniques. Precisely mapping the boundaries of a population is another time consuming task, but essential for the effective estimates of population census size. Currently, population boundaries are generally delineated by have practitioners walk distances (e.g. 1km) in several directions searching for more plants. However, as the distance from a central location increases the amount of area to survey increases IN A FASHION... This is essentially problematic for rare species with many small clusters of individuals, which require considerable survey effort to chance upon.

2 | METHODS

Study Species & System

Eriogonum coloradense Small (Polygonaceae) is a synoecious mat forming perennial herb endemic to the Central Southern Rocky Mountains in Colorado, U.S.A. It's known geographic range covers XX km^2, has 26 formally described populations, and is thought to occur across a range of elevation, slopes, aspects, soil types and habitats. The elevation range from which it is xx - xx, and the broad habitat types it's known from include: high elevation sagebrush steppe, sub-alpine grasslands, and alpine slopes. It is treated as an S3/G3 species by NatureServe, and a Tier 2 species by State Wildlife Action Plan by the Colorado Parks and Wildlife Service.

Data Acqusition

Dependent data were gathered from iNaturalist (XX records) and the Consortium of Southern Rockies herbaria (XX records) ('Southern rocky mountain herbaria portal'). These data were manually reviewed and 16 herbarium records which had low geolocation quality, or which were place in localities which did not match their herbarium labels were removed.

Digital Elevation Models at 3arc (type), and 1arc (type), and Digital Elevation Products at 1/3 arc, and 1m (), resolution were acquired from the United States Geological Survey and clipped to the domain of analysis (a rectangle buffered 10 mi. beyond any known population). The 3m dataset was created by bilinearly re-sampling the 1m DEP. These elevation products were used to create all geomorphology datasets using whiteboxTools. 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 (). Gray co-occurrence level matrices were produced using the glcm r package using 2023 NAIP aerial imagery, which underwent bilinear resampling to resolution, using default settings, but with windows of 5 in both directions ().

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 X 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 ca. 30-50m from the previous one until passing 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…

Comparision of Different Spatial Resolutions

Environmental Niche Models were generated at five different spatial resolutions, 3 (~72m), 1 (~24m), and 1/3 (~8m) arc-seconds, and 3 and 1m resolution. The details of modelling were similar for each resolution.

Records were thinned to the distance of an hypotenuse of a cell to avoid replicates (spThin). XXX Absence records were generated using the background function with method environmental distance from sdm (), 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.

compare the results of each spatial resolution at each of the 3 stages (naive, expert, adaptive)

Environmental Niche Modelling Results

here put the number of plots which were occupied in the adaptive modelling effort.

Species Occupancy

Plant Density

3 | RESULTS

Comparision of Different Spatial Resolutions

Ground Verification

Species Occupancy

Plant Density

Southern rocky mountain herbaria portal.

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.