Optimal Flowering Time Scripts & Data Organization

Last Updated: 4/28/2025

**Files that need to be downloaded separately/are not derived from the analysis are in bold.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **File Name** | **Original Name** | **Description** | **Inputs** | **Outputs** |
| 00\_analysis\_functions.R | 00\_analysis\_functions.R | Functions that are used in 01 and 02 to extract peaks from gaussian curves, subset parts of data, etc. |  |  |
| 01\_R2\_analyses.R | R2\_analyses.R | Function to fit Gaussian and linear models for all trials, calculate pseudo-R2 and AIC for each model. Meta-regressions on R2. | * 00\_analysis\_functions.R * **sunflower\_data\_simple\_v1.csv** | * R2\_AIC\_data.csv * R2\_indata.csv (for counties) * R2\_indata\_exact.csv (for exact location)   Meta-analysis (county):   * R2\_sample\_size.png * year\_r2\_county\_unadj.png * state\_r2\_county\_unadj.png |
| 02\_gaus\_opt\_flow\_time.R | 04\_gaus\_opt\_flow\_time.R | Functions for fitting Gaussian models to all locations, make diagnostic plots of all site years, and write out files with optimal flowering time, and original data.  Also runs analysis on absolute performance of trials.  Must run first half of 02\_R2\_analysis.R prior to running the functions in this script | * 00\_analysis\_functions.R * **sunflower\_data\_simple\_v1.csv** * R2\_AIC\_data.csv | For exact locations:   * gauss\_allexact\_mods.RData * optimum\_flowering\_time.csv * exact\_loc\_lookup.csv   For counties:   * gauss\_county\_mods.RData * optimum\_flowering\_time\_county.csv * county\_lookup.csv   Figures:   * linear\_gauss\_curves\_all.pdf * linear\_gauss\_curves\_county.pdf * single\_site\_yr\_examples.png * yield\_optimum\_relationship.png |
| 03\_clim\_yearly\_stats.R | 06\_clim\_analyses.R | Creates a yearly summary of climate parameters for each location from the Daymet data from 1980-2023.  Makes plots of correlations of the climate variables. PCAs of climate variables – need to run functions to make all climate variables for these to work in current file version. | * **daymet\_timeseries\_cleaned.csv** * **sunflower\_data\_simple\_v1.csv** * R2\_indata.csv | * yearly\_climate\_data.csv |
| 04\_clim\_analyses\_optimum.R | 06a\_clim\_analyses\_optimum.R | Does Bayesian analysis of which climate variables relate to optimal flowering time. Can work on both county and exact locations. Runs models for Day of Year, Days past Planting, with GDD70 and GDD fixed period. Does future projections, makes predictor-residual plots for correlated variables. | * optimum\_flowering\_time\_county.csv * [optimum\_flowering\_time.csv – if run for exact locations] * yearly\_climate\_data.csv | Files:   * optim\_clim\_sc\_county.csv * optim\_clim\_county.csv * [optim\_clim\_sc.csv * optim\_clim.csv – for exact loc]   Figures:   * dpp\_distribution.png * DOY\_opt\_clim.png * DOY\_future.png * fixed\_DOY\_opt\_clim.png * DPP\_opt\_clim.png * DPP\_future.png * fixed\_DPP\_opt\_clim.png * ALL\_resids.png   Results tables:   * table\_dpp.png * table\_all.png   Can also run this file for exact location, and the figures and tables will be written out in a different directory. All results tables will be combined into a single table when run by exact loc. |
| 05\_proportion\_optimal.R | 07\_proportion\_optimal.R | Multinomial regressions on whether a trial was ‘suitable’, ‘early’, ‘late’, or ‘unclear’. There are multinomial regressions for year, location (county or exact), and for climate. Figures are made for only the climate variables that come out as associated with the categorization. | * optim\_clim\_county.csv | * year\_county\_mult.png * climate\_mult\_full.png   Can also run for exact location and figures will be saved in a different directory. |
| 06\_variation\_DPP.R | variation\_DPP.R | Compares flowering timing of commercial and USDA/GRIN accessions in a histogram. | * **grin\_dayflow.csv** * **sunflower\_data\_simple\_v1.csv** | * variation\_DPP.png |
| 07\_ugly\_map.R | ugly\_map.R | Draws a map with all data points for each trial location. Colors points by climate and sizes points by how many trials are in each location. Also compares the subset of sites used in exact location and county location groupings. | * R2\_indata.csv * R2\_indata\_exact.csv * clim\_yearly\_out.csv | * ugly\_map.png |
| 08\_summary\_stats.R | summary\_stats.R | Calculates sample sizes of sites (counties), years, hybrids, etc. used in the analyses. The R2 data is the biggest dataset used, so these stats are calculated for that subset. A smaller subset of points was included for climate analyses, which is also calculated. | * R2\_indata.csv * optim\_clim\_sc\_county.csv |  |
| 09\_climate\_trends.R | climate\_trends.R | Makes a figure with the climate trends over time for each county. Also makes a map that visualizes the average climate for each location (not saved, and not updated to county grouping). | * yearly\_climate\_data.csv * optimum\_flowering\_time\_county.csv | * climate\_trends.png |

**List of figures and tables in manuscript**

Table 1. county/table\_dpp.png

Figure 1. county/ugly\_map.png

Figure 2. county/state\_r2\_county\_unadj.png

Figure 3. county/climate\_mult\_full.png

Figure 4. county/DPP\_opt\_clim.png

Figure 5. county/DPP\_future.png

Figure 6. variation\_DPP.png

**Supplement**

Table S1. county/table\_all.png

Table S2.  exact\_location/table\_all.png

Figure S1. climate\_trends.png

Figure S2. county/R2\_sample\_size.png

Figure S3. county/single\_site\_yr\_examples.png

Figure S4. county/year\_r2\_county\_unadj.png

Figure S5. county/year\_county\_mult.png

Figure S6. county/yield\_optimum\_relationship.png

Figure S7. county/dpp\_distribution.png

Figure S8. county/ALL\_resids.png

Figure S9. county/DOY\_opt\_clim.png

Figure S10.  exact location/DPP\_opt\_clim.png

Figure S11. county/fixed\_DPP\_opt\_clim.png

Figure S12. county/fixed\_DOY\_opt\_clim.png

Figure S13. county/DOY\_future.png

Figure S14. nass\_planting\_dates.png