Code from: Expanded potential growing region and yield increase for Agave americana with future climate

Predicting Agave Potential Yield

The GitHub repository (cct-datascience/agave-prediction) contains the code and documentation for our paper "Growing Region and Yield Increase for Agave americana with Future Climate" by Davis et al 2021

Data generated by code in this repository is archived in the University of Arizona Research Data Repository doi:10.25422/azu.data.16828279.

> Davis, Sarah C., John T. Abatzoglou, and David S. LeBauer. 2021. "Expanded Potential Growing Region and Yield Increase for Agave americana with Future Climate" Agronomy 11, no. 11: 2109. https://doi.org/10.3390/agronomy11112109

GitHub Repository Contents

Analysis Code

The **analysis/** directory contains code and data used in this paper. The intent is to provide a record of the steps - including data sources and the model structure. All of the code used in the analyses is provided, although it hasn't been run end-to-end as is so can not be considered truly reproducible. This is because many steps are slow, the WPDA data changed since it was originally downloaded, and both netCDF and Panoply can be finicky.

Analysis code was run in this order:

```
01-download_climate.R
01.1-combine_masks.R
02-append_terraclimate.sh
cd data/derived_data/
03-compute_biomass.sh
04-update_nc_metadata.sh
05-analysis.R
```

- `01-download_climate.R` downloads the TERRACLIMATE data to the `raw_data` directory.
- `01.1-combine_masks.R` combines urban and protected area polygons and generates polygons stored in /data/derived_data/not_suitable.gpkg
- `0.2-append_terraclimate.sh` combines TERRACLIMATE data into a single file for each scenario in the derived_data folder. `derived_data` named `TerraClimate4C.nc` and `TerraClimate19812010.nc`
- `03-compute.biomass.sh` implements the model under different climate scenarios with rainfed, irrigation, and rock mulch scenarios (described below)
- `04-update_nc_metadata.sh` corrects metadata in output netcdf files
- `05-analysis.R` applies the urban and protected area masks, generates masked versions of the maps as netCDF files, and calculates the potential area and total biomass for each scenario.
- `06-figures.Rmd` generates the tables using Panoply and also provides alternatives to generating maps in R, but these were not used.

Data Sources

Analyses start with the following datasets in the `analysis/data/raw_data/` folder, and generate outputs in `analysis/data/derived_data`. Raw data are not archived, as they can be accessed from the provider as described below.

Climate data used as model inputs

TERRACLIMATE data from Abatzoglou et al (2018) were ownloaded from: http://www.climatologylab.org/terraclimate.html

TERRACLIMATE is on a 1km grid for 1980-2010, +2C and +4C warming climates.

We accessed monthly mean values of: - precipitation (ppt) - minimum temperature (tmin) - solar radiation (srad)

In the files: - `TerraClimate19812010_ppt.nc` - `TerraClimate19812010_tmin.nc` - `TerraClimate2C_tmin.nc` - `TerraClimate4C_tmin.nc` - `TerraClimate2C_ppt.nc` - `TerraClimate4C_ppt.nc` - `TerraClimate19812010_srad.nc`

> Abatzoglou, J.T., S.Z. Dobrowski, S.A. Parks, K.C. Hegewisch, 2018, Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015, Scientific Data

Absolute Annual Minimum Temperature

The calculation of absolute annual minimum temperature is described in the section Calculating Absolute Minimum Annual Temperature. The variable `absmin` is found in the files `absmin19802010.nc` and `absmin4C.nc`.

Inputs to the calculation of absolute annual minimum temperature include tmin from TERRACLIMATE as well as `cadj` and `adjustC`. The variable `cadj` represents the perturbation from true coldest month to climatological coldest month. The variable `adjustC` represents the difference in climatological coldest night of the year and daily average lowest temperature of a month.

The variables `cadj`, `adjustC` and `absmin` were created by Dr. Abatzoglou for this analysis. Since they are the only inputs to the analysis described below that are not otherwise available, the original inputs are archived in the file `absmin_adjustments.zip`. The same data layers are also provided alongside `absmin` in the files `absmin19802010.nc` and `absmin4C.nc`.

Urban and protected areas

The following data was used to exclude present day 1) Urban Areas (Kelso and Patterson 2010; South 2017) and 2) Protected Areas (UNEP-WCMC and IUCN, 2021) from regional predictions of suitable land area and potential biomass:

- **Natural Earth**
    - Downloaded with the `rnaturalearth` R package. Andy South (2017). rnaturalearth: World Map Data from Natural Earth. R package version 0.1.0. https://CRAN.R-project.org/package=rnaturalearth
    - Kelso, Nathaniel Vaughn, and Tom Patterson. "Introducing natural earth data-naturalearthdata.com." Geographia Technica 5.82-89 (2010): 25.
- **Protected Areas** UNEP-WCMC and IUCN (2021), Protected Planet: The World Database on Protected Areas (WDPA) and World Database on Other Effective Area-based Conservation Measures (WD-OECM) [Online], April 2021, Cambridge, UK: UNEP-WCMC and IUCN. Available at: www.protectedplanet.net.

The file `WDPA_WDOECM_Apr2021.zip` that contains shape files was downloaded, unzipped into `raw_data/` folder, and converted into a geopackage using the script `01.1-combine_masks.R`. For convenience the aggregated, OGC-compliant geopackage format of this database is archived in the file `wdpa.gpkg`.

This dataset is updated regularly, and can be downloaded from https://www.protectedplanet.net.

Software Used

NCO: NetCDF Operators (Zender 2014) Zender, C. S. (2014), netCDF Operator (NCO) User Guide, Version 4.4.3, http://nco.sf.net/nco.pdf.

R R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

fasterize Noam Ross (2020). fasterize: Fast Polygon to Raster Conversion. R package version 1.0.3. https://CRAN.R-project.org/package=fasterize

rnaturalearth Andy South (2017). rnaturalearth: World Map Data from Natural Earth. R package version 0.1.0. https://CRAN.R-project.org/package=rnaturalearth

sf Pebesma, E., 2018. Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal 10 (1), 439-446, https://doi.org/10.32614/RJ-2018-009

exactextractr Daniel Baston (2021). exactextractr: Fast Extraction from Raster Datasets using Polygons. R package version 0.6.0. https://CRAN.R-project.org/package=exactextractr

Panoply PanoplyCL v 4.12.8 b0071 downloaded from https://www.giss.nasa.gov/tools/panoply/download/ can be used to recreate figures, see `.pcl` files in figures directory

Analysis

Raw Data (Input)

Files in the `/analysis/data/raw_data` are described in the Data Sources section below. These have not been archived.

- TERRACLIMATE files must be downloaded to this folder using the `01-download_climate.R` script
- WPDA Shapefiles can be downloaded from www.protectedplanet.net

Derived Data (Output)

The file `allbiomass.nc` provides predicted annual biomass increments for *Agave americana* under all climates and scenarios. It contains the model predictions of annual biomass production with units of Mg/ha/y for six scenarios on the same latitude and longitude dimensions as the TERRACLIMATE dataset.

variable	water management	climate
`biomass`	rainfed	historical (1981-2010)
`biomass_irrig`	irrigated	historical (1981-2010)
`biomass_rockmulch`	rainfed with rock mulch	historical (1981-2010)
`biomass4C`	rainfed	4C warming
`biomass4C_irrig`	irrigated	4C warming
`biomass4C_rockmulch`	rainfed with rock mulch	4C warming

Other intermediate files of interest:

- `absmin19812010.nc` and `absmin4C.nc`contain the variable absmin, the absolute minimum temperature. It also contains the variables `tmin`, `adjustC`, and `cadj` used to compute `absmin` as described in Calculating Absolute Minimum Annual Temperature.
- `coefs19812010.nc` and `coefs4C.nc` provide calculated values of α, β, and Γ for each grid cell under 1981-2010 and +4C climate normals for three water availability scenarios.
- `not_suitable.gpkg` polygons combining land classified as unsuitable for growing that is either protected (UNEP-WCMC and IUCN 2021) or urbanized (Kelso and Patterson 2010).

Analysis Steps

1. Download **TERRACLIMATE**

```
# ~ 1h @10Mbps
cd analysis
Rscript --vanilla ./01-download_climate.R
```

2. Download WPDA Protected Areas

Per the WPDA Data license this dataset can not be downloaded.

1. Go to https://www.protectedplanet.net/en/thematic-areas/wdpa
2. Download `WPDA_WDOECM_Apr2021_Public_all_shp.zip`
3. Unzip to `raw_data/WDPA_WDOECM_Apr2021/`

3. Generate mask of protected and urban areas

Combine WPDA with rnaturalearth Urban areas, generate polygons as 'masks.gpkg'

```
Rscript --vanilla ./01.1-combine_masks.R
```

NB this step sometimes does not work due to an error generated by the rnaturalearth API, described in https://github.com/ropensci/rnaturalearth/issues/29.

4. Append **TERRACLIMATE** and compute derived variables

```
./02-append_terraclimate.sh
```

5. Calculate PAR

$$\text{PAR} = 2 \times \text{SRAD}\frac{W}{m^{-2}} \times 110\frac{\text{lux}}{W} \times 10^{-3}\frac{\text{klux}}{\text{lux}} \times 18\frac{\mu\text{mol/m}^2\text{s}^2}{\text{klux}}$$

- SRAD: Mean daily shortwave radiation (W/m2) from TERRACLIMATE
- PAR: Photosynthetically active radiation in units of (umol photons / m2 / s)
- 110 is the average global annual luminous efficacy value of 110 lumens per W (Littlefair, 1985) in order to convert to lux (lumens per square meter)
- lux converted into the average daily PAR using the relationship 1 klux = 18 µmol photons m-2 s-1 of daylight.
- Multiply by 10-3 to convert from lux to klux.
- Multiply x2 to convert from mean PAR integrated over 24 hours to the flux over a 12 hours of daylight on average

6. Calculate Absolute Minimum Annual Temperature

$$\text{absmin} = \text{tmin} - \text{cadj} - \text{adjustC}$$

- `absmin`: absolute minimum annual temperature, the daily coldest minimum temperature
- `tmin`: average minimum temperature of the coldest month, climatological
- `adjustC`: is from ERA5 and represents the difference in climatological coldest night of the year and daily average lowest temperature of a month.
- `cadj`: perturbation from true coldest month to climatological coldest month

`cadj` is a minor adjustment to reflect the fact that ERA5 calculations did these calculations dynamically (coldest Tmin in a year minus Tmin of the coldest month) as opposed to using coldest Tmin of climatology. The former allows for the fact that in some years December might be the coldest, in others January. The climatology will tend to have a higher temperature of the coldest month than if these are calculated independently.

7. Run model

See detailed description in the next section.

Model Implementation

The EPI model is implemented using NetCDF Operators in the file analysis/03-compute.biomass.sh. The key equations of the model are provided below for reference; a full description is provided in Niechayev et al (2019) and the manuscript associated with this archive Davis et al (2021).

> Niechayev, N., Jones, A., Rosenthal, D., & Davis, S. (2019). A model of environmental limitations on production of Agave americana L. grown as a biofuel crop in semi-arid regions. Journal of Experimental Botany, 70, 6549-6559.

light index α:

$$\alpha = \begin{cases} 0, & \text{for } I < 0.6252 \\ -7 \times 10^{-7} I^2 + 0.0016\ I - 0.0001, & \text{for } I > 0.6252 \end{cases}$$

Where I is the daily photosynthetically active radiation for the month.

precipitation index β:

$$\beta = \begin{cases} 0, & \text{for } P \leq 10.219 \\ 0.0279\ P - 0.2851, & \text{for } 10.219 < P < 46.061 \\ 1, & \text{for } P \geq 46.061 \end{cases}$$

Where P is the average total monthly precipitation.

temperature index Γ:

$$\Gamma = \begin{cases} 0, & \text{for } -3 < T < 36 \\ -2 \times 10^{-7} T^5 + 0.13 \times 10^{-4} T^3 - 3.878 \times 10^{-3} T^2 + 0.1052 T + 0.352, & \text{otherwise} \end{cases}$$

Where T is the average daily temperature for the month.

monthly EPI:

$$\text{EPI}_{\text{month}} = \alpha \times \beta \times \Gamma$$

Five year total EPI is the sum of monthly EPIs multiplied by five:

$$\text{EPI}_{5 \text{ years}} = 5 \times \sum_{0}^{12} \text{EPI}_{\text{month}}$$

Predicted Biomass over a 5 year cropping cycle uses the parameters from

$$\text{Biomass} = 1.7607 \times \text{EPI} - 14.774$$

Simulating Rock Mulch

$$\beta_{\text{rock mulch}} = \beta * 1.83$$

Simulating Irrigation

Calculate monthly water demand:

Given $\beta = 0.0279 \times \text{P} - 0.2851$, find the values of P where $\beta \geq 1$?

$$\text{P}_{\beta=1} = \frac{1 + 0.2851}{0.0279} \simeq= 46.06093$$

- for each month if
  - $\alpha > 0$ AND
  - $\gamma > 0$ AND
  - P < 46.061mm
- calculate irrigation required = water demand $46.01 - P$
- annual irrigation demand = $\sum_{jan}^{dec} 46.01 - \text{P}$ calculate sum over all months

Simulating Freezing Mortality

For any grid cell with absolute minimum annual temperature < -10, set biomass to 0.

Figures

The file `figures.zip` contains all map figures generated using Panoply.

The folder also contains Panoply (plotting software) configuration files ending in `*.pcl`. These files provide information about Panoply software settings used to generate the figures using the command line PanoplyCL software that is available on request from the Panoply software author, Dr. Robert Schmunk.

Licenses

- **Code :** BSD 3-Clause

- **Data :** are 'public domain' CC-0: attribution requested in reuse

- TERRACLIMATE is public domain

- WPDA is available for public use, but not redistributable