Select Seed Lots under Climate Change Scenarios

Reed Benkendorf, Rob Massatti & Daniel Shyrock

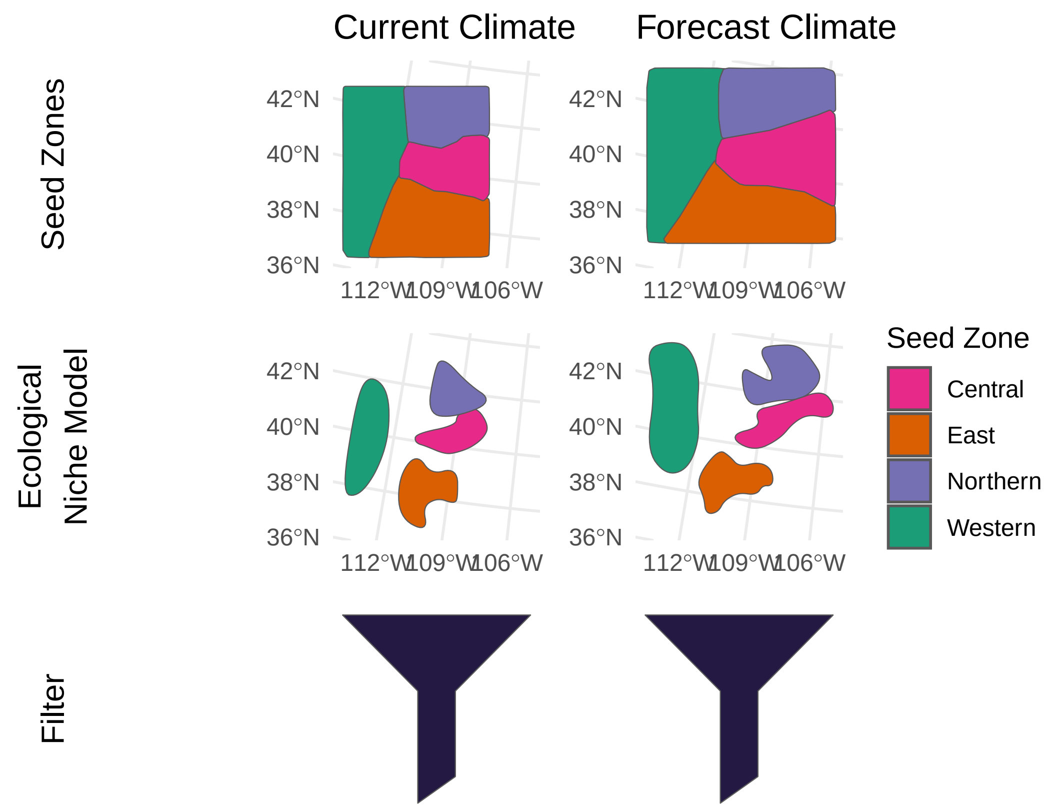

Figure 1: A general workflow for forecasting seed zones into future climate scenarios. The panels at left show the results for creating seed zones in space, clipping them to a species range, and then applying various filters to select priority areas for seed collection based on anticipated restoration needs. The process is repeated using climate forecasts in the panels at right. See `filter` figure to understand examples of data sets which may be used to filter data sets.

Seed Zones

When using a common garden type approach, and fitting fitting regression models to morphotypes from a CCA analysis avoid over fitting models to ensure the ability to generalize across both feature space. This will be particularly important for establishing seed zones under models of climate change, where we will be unable to update them in an iterative-adaptive framework. If using landscape genetic delineated seed transfer zones, consider fitting an ENM to populations found within each individual seed zone, and predicting them to future scenarios.

Ecological Niche Model

An array of statistical approaches and guidance have been developed which treat the (generally binary) occurrence of a species as a function of environmental variables to create ecological niche models (Elith & Leathwick 2010). An ENM can be fit to current species distribution records (from sources such as BISON, or GBIF), and predicted onto gridded surfaces under both current and future climate scenarios (see **Forecasting seed zones with climate models**), although these forecasts come with a range of assumptions (Dormann 2007, Araújo et al. 2005). Thresholding the probability surfaces allows for creating a binary mask which allows for restricting seed zones to areas which may support the species (Liu et al. 2005), and create more accurate estimates of realized seed zone sizes under current and future climate regimes. While this practice should be easily implementable for models under current conditions, which are often already limited to areas within the samples used for common garden (Johnson et al. 2017), although some modifications for clipping seed zones to forecast ranges, such as mechanistic based modelling, may need to be utilized (Urban et al. 2016).

Filter ('Sieve') Areas

Not all areas across the portions of seed zones may be used for seed collection, or restoration. For example, certain areas may have undergone transitions into an annual grassland states, prohibiting the possibilities for both seed development or effective restoration with our current skills. During the future time period plant materials are unlikely to be able to self-migrate to projected suitable habitat, hence these areas will only ever be sinks for developed plant materials, unless it becomes accepted practice to collect seed from seeded populations. In our simple conceptual figure, which can be considered as seeking to determine the seed needed per zone,

Environmental filters can be effectively set by masking raster pixels into 0/1 conditions, whereby 0 or masked cells are removed from downstream analysis, and 1 pixels allow the inclusion of the pixel in downstream analyses.

National Land Cover Dataset (NLCD) Details broad land cover classes, and can be used to remove areas which have been converted to annual grass. Updated annually in the conterminous US.

Protected Areas of the US Database (PADUS) Can be used to restrict ranges of analysis to public lands, by administrative agency (e.g. State, Federal, NGO) or agency (e.g. BLM, Forest Service) or combinations thereof. Updated every few years.

Historic Seeding Can be used to restrict seed collections from historically seeded areas. Updated sporadically.

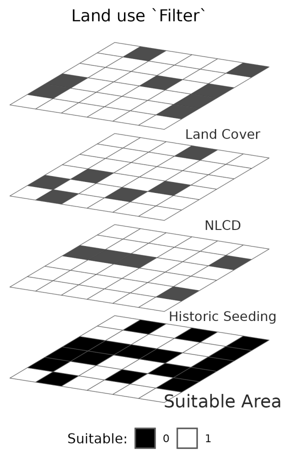

Figure 2: Conceptual figure indicating the use of filters to 'sieve' areas for priority germplasm development.

Forecasting seed zones with climate models

A variety of raster data sets exist, which interpolate current climate patterns, and forecast future climate conditions under various CO_2 relative concentration pathways (RCP) or shared socioeconomic pathways (SSPs) (see Noce et. al 2020). Three data sets, which may be of interest, use downscaling approaches to generate high resolution surfaces (WorldClim 2.0, CHELSA, and climateNA) and two of these data sets, ClimateNA (to ~30m resolution) and CHELSA (~ 1 km resolution), offer bioclim variables for a range of time periods for download (Brun et al. 2022, Fick & Hijman 2017, Mahony et al. 2022, Wang et al. 2016). Notably climateNA is the data product used by the Climate Smart & Seed Lot selection tools and models fit using these data sets should offer the maximum benefits for determining the most suitable origin locations for specific restoration projects (St. Clair et al. 2022).

Models of seed zones can be fit under both current conditions, and at the time points offered by the above models. For users unsure of which data products to use, i.e. which global climate model for a set of SSPs or RCPs, the same statistical model can be fit to a variety of forecasts and the mean (or in certain contexts the mode) and the predictions ensembled to create a single prediction, as well as estimates of uncertainty (see Dormann et al. 2018).

Preventing model extrapolation

Complications with interpreting ecological niche models (ENMs) in climate shifted space, where novel combinations of climate variables may generate feature spaces training data were not sampled from, have fostered approaches for detecting spatial extrapolation in geostatistical models developed from sparse data sets as is common in ecology (Elith et al. 2010, Owens et al. 2013). These methods seek to identify areas in feature space and map them to geographic space (pixels) on raster surfaces where the predictive ability of models are likely to deteriorate relative to the globally acquired performance metrics (e.g. RMSE, MAE, AUC-Pr), and under which the model should be used, at best, extra cautiously for decision making (Conn et. al 2015, de Bruin et al. 2022, Velazco et al. 2023, Wang et al. 2023). These approaches are now considered best practice for ENM modelling and the many alternatives should be investigated for incorporation to seed zone development modelling, although notably most of the recent developments pertain to machine learning approaches, and niche modelling applications.

Metrics for targeting native plant materials development

Under our approach numerous metrics could be used by native seed development personnel to prioritize projects. For example, the total area covered by a seed zone, statistical metrics of dispersion for resilience of seed zones (e.g. mean or median resilience values of each seed zone), the difference in size between current and forecast seed zones (e.g. zones may expand or contract under scenarios). Within zones, areas can be targeted for native seed collection and development which minimize the geographic distance between current and predict seed zones, or environmental similarity between areas currently within a seed zone and projected to be within a seed zone.

Example of metrics which can guide germplasm development

Areas in the left column have been `filtered`

All Areas	Filtered Areas
Area	
SZ t_0	SZ remnant t_0
SZ t_1	SZ remnant t_1
SZ pct. change t_0 -> t_1	SZ remnant pct. change t_0 -> t_1
Resilience	
mean SZ resilience t_0[1]	mean SZ resilience (- converted areas) t_0[2]
mean SZ resilience t_1	mean SZ resilience (- converted areas) t_1
mean SZ resilience t_0 -> t_1	mean SZ resilience (- converted areas) t_0 -> t_1

[1] While 'mean' is displayed in line, other summary statistics may be of interest.

[2] Burned area simulations can be used to estimate the extent of remnant lands in the future.

Figure 3: Metrics which can be used to relate the need for germplasm development under both current and forecast climate scenarios.