Empirical seed transfer zones require conventions for data sharing to increase their utility

Brianna Wieferich, Reed Benkendorf
E. Woodworth (wu-naut) Brianna.Wieferich@usda.gov
chicago botanic garden

BACKGROUND To increase the availability of locally adapted germplasm, empirical Seed Transfer Zones (eSTZs) are being used to guide restoration and agricultural admixture decisions. eSTZs are based on common garden, genetic, or climate similarity data, and require considerable effort to develop. Despite the rigor with which these studies are conducted, inconsistencies exist in data products derived from them which can hinder their utility and adoption. Here we showcase the inconsistencies in eSTZ data products, offer suggestions for standardization, and report on an R package to implement these suggestions.

Objectives

Develop:
1) file naming conventions
2) field naming standards in vector data
3) cartographic standards
4) directory structure conventions
5) Implement suggestions in an R package 'eSTZwritR'

Methods Reviewed all eSTZs on the Western Wildland Environmental Threat Assessment Center (WWETAC) website as of May 1, 2024.

Each data product's: file name structure, field naming conventions, and directory structure, were analyzed. All scoring was done by hand, and all analyses were carried out in R 4.2.1.

Results

Shapefile Properties

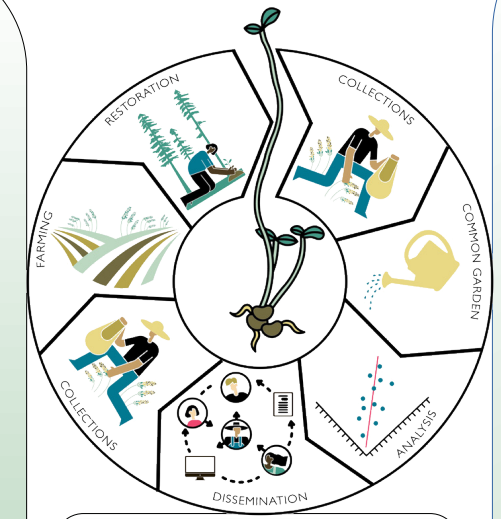
- Different polygon types used for treating zones
- Field denoting seed zone was generally lacking

File Names (shapefile)

- Consensus to use USDA NRCS-Plants code
- Many files did not mention the data (seed zones)
- Majority of files lack a specified geographic extent

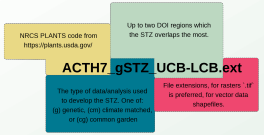
Map Components

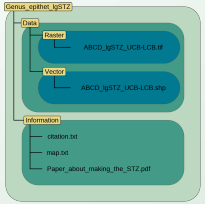
- Several essential cartographic elements (title, statement on data sources, legend for the seed zones) were missing from many products.



Suggestions

File naming convention





Example field names in a shapefile

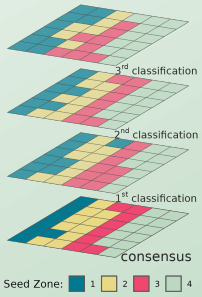
ID	SeedZone	SZName	AreaAcres	BIO1_R	BIO2_mean
1	1	Salt Desert	12340	20.2	5.1
2	2	Desert Scrub	14230	19.1	7.1
3	3	Pinyon-Juniper/Oak Brush	30142	15.1	10.1
4	4	Montane	9872	12.3	12.3

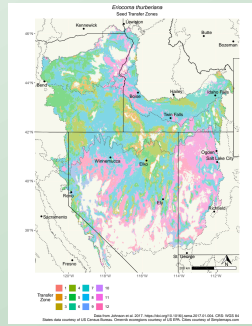
The first four (blue) fields should be in every file. More fields are optional.

- Use UpperCamelCase for field names whenever possible.
- The four field names above should be in every file.
- Do you need to include extra variables? If so place them "after" these four fields.
- If using Bioclim variables, use their official bioclim designations - e.g. BIO8.
- If using a statistic use: min, max, mean, R, SD; prefixed with "_" e.g. "BIO8_R" .
- Have SeedZone numbers in ascending order from the hottest and driest areas.
- Descriptive names (SZName) e.g. "SW montane" are useful additions to the Seed Zones.
- "ID" reflects each polygon geometry - if using multipolygons the ID will be "SeedZone" .

These suggestions can be implemented using an R package, 'eSTZwritR'. Or by consulting it's webpage. We look forward to collaboratively developing standards in the future!

Three model predictions from a classifier and a consensus layer





Conclusions Myriad discrepancies exist in the way that eSTZs are being distributed. Here we present standards for the scientists developing eSTZs to use in order to standardize the data products they develop. We provide an R package to implement our suggestions.