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 truthed 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 of

*Author for Correspondence: rbenkendorf@chicagobotanic.org

these models. Recent advances in remote-sensing technologies have allowed for well-designed ground truthing 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 truthed 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 tested on the ground, 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.

Application of high resolution (e.g. $< 1/3$ arc second, ~10m at 35*N) is highly desirable, but may incur enormous computational costs dependent on the domain of analysis.

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

< p3 > Improving Model Performance - Iterative adaptive-nice based sampling *ENM doesn't make the cut?* - move this into methods...

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 these models generally have high-performance metrics while tested on small subsets of hold-out data, they are unlikely to detect many new populations during ground truthing. 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 sampling design to collect multiple estimates

2 | METHODS