

Seed lot selection under climate change

Reed Benkendorf, Rob Massatti & Daniel Shyrock

Seed Zones

When using a common garden approach and fitting regression models to morphotypes derived from a canonical correspondence analysis, care should be taken to avoid overfitting. Overfitted models reduce the ability to generalize across environmental and geographic feature space (Wenger and Olden 2012; Yates et al. 2023). This consideration is especially important when developing seed zones under future climate scenarios, as opportunities to update these models in an iterative or adaptive framework may be limited (although using space-for-time substitution approaches may be possible, (Lovell et al. 2023)). If seed transfer zones are delineated using landscape genetic data, ecological niche models (ENMs) can be fit to populations within each zone and projected under future climate scenarios to guide seed sourcing and deployment.

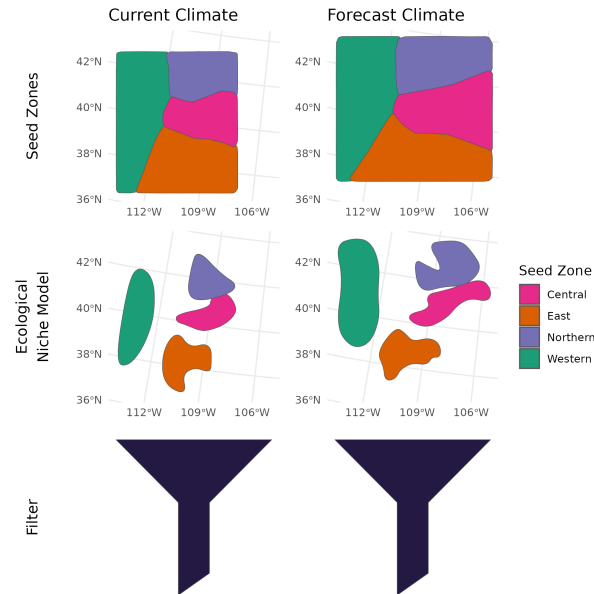


Figure 1: A general workflow for forecasting seed zones under future climate scenarios. The left panels illustrate the process of delineating seed zones, constraining them to a species' current range, and applying environmental filters to identify priority areas for seed collection based on anticipated restoration needs. The workflow is repeated using climate forecast data, as shown in the right panels. See the 'Filter' figure for examples of datasets commonly used to apply environmental filters.

Ecological Niche Model

Ecological niche modeling involves using species occurrence data and environmental predictors to statistically estimate a species' potential distribution (Elith & Leathwick, 2009). ENMs can be developed using occurrence data from sources such as BISON or GBIF and projected onto current or future climate surfaces (see *Forecasting Seed Zones with Climate Models*). However these model projections have multiple assumptions and uncertainties which should be understood before their usage (Dormann, 2007; Araújo et al. 2005).

By thresholding probability surfaces, binary masks can be created to limit seed zones to areas likely to support a species (Liu et al. 2005), resulting in estimates can maximize an objective for the extent of seed zones under both current and future conditions. While current-condition models are often naturally limited to sampled areas (e.g., common garden sites; Johnson et al., 2017), modifying future zone extents may require mechanistic modeling or additional post-processing (Urban et al., 2016).

Filter ('Sieve') Areas

Not all areas within a seed zone may be suitable for seed collection or restoration. For example, some landscapes may have transitioned to annual grassland states, limiting both seed viability and restoration potential. In many cases, projected suitable habitat may be inaccessible to natural seed dispersal (assisted climate migration), leaving these areas as long-term sinks unless seed collection from previously seeded populations becomes an accepted practice. However, environmental filters can be implemented by masking raster layers into binary surfaces (1 = included, 0 = excluded), effectively guiding the inclusion of pixels in downstream analyses.

National Land Cover Dataset (NLCD) Provides land cover classifications and can be used to exclude areas converted to annual grasslands. Updated annually in the conterminous U.S.

Protected Areas of the US Database (PADUS) Can restrict analyses to public lands by managing entity (e.g., State, Federal, NGO) or agency (e.g., BLM, USFS). Updated every few years.

Historic Seeding Can be used to avoid seed collection from historically seeded areas. These data are updated sporadically.

Forecasting seed zones with climate models

A variety of gridded climate datasets exist for both current conditions and future projections based on Representative Concentration Pathways (RCPs) or Shared Socioeconomic Pathways (SSPs) (Noce, Caporaso, and

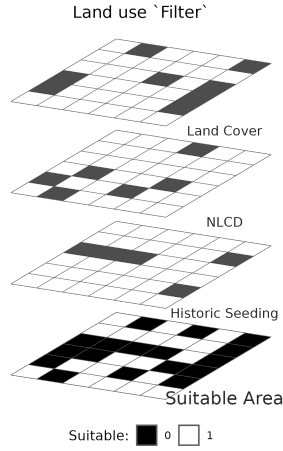


Figure 2: Conceptual illustration of environmental filtering used to identify priority areas for germplasm development. Filters are applied to exclude unsuitable areas based on ecological or land use criteria, effectively sieving the landscape to support targeted seed collection and restoration planning.

Santini 2020). High-resolution datasets such as WorldClim 2.0, CHELSA, and ClimateNA offer downscaled bioclimatic variables across historical and future time periods (Fick and Hijmans 2017; Brun et al. 2022; Mahony et al. 2022). Notably, ClimateNA (30 m resolution) is used in tools like the Climate-Smart and Seedlot Selection Tool, making it particularly relevant for site-specific restoration planning (St. Clair et al. 2022).

Seed zones can be modeled under both current and forecasted conditions. When modellers are uncertain about which climate projections to rely on, a single statistical model can be applied across multiple climate forecasts and the predictions ensembled to estimate both consensus predictions and uncertainty (Dormann et al. 2018).

Preventing model extrapolation

Projecting ecological niche models into future climate scenarios can lead to extrapolation into environmental conditions not represented in the training data. This can degrade model performance, especially in sparsely sampled feature spaces (Elith, Kearney, and Phillips 2010; Owens et al. (2013)]. Several methods, largely relying on cross validation, have been developed to detect spatial extrapolation and identify regions of low model reliability (Roberts et al. 2017). These tools map regions in geographic space where model performance may be unreliable relative to global performance metrics (e.g., RMSE, MAE, AUC-PR). In such areas, models should be interpreted cautiously (De Bruin et al. (2022); Wang, Khodadadzadeh, and Zurita-Milla (2023)). These extrapolation-detection tools are now considered best practice in ENM modeling, particularly for machine learning-based approaches, determining whether these approaches can be incorporated into

genecological studies is warranted.

Metrics for targeting native plant materials development

Several metrics can help prioritize native seed development projects, including:

- Total Area of Seed Zones: Useful for estimating potential seed demand.
- Resilience Metrics: Summary statistics (mean, median) for resilience indicators within each zone.
- Change in Zone Size: Degree of expansion or contraction between current and forecast zones.

What is important is that the extent of seed zones are considered in context, such as public land management, the expectations for natural re-vegetation of areas within a zone, and the change in area between current and climate change projections. Several participants in our workshop mentioned that it take many (3-10) to get a seed source reliably into production. Accordingly materials desired for 2040, should be field collected within the next decade.

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 $t_0 \rightarrow t_1$	SZ remnant pct. change $t_0 \rightarrow t_1$
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 $t_0 \rightarrow t_1$	mean SZ resilience (- converted areas) $t_0 \rightarrow t_1$

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: Key metrics for evaluating germplasm development needs under both current and forecast climate scenarios. These metrics support prioritization of seed collection and development efforts by quantifying spatial, ecological, and resilience-based factors within and across seed zones.

Works Cited

- Brun, Philipp, Niklaus E Zimmermann, Chantal Hari, Loïc Pellissier, and Dirk Nikolaus Karger. 2022. “Global Climate-Related Predictors at Kilometer Resolution for the Past and Future.” *Earth System Science Data* 14 (12): 5573–603.
- De Bruin, Sytze, Dick J Brus, Gerard BM Heuvelink, Tom van Ebbenhorst Tengbergen, and Alexandre MJ-C Wadoux. 2022. “Dealing with Clustered Samples for Assessing Map Accuracy by Cross-Validation.” *Ecological Informatics* 69: 101665.
- Dormann, Carsten F, Justin M Calabrese, Gurutzeta Guillera-Arroita, Eleni Matechou, Volker Bahn, Kamil Bartoń, Colin M Beale, et al. 2018. “Model Averaging in Ecology: A Review of Bayesian, Information-Theoretic, and Tactical Approaches for Predictive Inference.” *Ecological Monographs* 88 (4): 485–504.
- Elith, Jane, Michael Kearney, and Steven Phillips. 2010. “The Art of Modelling Range-Shifting Species.” *Methods in Ecology and Evolution* 1 (4): 330–42.
- Fick, Stephen E, and Robert J Hijmans. 2017. “WorldClim 2: New 1-Km Spatial Resolution Climate Surfaces for Global Land Areas.” *International Journal of Climatology* 37 (12): 4302–15.
- Lovell, Rebecca SL, Sinead Collins, Simon H Martin, Alex L Pigot, and Albert B Phillimore. 2023. “Space-for-Time Substitutions in Climate Change Ecology and Evolution.” *Biological Reviews* 98 (6): 2243–70.
- Mahony, Colin R, Tongli Wang, Andreas Hamann, and Alex J Cannon. 2022. “A Global Climate Model Ensemble for Downscaled Monthly Climate Normals over North America.” *International Journal of Climatology* 42 (11): 5871–91.
- Noce, Sergio, Luca Caporaso, and Monia Santini. 2020. “A New Global Dataset of Bioclimatic Indicators.” *Scientific Data* 7 (1): 398.
- Owens, Hannah L, Lindsay P Campbell, L Lynnette Dornak, Erin E Saupe, Narayani Barve, Jorge Soberón, Kate Ingenloff, et al. 2013. “Constraints on Interpretation of Ecological Niche Models by Limited Environmental Ranges on Calibration Areas.” *Ecological Modelling* 263: 10–18.
- Roberts, David R, Volker Bahn, Simone Ciuti, Mark S Boyce, Jane Elith, Gurutzeta Guillera-Arroita, Severin Hauenstein, et al. 2017. “Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure.” *Ecography* 40 (8): 913–29.
- St. Clair, John Bradley, Bryce A Richardson, Nikolas Stevenson-Molnar, Glenn T Howe, Andrew D Bower, Vicky J Erickson, Brendan Ward, Dominique Bachelet, Francis F Kilkenny, and Tongli Wang. 2022. “Seedlot Selection Tool and Climate-Smart Restoration Tool: Web-Based Tools for Sourcing Seed Adapted to Future Climates.” *Ecosphere* 13 (5): e4089.
- Wang, Yanwen, Mahdi Khodadadzadeh, and Raúl Zurita-Milla. 2023. “Spatial+: A New Cross-Validation

- Method to Evaluate Geospatial Machine Learning Models.” *International Journal of Applied Earth Observation and Geoinformation* 121: 103364.
- Wenger, Seth J, and Julian D Olden. 2012. “Assessing Transferability of Ecological Models: An Underappreciated Aspect of Statistical Validation.” *Methods in Ecology and Evolution* 3 (2): 260–67.
- Yates, Luke A, Zach Aandahl, Shane A Richards, and Barry W Brook. 2023. “Cross Validation for Model Selection: A Review with Examples from Ecology.” *Ecological Monographs* 93 (1): e1557.