Climate variability indices for ecological and crop models in R: the `climatrends` package

A preprint

Kauê de Sousa[1,2][*], Jacob van Etten[2], Svein Øivind Solberg[1]
[1] Department of Agricultural Sciences, Inland Norway University of Applied Sciences, 2318 Hamar, Norway
[2] Bioversity International, 00054 Maccarese, Rome, Italy
[*] Correspondence should be addressed to: kaue.desousa@inn.no

Summary

Abiotic factors play an important role in most ecological and crop systems that depend on certain levels of temperature, light and precipitation (and their interplay) to initiate important physiological events (Schulze et al., 2019). In the walk of climate change, understanding how these factors drive the physiological processes is a key approach to provide recommendations for adaptation and biodiversity conservation. The package `climatrends` aims to provide the methods in R (R Core Team, 2020) to compute precipitation and temperature indices that serve as input for climate and crop models (Kehel, Crossa, & Reynolds, 2016; van Etten et al., 2019), trends in climate change (Aguilar et al., 2005; de Sousa et al., 2018) and applied ecology (Prentice et al., 1992; Liu & El-Kassaby, 2018).

Methods and features

Implementation

Six main functions are provided (Table 1), `crop_sensitive()`, `ETo()`, `GDD()`, `late_frost()`, `rainfall()` and `temperature()` with a default method for numeric 'vector' and additional methods implemented via the package `methods` (R Core Team, 2020) for classes 'matrix' (or array), 'data.frame', and 'sf' (of geometry POINT or POLYGON) (Pebesma, 2018). The last two methods are designed to fetch data from cloud sources, until now from the packages `nasapower` (Sparks, 2018) and `chirps` (de Sousa, Sparks, Ashmall, van Etten, & Solberg, 2020).

Table 1: Main functions available in climatrends.

Function	Definition
crop_sensitive()	Compute crop sensitive indices
ETo()	Reference evapotranspiration using the Blaney-Criddle method
GDD()	Compute growing degree-days
late_frost()	Compute the occurrence of late-spring frost
rainfall()	Precipitation indices
temperature()	Temperature indices

These functions started as a set of scripts to compute indices in citizen science trials. In these trials. Aiming to capture the environmental variation across different sites, which can differ since each data-point generally have a different starting day and duration, the arguments `day.one` and `span` are vectorised and may be variable across data-points. For time series analysis, where fixed periods are defined across many locations, the indices can be adjusted with the argument `last.day` linked to the argument `day.one`.

Temperature and precipitation indices

The package `climatrends` computes 12 temperature indices and 10 precipitation indices that were suggested by previous research on climatology and crop science (Aguilar et al., 2005; Kehel et al., 2016). The indices computed by the functions `temperature()` and `rainfall()` are described in Table 2.

Table 2: Temperature and precipitation indices available in climatrends.

Index	Definition	Unit
maxDT	Maximun day temperature	°C
minDT	Minimum day temperature	°C
maxNT	Maximun night temperature	°C
minNT	Minimum night temperature	°C
DTR	Diurnal temperature range (mean difference between DT and NT)	°C
SU	Summer days, number of days with maximum temperature > 30 °C	days
TR	Tropical nights, number of nights with maximum temperature > 25 °C	days
CFD	Consecutive frosty days, number of days with temperature < 0 °C	days
WSDI	Maximum warm spell duration, consecutive days with temperature > 90th percentile	days
CSDI	Maximum cold spell duration, consecutive nights with temperature < 10th percentile	days
T10p	The 10th percentile of night temperature	°C
T90p	The 90th percentile of day temperature	°C
MLDS	Maximum length of consecutive dry day, rain < 1 mm	days
MLWS	Maximum length of consecutive wet day, rain >= 1 mm	days
R10mm	Heavy precipitation days 10 >= rain < 20 mm	days
R20mm	Very heavy precipitation days rain >= 20	days
Rx1day	Maximum 1-day precipitation	mm
Rx5day	Maximum 5-day precipitation	mm
R95p	Total precipitation when rain > 95th percentile	mm
R99p	Total precipitation when rain > 99th percentile	mm
Rtotal	Total precipitation in wet days, rain >= 1 mm	mm
SDII	Simple daily intensity index, total precipitation divided by the number of wet days	mm/days

Growing degree-days

Growing degree-days (gdd) is an heuristic tool in phenology that measures heat accumulation and is used to predict plant and animal development rates (Prentice et al., 1992). Growing degree-days are calculated by taking the integral of warmth above a base temperature (T_0). The function `GDD()` applies by default the following equation.

Equation [1]

$$GDD = \frac{T_{max} + T_{min}}{2} - T_0$$

where T_{max} is the maximum temperature in the given day, T_{min} is the minimum temperature in the given day and T_0 is the minimum temperature for growth (as per the physiology of the focal organism or ecosystem averages).

Additionally, the function `GDD()` offers three modified equations designed for cold environments and for tropical environments. For cold environments, where T_{min} may be lower than T_0, there are two modified equations that adjust either T_{mean} (variant a) or T_{min} (variant b). The variant a changes T_{mean} to T_0 if $T_{mean} < T_0$ and is expressed as follow.

Equation [2]

$$GDD = max\left(\frac{T_{max} + T_{min}}{2} - T_0,\ 0\right)$$

The variant b, is calculated using Equation 1, but adjusts T_{min} or T_{max} to T_0 if $T < T_0$, the equation is adjusted as follows.

Equation [3]

$$T < T_0 \ \rightarrow \ T = T_0$$

where T may refer to T_{min} and/or T_{max} when the condition of being below T_0 applies.

For tropical areas, where the temperature may surpass a maximum threshold ($T_{0_{max}}$), resulting in limited development, the minimum temperature is adjusted using Equation 3 and the maximum temperature is adjusted to a maximum base temperature as follow.

Equation [4]

$$T_{max} > T_{0_{max}} \ \rightarrow \ T_{max} = T_{0_{max}}$$

where $T_{0_{max}}$ is the maximum base temperature for growth, defined in `GDD()` using the argument `tbase_max`.

These modified equations are defined as 'a', 'b' and 'c', respectively, and can be selected using the argument `equation`.

By default, the function returns the degree-days that is accumulated over the time series using Equation 1. Additionally, the function may return the daily values of degree-days or the number of days that a given organism required to reach a certain number of accumulated degree-days. These values are defined by 'acc', 'daily' or 'ndays' and can be adjusted using the argument `return.as`. The required accumulated gdd is defined with argument `degree.days`. For example, the Korean pine (*Pinus koraiensis*) requires 105 $°C$ accumulated gdd to onset the photosynthesis (Wu, Guan, Yuan, Wang, & Jin, 2013). In that case, `GDD()` will calculate the growing degree-days (gdd) and sum up the values until it reaches 105 $°C$ and return the number of days required in the given season (GDD_r), as follows.

Equation [5]

$$\parallel GDD_r \parallel = \ ggd_1 \ + \ ... \ + \ gdd_n$$

where GDD_r is the length of the vector with accumulated degree-days from day 1 to n.

Late-spring frost

Late-spring frost is a freezing event occurring after a substantial accumulation of warmth. Frost damage is a known issue in temperate and boreal regions, it is associated with the formation of extracellular ice crystals that cause damage in the membranes (Lambers, Chapin III, & Pons, 2008). Freezing occurring after an advanced phenological stage during spring may harm some plant species, resulting in lost of productivity in crop systems (Trnka et al., 2014) and important ecological impacts (Zohner et al., 2020).

The function `late_frost()` supports the computation of late-spring frost events. The function counts for the number of freezing days with minimum temperature below a certain threshold (argument `tfrost`). And returns the number of days spanned by frost events (temperature below `tfrost`), latency (event with no freezing temperature but also no accumulation of growing degree-days) and warming (when growing degree-days are accumulated enabling the development of the target organism). Additionally the function returns the first day of the events. The function calculates the growing degree-days applying the variant b (Eq. 3), which can be adjusted using the argument `equation` passed to `GDD()` as explained in the later section. The main inputs are a vector with maximum and minimum temperatures to compute the degree-days, a vector of dates (argument `date`), and, if needed, the `tbase` and `tfrost`, set by default to 4 and -2 $°C$.

Crop-related indices

Two functions in **climatrends** are mainly designed to capture the effects of climate on the development and stress of crop species, `crop_sensitive()` computes indices that aim to capture the changes in temperature extremes during key phenological stages (e.g. anthesis), and `ETo()` computes the reference evapotranspiration.

The crop sensitive indices available in **climatrends** are described in Table 3. These indices were previously used in crop models to project the impacts of climate change on crop yield (Trnka et al., 2014; Challinor, Koehler, Ramirez-Villegas, Whitfield, & Das, 2016). Each index has a default temperature threshold(s) which can be adjusted by using the arguments `*.threshold`. Where the `*` means the index. For example, to change the defaults for hts_max (high temperature stress), a vector with the temperature thresholds is passed through the argument `hts_max.thresholds`.

The reference evapotranspiration measures the influence of the climate on a given plant's water need (Brouwer & Heibloem, 1986). The function `ETo()` applies the Blaney-Criddle method, a general theoretical method used when only air-temperature is available locally. It should be noted that this method is not very accurate and aims to provide the order of magnitude of evapotranspitation. The reference evapotranspiration is calculated using the following equation.

Equation [6]

$$ETo = p \times \left(0.46 \times \frac{T_{max} + T_{min}}{2} + 8\right) \times K_c$$

Where p is the mean daily percentage of annual daytime hours, T_{max} is the maximum temperature, T_{min} is the minimum temperature, and K_c is the factor for organism water need.

The percentage of daytime hours (p) is calculated internally by the 'data.frame' and 'sf' methods in `ETo()` using the given latitude (taken from the inputted `object`) and date (taken from the inputted `day.one`). It matches the latitude and date with a table of daylight percentage derived from Brouwer and Heibloem (1986). The table can be verified using `climatrends:::daylight`.

Table 3: Crop sensitive indices computed by climatrends.

Index	Definition	Default thresholds
hts_mean	High temperature stress using daily mean temperature, and given as percentage number of days a certain threshold is exceeded	32, 35, 38 °C
hts_max	High temperature stress using daily max temperature, and given as percentage number of days a certain threshold is exceeded	36, 39, 42 °C
hse	Heat stress event, and given as percentage number of days a certain threshold is exceeded for at least two consecutive days	31 °C
hse_ms	Heat stress event, and given the maximum number of days a certain threshold is exceeded for at least two consecutive days	31 °C
cdi_mean	Crop duration index using daily mean temperature, and given as max(Tmean - threshold, 0)	22, 23, 24 °C
cdi_max	Crop duration index using daily max temperature, and given as max(Tmax - threshold, 0)	27, 28, 29 °C
lethal	Lethal temperatures, defined as percentage of days during the timeseries where daily mean temperature exceeds a given threshold	43, 46, 49 °C

Examples

Common bean

During five growing seasons (from 2015 to 2017) in Nicaragua, van Etten et al. (2019) conducted a crowdsourcing citizen-science experiment testing 11 common bean varieties (*Phaseolus vulgaris* L.) in 842 farmer-managed plots. Sets of three varieties were allocated randomly to farms as incomplete blocks. A Plackett–Luce model was used to analyse the data, this model estimates for each variety the probability that it wins, beating all other varieties in the set (Turner, van Etten, Firth, & Kosmidis, 2020). An earlier version of `climatrends` was used in this research to capture the seasonal climate variation, here we reproduce part of this analysis regarding calculation and application of the climate indices. The approach here is slightly different because it considers the growing-degree days from planting date to maturity (the earlier study used planting date to the end of reproductive stage) and add new indices to illustrate the package implementation.

The data used here is available as supplementary material (**zenodo ref**) as `cbean`. This contains a data.frame with a Plackett-Luce grouped rankings, the geographical coordinates of each sampled plot and the planting dates from where each farmer decided to start the experiment. The planting dates differ from each other in the same season. The temperature data used was the land surface temperature MODIS (MYD11A2) (Wan, Hook, & Hulley, 2015) and is storaged as an array with two layers (1st for the day and 2nd for the night temperatures). Each column corresponds to the dates (from 2015-09-10 to 2017-06-09) and the rows corresponds to the rows in the `cbean` data.frame.

Since the phenological stages were not available, we estimate these stages based on the amount of growing degree-days required to reach a given stage using the function `GDD()`. For common beans, we define 900 degree-days, from planting date to maturity (de Medeiros, Daniel, & Fengler, 2016). The input data is the array with the temperature data, the vector with planting dates (`cbean$planting_date`), the required amount of degree-days passed to the argument `degree.days` and the character string 'ndays' specifying that the function must return the values as number of days. `GDD()` calls internally the function `get_timeseries()` which will match the given dates in `day.one` with the column names in the array and concatenate the values for each row. Then `GDD()` computes the degree-days for the time series and return the length of the vector where the accumulated gdd reached the pre-defined threshold (900).

The degree-days spanned from 54 to 100 days as shown in Fig. 1a. For simplicity we take the average per season and use this vector to compute the temperature indices.

```r
library("climatrends")
library("PlackettLuce")
library("tidyverse")

# compute the number of days required to accumulate
# gdd from planting date to maturity
gdd <- GDD(modis,
           day.one = cbean$planting_date,
           degree.days = 900,
           return.as = "ndays")

# add gdd to the cbean data and take the average
# of gdd per season
cbean %<>%
  mutate(gdd = gdd$gdd) %>%
  group_by(season) %>%
  mutate(gdds = as.integer(mean(gdd)))
```

To compute the temperature indices we use the array with tempeture data, the vector with planting dates, and the seasonal averaged degree-days passed as a vector using the argument `span`. `temperature()` concatenates the data from the given `day.one` to the given `span` and compute the indices for each row.

In van Etten (2019), a forward variable selection was applied to retain the most representative covariates based on the deviance reduction. This analysis retained the maximum night temperature (maxNT) as the most representative covariate. To illustrate how the Plackett-Luce trees can grow in complexity as we add more indices, we included the summer days (SU, number of days with maximum day temperature $> 30\ °C$) together with maxNT.

```r
# compute the temperature indices from planting date to the
# number of days required to accumulate the gdd in each season
temp <- temperature(modis,
                    day.one = cbean$planting_date,
                    span = cbean$gdds)

# combine the indices with the main data
cbean <- cbind(cbean, temp)

# fit a Plackett-Luce tree
plt <- pltree(G ~ maxNT + SU, data = cbean, minsize = 50)
```

Across-season distribution of maxNT captured for each sample plot in this experiment is shown in Fig. 1b. The data has a bimodal distribution which is reflected in the splitting value ($18.7\ °C$) for the Plackett-Luce trees in Fig. 1c. The upper node splits with 49 summer days (SU). We can interpret these results as that differences in growing performance of common beans is led by a considerable amount of diurnal temperature above a warmer threshold of $30\ °C$ (in this case >70% of the growing days) and warmer nights ($> 18.7\ °C$).

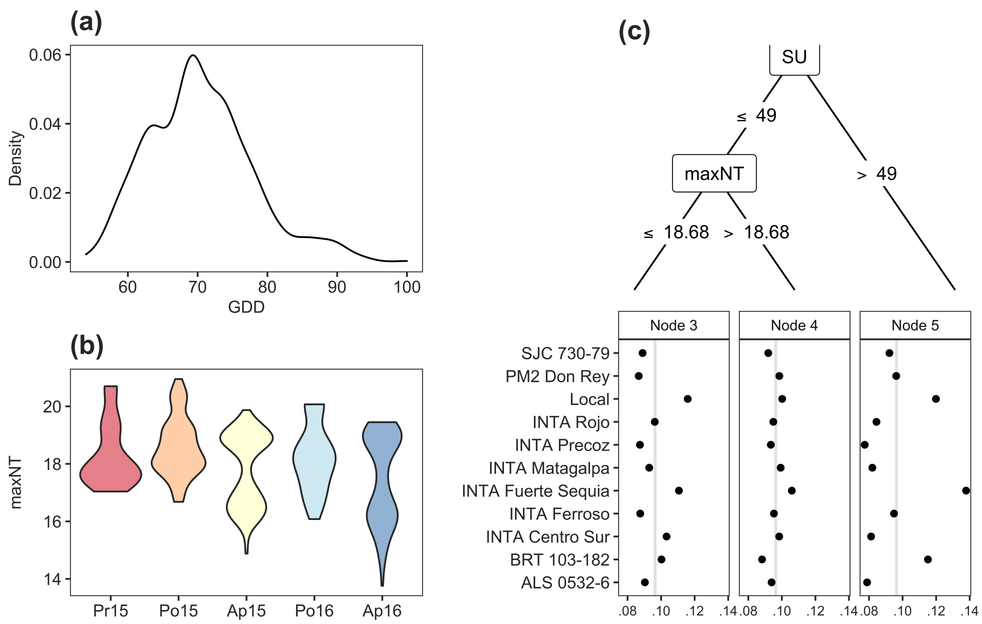

Fig. 1. Application of climatrends functions to support the analysis of a citizen-science data testing 11 common bean varieties in Nicaragua. (A) Days required to reach 900 growing-degree days from planting date calculated using the function GDD(). (B) Maximum night temperature (°C) distributed across seasons computed using the function temperature(). (C) Plackett-Luce Tree showing the probability of one common bean variety has to win against the others (axys X) in three different nodes splitted with the summer days (day temperature > 30 °C) and maximum night temperature (°C). Note: the first season (primera, Pr) spans from May to August, the second (postrera, Po) from September to October, and the third (apante, Ap) from November to January.

Trends in climate variability in Norway and Sweden

We randomly selected 100 points in hexagonal within the coordinates 7° and 17° W, and 59 ° and 63 ° N, that comprises Norway and Sweden before the Arctic Circle. We compute the temperature indices from 2000-01-01 to 2019-12-31 using the function `temperature()` with the method for objects of class 'sf'. The temperature data is fetched from from the NASA Langley Research Center POWER Project funded through the NASA Earth Science Directorate Applied Science Program (https://power.larc.nasa.gov/), using the R package `nasapower` (Sparks, 2018).

```r
library("climatrends")
library("sf")
library("nasapower")

# create a polygon within the coordinates 7, 17, 59, 63
e <- matrix(c(7, 59, 17, 59, 17, 63,
              7, 63, 7, 59),
            nrow = 5, ncol = 2, byrow = TRUE)

e <- st_polygon(list(e))

# sample 100 points in the hexagonal type
p <- st_sample(e, 100, type = "hexagonal")
p <- st_as_sf(p, crs = 4326)

# compute the temperature indices using the random points
temp <- temperature(p,
                    day.one = "2000-01-01",
                    last.day = "2019-12-31",
                    timeseries = TRUE,
                    intervals = 365)
```

We then select the indices CSDI (cold spell duration of night temperature), WSDI (warm spell duration of day temperature), and their associated indices the T10p (the 10th percentile of night temperature) and T90p (the 90th percentile of day temperature), in Figure 2. Plots are generated with `ggplot2` (Wickham, 2016) and `patchwork` (Pedersen, 2020).

The trends show a decrease in the cold spell duration (number of consecutive cold nights bellow the 10th percentile) and warm spell duration (number of consecutive warm days above the 90th percentile). However, the values of the percentiles show an increase over the time series. The T10p index shows a decrease around the year of 2010, but again rises up to the a value around the -10 °C, meaning that the could nights are becoming a bit warmer over the time. The T90p index also shows an increase in the temperature across the sampled area, with the average 90th percentile rising from ~ 16 °C to ~ 18 °C over the time series.

Further development

The package can support the integration with other datasets as they become available in `R` via API client packages. Also new indices related to the physiology of crops could be implemented.

Acknowledgements

This work was supported by The Nordic Joint Committee for Agricultural and Food Research (grant num. 202100-2817). We thank Julian Ramirez-Villegas and Marcel Schrijvers-Gonlag for the useful insights and discussion that helped in the development of this study.

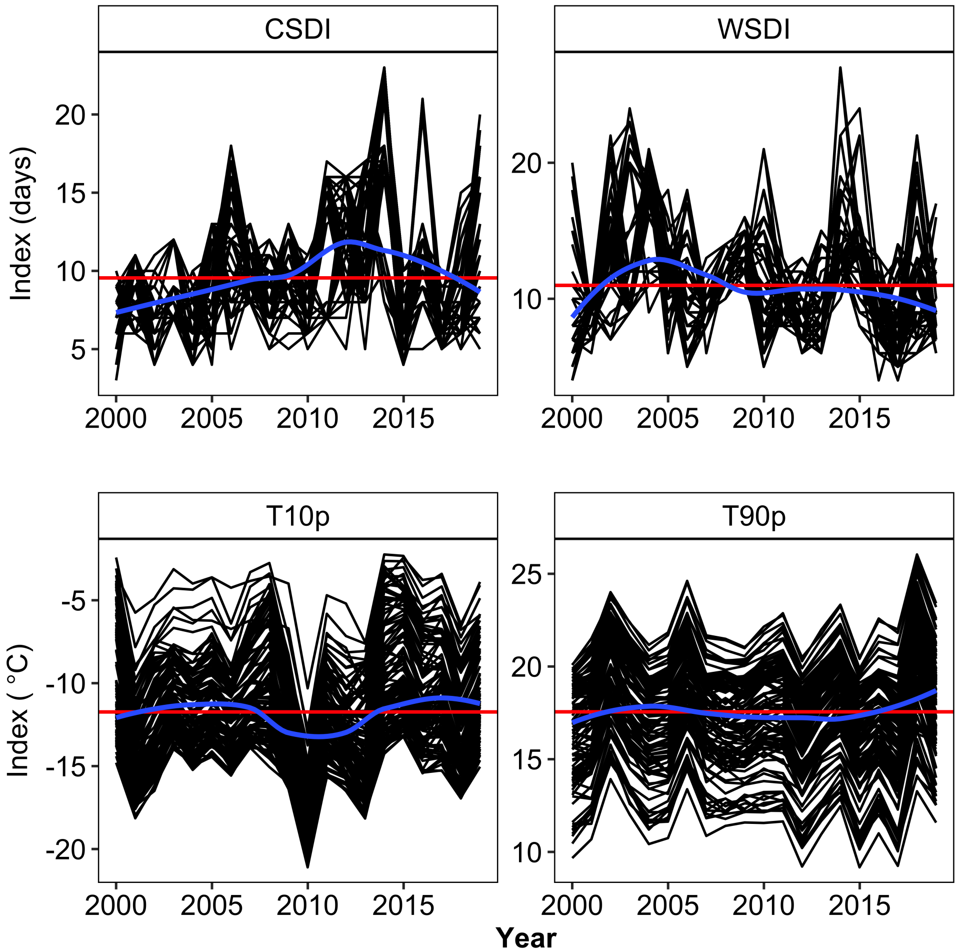

Fig. 2. Trends in temperature indices across Southern Norway and Sweeden from 2000 to 2019. CSDI, maximum cold spell duration, consecutive nights with temperature < 10th percentile. WSDI, maximum warm spell duration, consecutive days with temperature > 90th percentile. T10p, the 10th percentile of night temperature. T90p, the 90th percentile of day temperature. Red line indicates the historical mean of each index in the time series. Blue line indicates the smoothed trends in each index using the 'loess' method.

Data availability statement

To explore the latest functionalities of `climatrends`, please check the package's updates at CRAN (https://cran.r-project.org/package=climatrends).

References

Aguilar, E., Peterson, T. C., Obando, P. R., Frutos, R., Retana, J. A., Solera, M., … Mayorga, R. (2005). Changes in precipitation and temperature extremes in Central America and northern South America, 1961–2003. *Journal of Geophysical Research*, *110*(D23), D23107. doi:10.1029/2005JD006119

Brouwer, C., & Heibloem, M. (1986). *Irrigation water management: Irrigation water needs* (Training m). Rome, Italy: Food; Agriculture Organization of The United Nations. Retrieved from http://www.fao.org/3/S2022E/s2022e00.htm

Challinor, A. J., Koehler, A.-K., Ramirez-Villegas, J., Whitfield, S., & Das, B. (2016). Current warming will reduce yields unless maize breeding and seed systems adapt immediately. *Nature Climate Change*, *6*(10), 954–958. doi:10.1038/nclimate3061

de Medeiros, G. A., Daniel, L. A., & Fengler, F. H. (2016). Growth, Development, and Water Consumption of Irrigated Bean Crop Related to Growing Degree-Days on Different Soil Tillage Systems in Southeast Brazil. *International Journal of Agronomy*, *2016*, 8065985. doi:10.1155/2016/8065985

de Sousa, K., Casanoves, F., Sellare, J., Ospina, A., Suchini, J. G., Aguilar, A., & Mercado, L. (2018). How climate awareness influences farmers' adaptation decisions in Central America? *Journal of Rural Studies*, *64*, 11–19. doi:10.1016/j.jrurstud.2018.09.018

de Sousa, K., Sparks, A., Ashmall, W., van Etten, J., & Solberg, S. (2020). chirps: API Client for the CHIRPS Precipitation Data in R. *Journal of Open Source Software*, *5*(51), 2419. doi:10.21105/joss.02419

Kehel, Z., Crossa, J., & Reynolds, M. (2016). Identifying Climate Patterns during the Crop-Growing Cycle from 30 Years of CIMMYT Elite Spring Wheat International Yield Trials. In A. Bari, A. B. Damania, M. Mackay, & S. Dayanandan (Eds.), *Applied mathematics and omics to assess crop genetic resources for climate change adaptive traits* (pp. 151–174). CRC Press.

Lambers, H., Chapin III, F. S., & Pons, T. L. (2008). *Plant Physiological Ecology* (Second Edi, p. 620). doi:10.1007/978-0-387-78341-3

Liu, Y., & El-Kassaby, Y. A. (2018). Evapotranspiration and favorable growing degree-days are key to tree height growth and ecosystem functioning: Meta-analyses of Pacific Northwest historical data. *Scientific Reports*, *8228*. doi:10.1038/s41598-018-26681-1

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

Pedersen, T. L. (2020). *patchwork: The Composer of Plots*. Retrieved from https://CRAN.R-project.org/package=patchwork

Prentice, I. C., Cramer, W., Harrison, S. P., Leemans, R., Monserud, R. A., & Solomon, A. M. (1992). Special Paper: A Global Biome Model Based on Plant Physiology and Dominance, Soil Properties and Climate. *Journal of Biogeography*, *19*(2), 117. doi:10.2307/2845499

R Core Team. (2020). R: A language and environment for statistical computing. version 4.0.0. Vienna, Austria: CRAN R Project. Retrieved from https://www.r-project.org/

Schulze, E.-D., Beck, E., Buchmann, N., Clemens, S., Müller-Hohenstein, K., & Scherer-Lorenzen, M. (2019). *Plant Ecology* (Second Edi, p. 925). Berlin, Germany: Springer. doi:10.1007/978-3-662-56233-8

Sparks, A. H. (2018). nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and Climatology Data Client for R. *Journal of Open Source Software*, *3*(30), 1035. doi:10.21105/joss.01035

Trnka, M., Rötter, R. P., Ruiz-Ramos, M., Kersebaum, K. C., Olesen, J. E., Žalud, Z., & Semenov, M. A. (2014). Adverse weather conditions for European wheat production will become more frequent with climate change. *Nature Climate Change*, *4*(7), 637–643. doi:10.1038/nclimate2242

Turner, H. L., van Etten, J., Firth, D., & Kosmidis, I. (2020). Modelling rankings in R: the PlackettLuce package. *Computational Statistics*. doi:10.1007/s00180-020-00959-3

van Etten, J., de Sousa, K., Aguilar, A., Barrios, M., Coto, A., Dell'Acqua, M., … Steinke, J. (2019). Crop variety management for climate adaptation supported by citizen science. *Proceedings of the National Academy of Sciences*, *116*(10), 4194–4199. doi:10.1073/pnas.1813720116

Wan, Z., Hook, S., & Hulley, G. (2015). MYD11A1 MODIS/Aqua Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V006 [Data set]. *NASA EOSDIS LP DAAC*. doi:10.5067/MODIS/MYD11A1.006

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Wu, J., Guan, D., Yuan, F., Wang, A., & Jin, C. (2013). Soil Temperature Triggers the Onset of Photosynthesis in Korean Pine. *PLoS ONE*, *8*(6), e65401. doi:10.1371/journal.pone.0065401

Zohner, C. M., Mo, L., Renner, S. S., Svenning, J.-C., Vitasse, Y., Benito, B. M., … Crowther, T. W. (2020). Late-spring frost risk between 1959 and 2017 decreased in North America but increased in Europe and Asia. *Proceedings of the National Academy of Sciences*, 201920816. doi:10.1073/pnas.1920816117