

Goal

This vignette introduces the CERIS R package developed by Li et al. 2021 to model, explain, and predict trait dynamics from a multi-environment trial (MET). The join regression analysis (JRA) is one of the popular frameworks for MET analysis. The key limitation of JRA is that it is based on environmental mean, the average performance of the whole population at each environment. CERIS, Critical Environmental Regressor through Informed Search, identifies an environmental index to replace the environmental mean.

Outline

Here an environment is defined by the combination of location and year. When multiple planting dates are involved, *i.e.*, the same population is planted in the same location at different time points in the same year, each combination of location, year, and planting date is counted as an environment.

Three components required for CERIS are:

1. The performance records for a population with n individuals from m environments
2. The candidate pool for environmental index
3. Genotype files (optional, only if performance predicting of new genotypes is conducted.)

CERIS will use the first requirement to calculate the environmental mean from each environment. The environmental mean vector will have m numbers.

The candidate pool is built based on the combinations of growing periods and environmental parameters. A growing period is defined as x to y days-after-planting (DAP). Because of the varied planting dates among environments, the same growing period can vary among environments. For example, the same 1 to 10 DAP may correspond to May 1st to May 10th in one environment, but May 16th to May 25th in another environment. Environmental parameters can be common individual environmental factors such as daily maximum temperature, daily minimum temperature, day length, precipitations, and others that can be directly measured from any environment. Environmental parameters can also be calculated based on two or more environmental factors. Here we provide some environmental parameters derived from temperature and day length.

$GDD = \frac{1}{2}(T_{max} + T_{min}) - T_{base}$
$PTT = GDD \times DL$
$PTR = GDD \div DL$
$PTS = (T_{max}^2 - T_{min}^2) \times DL^2$

DL is the day length (day time in hours), which is calculated based on the daylength function in the *geosphere* package based on latitude coordinate. T_{max} is the maximum temperature, T_{min} is the minimum temperature, while T_{base} is adjusted based on species. A typical daily temperature source is NOAA GHCN (<https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/global-historical-climatology-network-ghcn>). Alternatively, you can build the environmental index pool from your own in-field weather stations.

Three essential input files are required for CERIS. 1). Trait records from the MET; 2). Meta information for the MET; 3). The environmental index pool file. The corresponding files in the

examples are: Trait_records.txt, Env_meta_table.txt, and 7Envs_envParas_DAP122. Follow the format to compile your own input files.

The CERIS R package has two scripts, CERIS_Workshop.r for the main function and Sub_functions_Workshop.r for subfunctions. After your own input files are formatted and stored in the corresponding directory, a few places you need to modify in CERIS.r based on your own settings are highlighted.

The directory structure.

```
##### Top_dir
##### |----0Maize (282DP as an example)
##### | |--8Envs_envParas_DAP80.txt
##### | |--Env_meta_table.txt
##### | |--Trait_records.txt
##### | |--Genotype.txt
##### |----CERIS.r
##### |----Sub_functions.r
```

In line 22, modify the Tbase based on your own species.

In line 25, modify the top_dir based on your own setting.

In line 34, modify the trait you want to analyze.

Run the CERIS.r script to line 77. Open the output files to determine the best environmental index, which should have the strongest correlation with environmental mean. Then modify the values from line 95 to 97 based on the searching results. Run the rest of the script.

After running all the lines. A new folder under 0Maize will be created for each trait. The most important output files are DTA_8Envs_PTTPTR_0LOO_cor.txt and MaxR_DTA_8Envs_0LOO.png. DTA_8Envs_PTTPTR_0LOO_cor.txt stores all the correlations between environmental mean and environmental indices. Each row is for one growing period. The fifth to ninth columns are the original correlations and corresponding to the top row, while the tenth to fourth columns are the negative values and corresponding to the bottom row in MaxR_DTA_8Envs_0LOO.png. In this DTA example, it is PTR from x to y DAP having the strongest but negative correlation.