ToC & Notes

Devin Gamble

12/23/2020

Contents

Project Overview

To understand how climate drives variation in reproductive phenology, we are utilizing over 1,700 herbarium records of the widely distributed annual wildflower Nemophila menziesii. By modelling specimen DOY in response to climate and collection location, we can describe climate-driven changes in phenology across space and time. Additionally, including a metric (Phenological Index – PI) to account for the phenophase of digitized specimens will allow for models with greater predictive power and the capacity to investigate shifts in specific phenophases. Finally, understanding variation in phenological sensitivity to climate will aid us in predicting how future climates will impact Nemophila menziesii as well as similar species' ecology, reproduction, and persistence.

Data and organization:

Through separate rounds of cleaning and duplicate removal, georeferencing, climate data extraction, and scoring, there are several different csv files with *Nemophila* specimen data present in this project repository. They are listed below for clarity, with details on what they contain and how recently they were made.

Specimen Records: These csv files can be found in the **data_cleaning** folder (within additional folders). They include

CCH1 Data:

- `Nemo_CCH1_02182020.csv` - CCH1 data set standardized by Ivana Gomez
- `nemo_cch1_cleaned.csv` - cleaned (by DG) CCH1 data set
- georeferenced_cch1_splits (folder) - cleaned cch1 data split by herbarium and georeferenced; combined later with CCH2 data

CCH2 Data:

- `occurrences.csv` - the original specimen data downloaded (CCH2 12-19-19 download)
- `nemo_cch2_Dec2019.csv` - cleaned up from the occurrences data
- `nemo_cch2_cleaned_V2.csv` - further cleaning, removing duplicates. This was then split up into subsets by herbarium for georeferencing & scoring.
- splits_for_PI (folder) has most recent cleaned CCH2 splits ('_pi.csv' files are more up to date than '_p.csv' files! Dupes & odd things removed, data re-split by CCH2 herbaria *after* running `combine_cch1_cch2_V2.R`)

Combined CCH1 & CCH2: – See `combine_cch1_cch2_V2.R` for more details.

- `cch1&2_combined.csv` - first combination, NOT cleaned
- `nemo_all_1.csv` - first *updated* combination of CCH1 and CCH2 records
- **Most up-to-date**: `nemo_all_2.csv` - second combination of CCH1 and CCH2 records, added elevation (m) from 'elevatr' package. Used to make latest PI scoring splits.

Specimen Images (CCH2) 'cch2_images' folder contains all images for CCH2 specimens at the time of initial downlaoding. Following the cleaning of CCH2 specimens, images were split up to match the splits made for PI scoring. The folder 'cch2_images_n' contains these folders and the images that were scored.

Climate Data: – **climateNA** folder
– See `get_cna_all.Rmd` for more details. Uses 'nemo_all_2.csv' as input to create ClimateNA compatible csv 'all_CNA_s.csv'.

- First-attempt CCH1 and 2 climate data can be found in their respective folders.
- Direct ClimateNA downloads are in the "cna_rawfiles" folder within the "all_CNA" folder.
- These were cleaned and combined to get `norms_raw_51_and_81.csv`, `YoC_Y_and_L1.csv`, `100y_avg.csv`

- Anomalies calculated in 'get_cna_anoms.Rmd', creates `climNA_anoms.csv`

- Most up-to-date: `climNA_nemo_all.csv` - combined all previous climate files, saved in "all_cna" folder

All-in-one, most up-to-date csv file: `nemo_full_1901_2019.csv`

- Combines specimen data from `nemo_all_2.csv` (with error distance) and climate data from `climNA_nemo_all.csv` (with PPT vars log-transformed), filtering for years only from 1901-2019. Use `nemo_full_1861_2019.csv` for full temporal breadth.

Project Components

The following sections describe the different objectives, analyses, and their pertinent Rmd files.

Data Cleaning

Folder: **data_cleaning**

This folder contains several scripts designed to clean and standardize the raw data downloaded from CCH1 and CCH2, as well as scripts to split up data for georeferencing and PI scoring. It also contains Rmds (`image_dl.Rmd` and `image_dl2.Rmd`) to extract and download images from the CCH2 herbarium records. On a (my) local machine, this folder contains all of these downloaded images. Since these files are not central to the actual analysis of data, they are not described further here.

Climate Data

Folder: **climateNA**

These folders contain markdowns, csvs, and raw data used to organize the variables extracted from ClimateNA. "all_cna" is the main folder. The cch1 and cch2-specific folders were early attempts at climate data extraction.

Markdown Files:

- `get_cna_all.Rmd`
  - Text and code for the extraction of date from ClimateNA. Includes instructions on obtaining YoC, Normal, and time-series data.
- `get_cna_anoms.Rmd`
  - similar to the above file, this describes the procedure for obtaining climate anomalies.

Contact DG for a combined Rmd that covers the extraction and calculation of all these climate variables.

Data Exploration

Folder: **analyses**

Markdown Files:

- `EDA1.Rmd`
  - Explores non-climatic trends in the combined herbarium data. Has a few simple visualizations and raster code for a rough distribution map.
- `EDA2.Rmd`
  - More comprehensive exploratory analysis, including: Adding collector information back into combined dataset, creating error-distance subsets, examining variable correlations, creating basic pheno-climatic models, examining temporal change in climate variables, interaction plots, a couple 'nice' figures.
  - This file is long and not very well-organized. It mainly serves as a sketch pad for analyses that were (or would be) refined later.

  - First writing of the complete `nemo_full_1901_2019.csv` and `nemo_full_1861_2019.csv`, after log-transforming PPT variables!

- `EDA_climate.Rmd`
  - Simple plots and distributions of different climate variables. Raster maps.

Phenological Senseitivity

Folder: **analyses**

In addition to some analyses on phenological sensitivity in the EDA markdowns, these files contain more polished and complete analyses on pheno-climatic models.

Markdown Files:

- `sensitivity_1`
  - Analysis of pheno-climatic models
  - Large chunk of code copied over from Climate_1.Rmd (since removed)

Error Distance

Folder: **analyses/error_distance**

This folder contains Rmds and data for analyzing the effect of coordinate uncertainty on estimates of phenological sensitivity. The folder 'new_coords' contains csvs of simulated climate data for high-confidence records.

- `bias_error.Rmd`
  - Exmaines the effect of error distance on pheno-climatic models by (1) comparing models run with different subsets of data and (2) by including error_distance as a covariate. Considers 100Y and anomalous conditions. Also some exploration of other trends (e.g., georeferenced specimens, collector bias).
- `bias_error2.Rmd`
  - Bootstrap resampling analyses to emphasize confidence in the previous analyses (split by data subsets). Compare bootstrapped estimates among 100Y and anomalous climate conditions
- `bias_error3.Rmd`
  - Simulated Data for new lat/long + elevation, dispersed by one of five distances from original coordinates. Format new climate data into combined csv files.
- `bias_error3b.Rmd`
  - Regression analysis & comparison of simulated coordinate data – continuation of the previous file.

Other files

These are project files not mentioned in the above components:

- `AC_BO_phenology.Rmd`
  - Calculates and shows the meand DOY and cliamtic conditoins at AC and BO populations.