Clim2Agri

CGIAR INITIATIVE ON Foresight

User guide for extracting and sub-setting geospatial climate data from gridded products for their use in crop modelling and agricultural research

version 1.0

Carlo Montes, Francisco Gómez

December 2023

Contents

Overview

Multiple information platforms currently provide fully open-access data and services from satellite Earth observations, ground observations and model outputs (Buontempo et al., 2020). These platforms are largely the result of extensive international programs coordinated by specialized agencies such as the National Aeronautics and Space Administration (NASA) of the United States (e.g., Rienecker et al., 2011), the European Centre for Medium Range Weather Forecasts (ECMWF) (e.g., C3S, 2019), or outputs from research and development programs led by research institutions associated with universities or national services. The data generated by these active platforms are provided following multiple formats and structures, which may differ considerably depending on the objectives and applications for which they were originally conceived. For example, the increasingly demanded outputs from General Circulation Models (GCM) from both historical periods and future projections are typically very large datasets of three-dimensional fields at high temporal resolutions (i.e., daily time steps) that can be difficult to handle if computing capacities are not appropriate. Moreover, agricultural research applications usually require surface weather variables (i.e., 2-m height), which in the case of GCMs must be extracted as the lower of multiple vertical levels at which the atmosphere has been discretized. Added to the above, it is common for the agricultural research community to use data formats different from those in which the original products were generated, which in some cases can represent a bottle neck since the data processing tools are not necessarily those designed to process, for example, geophysical data (e.g., Network Common Data Form NetCDF, Hierarchical Data Format HDF, General Regularly-distributed Information in Binary form GRIB).

In response to the above-described problem, a set of data processing tools (Clim2Agri) that allow the agricultural research and development community to simply and efficiently access data from open access platforms to be used in applications such as modeling or analysis of observations was generated. Clim2Agri is a Python-based application allowing the extraction and subsetting of agriculturally-relevant climate and land surface data from multiple gridded products (e.g., atmospheric reanalysis, satellite-derived products), from multiple formats (e.g., netCDF4, HDF), and their transformation into formats that are required by the agriculture research and crop modeling community (e.g., csv, txt). The platform is hosted at the CGIAR Foresight Initiative GitHub repository[1] and will allow users to extract the data

[1] https://github.com/ForesightInitiative/clim2agri

of interest (geospatial, single point, and time series subset), select a specific area using a shapefile, and extract the data in the selected format.

Based on feedback from the CGIAR agricultural research and crop modeling community, Clim2Agri incorporates data from multiple platforms and output formats. The user will select in the Python code the variables to be extracted and the data product to generate the output file for a specific geographical point. The user provides the geographical latitude and longitude for which the data from the closest grid point will be extracted. Clim2Agri generates output files formats according to specific crop models, which is also part of the input parameters (e.g., DSSAT, AquaCrop, APSIM). All the combinations of variables and products are supported and explained in this document.

The CGIAR Initiative on Foresight combines state-of-the-art analytics, innovative use of data, and close engagement with national, regional, and global partners to offer better insights into alternative transformation pathways that can inform choices and sharpen decision-making today, leading to more productive, sustainable and inclusive food, land and water systems in the future.

The foresight initiative is dedicated to making data and metrics and models and tools that are needed for foresight analysis available, accessible, transparent, and usable for decision-makers regarding policies and investment decisions with respect to the transformation of food, land and water systems under global change and the on-going climate crisis. As part of this endeavor for radical democratization of foresight analysis the Foresight Initiative provides as many tools and models through the Initiative's GitHub repository https://github.com/ForesightInitiative.

Purpose

The purpose of this report is to provide the user documentation of Clim2Agri.

Structure of the report

The report provides a brief description of some key crop growth models. It then describes the main forcing variables and the related input files for crop models. The report then discusses some key gridded meteorological products. The report then goes into the configuration of the environment for Clim2Agri. It discusses how to access meteorological data and how to actually do all the steps.

Details of the crop model input files are found in the Appendix.

Version history

This is the initial version of the documentation for Clim2Agri version 1.0

Intended audience

Crop modellers.

Factsheet

Key	Value
General information	
Name	Clim2Agri
Version	1.0
Language	Python and Jupyter
Contact	Carlo Montes < c.montes@cgiar.org >
Status	Released
Latest release date	December 2023
System requirements	
Operating system	any
Software requirements	Python plus Jupyter Notebooks
General information	
Key features	Preparation of crop model input files
Key inputs	Gridded meteorological data
Key outputs	Crop model input files
License info and acknowledgements	
License	GPL-3.0 license
Acknowledgements	Funding was provided by the CGIAR foresight Initiative

Crop Simulation models

Crop simulation models consist of a set of mathematical formulations used to simulate crop growth, yields, water consumption, among other processes. Their main application is to reduce the uncertainty about future production, and improve irrigation management. The input variables required to run crop models is highly variable, although according to the literature2 there is a set of main input variables that should always be used as input. Regarding the meteorological data, such variables correspond to temperature (daily minimum and maximum), daily global solar radiation and rainfall. Nevertheless, there are more advanced and robust models that also require data such as wind speed, evapotranspiration, vapor pressure, among others, which depends, for example, on the number of processes simulated by the model.

There is a large number of crop models being used nowadays. Below we show those among the most cited so far, and provide a brief description obtained from their respective website.

APSIM

Reference: https://doi.org/10.1016/0308-521X(94)00055-V

Website: http://www.apsim.info

Description: "The Agricultural Production Systems sIMulator (APSIM) is internationally recognized as a highly advanced platform for modeling and simulation of agricultural systems. It contains a suite of modules that enable the simulation of systems for a diverse range of plant, animal, soil, climate and management interactions. APSIM is undergoing continual development, with new capability being added to APSIM Next Generation. Its development and maintenance is underpinned by rigorous science and software engineering standards. The APSIM Initiative was established in 2007 to promote the development and use of the science modules and infrastructure software of APSIM. The current members are CSIRO, The State of Queensland, The University of Queensland, AgResearch Ltd. (NZ), University of Southern Queensland, Iowa State University, (US) and Plant and Food Research (NZ)."

Aquacrop

Reference: https://doi.org/10.2134/agronj2008.0139s

Website: http://www.fao.org/aquacrop

Description: "AquaCrop is a crop growth model developed by FAO's Land and Water Division to address food security and assess the effect of the environment and management on crop production. AquaCrop simulates the yield response of herbaceous crops to water and is particularly well suited to conditions in which water is a key limiting factor in crop production. AquaCrop balances accuracy, simplicity and robustness. To ensure its wide applicability, it uses only a small number of explicit parameters and mostly intuitive input variables that can be determined using simple methods."

CropSyst

Reference: https://doi.org/10.1016/S1161-0301(02)00109-0

Website: http://modeling.bsyse.wsu.edu/CS_Suite_4/CropSyst/index.html

[2] https://link.springer.com/chapter/10.1007/978-94-017-3624-4_2

Description: "CropSyst is a user-friendly, conceptually simple but sound multi-year multi-crop daily time step simulation model. The model has been developed to serve as an analytic tool to study the effect of cropping systems management on productivity and the environment. The model simulates the soil water budget, soil-plant nitrogen budget, crop canopy and root growth, dry matter, yield, residue and decomposition, and erosion. Management options include: cultivar selection, crop rotation (including fallow years), irrigation, nitrogen fertilization, tillage operations (over 80 options), and residue management."

DSSAT

Reference: https://doi.org/10.1016/S1161-0301(02)00107-7

Website: http://dssat.net/

Description: "Decision Support System for Agrotechnology Transfer (DSSAT) is software application program that comprises dynamic crop growth simulation models for over 42 crops. DSSAT is supported by a range of utilities and apps for weather, soil, genetic, crop management, and observational experimental data, and includes example data sets for all crop models. The crop simulation models simulate growth, development and yield as a function of the soil-plant-atmosphere dynamics. DSSAT has been applied to address many real-world problems and issues ranging from genetic modeling to on-farm and precision management, regional assessments of the impact of climate variability and climate change, economic and environmental sustainability, and food and nutrition security. DSSAT has been used for more than 30 years by researchers, educators, consultants, extension agents, growers, private industry, policy and decision makers, and many others in over 187 countries worldwide."

EPIC

Reference: https://elibrary.asabe.org/abstract.asp?aid=31032 doi: 10.13031/2013.31032

Website: https://epicapex.tamu.edu/

Description: "The Environmental Policy Integrated Climate (EPIC) model was developed to estimate soil productivity as affected by erosion and simulates approximately eighty crops with one crop growth model using unique parameter values for each crop. It can be configured for a wide range of crop rotations and other vegetative systems, tillage systems, and other management strategies. It predicts effects of management decisions on soil, water, nutrient and pesticide movements, and their combined impact on soil loss, water quality, and crop yields for areas with homogeneous soils and management."

Oryzav3

Reference: https://doi.org/10.1016/j.agrformet.2017.02.025
https://doi.org/10.1016/j.agsy.2004.09.011

Website: https://sites.google.com/a/irri.org/oryza2000/home?authuser=0

Description: "ORYZA version 3 (ORYZA v3), or simply ORYZA, is an ecophysiological model which simulates growth and development of rice including water, C, and N balance (Bouman et al., 2001[3]; IRRI, 2013) in lowland, upland, and

[3] Bouman, B.A.M., Kropff, M.J., Tuong, T.P., Wopereis, M.C.S., Ten Berge, H.F.M., & Van Laar, H.H. (2001). ORYZA2000: modeling lowland rice. International Rice Research Institute, Los Baños, Philippines, and Wageningen University and Research Centre, Wageningen, Netherlands, 235 pp.

aerobic rice ecosystems. It works in potential, water-limited, nitrogen-limited, and NxW-limited conditions. And it was calibrated and validated for 18 popular rice varieties in 15 locations throughout Asia."

STICS

Reference: http://dx.doi.org/10.1051/agro:19980501

Website: https://www6.paca.inrae.fr/stics_eng

Description: "The crop model known as STICS (Simulateur mulTIdisciplinaire pour les Cultures Standard, or multidisciplinary simulator for standard crops) was created at INRA, the French national institute for agricultural research, in 1996 on the initiative of Nadine Brisson with informatics support from Dominique Ripoche. The creators built the STICS model from "pieces" by bringing together the GOA (plant), BYM (water), and LIXIM (nitrogen) models, which had essentially been produced by two INRA teams, one in Avignon (with Nadine Brisson) and one in Laon (with Bruno Mary). At the beginning, STICS simulated two main crops, wheat and corn, and was used for the first time as part of the ECOSPACE project (1997) to simulate agricultural production and nitrate leaching on the basis of soil heterogeneity."

WOFOST

Reference: https://doi.org/10.1111/j.1475-2743.1989.tb00755.x

Website:https://www.wur.nl/en/research-results/research-institutes/environmental-research/facilities-tools/software-models-and-databases/wofost.htm

Description: "WOFOST (WOrld FOod STudies) is a simulation model for the quantitative analysis of the growth and production of annual field crops. It is a mechanistic, dynamic model that explains daily crop growth on the basis of the underlying processes, such as photosynthesis, respiration and how these processes are influenced by environmental conditions."

Main forcing variables for crop models

The table below contains the selected crop models and main input meteorological variables. Each variable has its units

Table 1: Variables and corresponding units for the selected crop models.

Variable and units		APSIM	Aquacrop	CropSyst	DSSAT	EPIC	Orizav3	STICS	WOFOST
Min T	°C	✓	✓	✓	✓	✓	✓	✓	✓
Max T	°C	✓	✓	✓	✓	✓	✓	✓	✓
Pp	mm/day	✓	✓	✓	✓	✓	✓	✓	✓
Radiation	MJ/m^2	✓	✓**	✓*	✓	✓	✓, kJ/m^2	✓	✓, kJ/m^2
Wind speed	m/s		✓**	✓*	✓*, km/day	✓	✓	✓*	✓, 2m
RH			✓**	✓*, max		✓		✓*	
Vapor pressure	kPa						✓		✓
ETo	mm		✓**					✓*	
T avg	°C				✓				
T amplitude	°C				✓				
Ref. HWe	m				✓*				
Ref. HWi	m				✓				
T dew	°C				✓*				
PAR	$moles/m^2/day$				✓*				

next to the name of the variable, but when the model requires the input in different units, this is indicated in the corresponding box. Table 2 provides a glossary of the notation used.

The definition of the variables used in Table 1 can be found below in Table 2.

Table 2: Definition for the short-hand notation used in Table 1.

T	Temperature
Pp	Precipitation
RH	Relative Humidity
ETo	Reference evapotranspiration
Avg	Average (T avg in DSSAT refers to the average for the whole period)
Ref. HWe	Reference height for weather measurement
Ref. HWi	Reference height for wind measurement
PAR	Photosynthetic active radiation
*	Optional input
**	Aquacrop requires ETo, but it can be replaced by radiation, windspeed, and vapor pressure.

.

Input files for crop models

The input files vary largely among crop models, although all of them have a similar structure. The model input files typically resemble a text file where the first couple of rows are filled with metadata such as the name of the station, location, etc. There is also a series of columns containing the necessary information for the model to run, as shown in the previous section. The extension of the file varies accordingly. It was not mentioned before, but the date of the measurement/data is also mandatory in such input files.

Below the reader will find a snapshot with the above-mentioned structure, which does not necessarily coincide with any particular model but at least gives an idea of how the data files should be prepared. However, in Appendix A the reader is going to find screenshots of the different input files that each of the crop models requires.

weather.txt

```
##  The metadata is located here
##
##  Name of the station
##  Location
##  Altitude
##  Measurement period
##
##  Any comments regarding the data e.g. which crop it corresponds, etc.
##


## Then some models may require a summary row with the averages of each and/or other variables

    Temperature,    Windspeed,  Precipitation
    (°C)       ,    (m/s)    ,  (mm/day)
    15         ,    10       ,  0.5

## Afterwards, typically you will find the data

        Date,   Max. Temperature,   Min. Temperature,   Precipitation,  Windspeed
(DD-MM-YYYY),   (°C)            ,   (°C)            ,   (mm/day)     ,  (m/s)
  05-06-2023,   25              ,   10              ,   0            ,  5
  06-06-2023,   27              ,   8               ,   0            ,  3
  07-06-2023,   24              ,   9               ,   0            ,  2
  08-06-2023,   22              ,   12              ,   1            ,  2
  09-06-2023,   23              ,   10              ,   5            ,  3
  10-06-2023,   25              ,   13              ,   0            ,  5
  11-06-2023,   28              ,   15              ,   0            ,  8
  ...       ,   ...             ,   ...             ,   ...          ,  ...
```

Figure 1: Example of a generic meteorological input file for a crop model.

Gridded meteorological products

Gridded meteorological products are generated and delivered by different agencies around the world to provide meteorological data in a common grid format for multiple applications. Each product contains different weather variables, obtained by different approaches (satellite information, weather stations, model outputs), and with different grid size (spatial resolution), temporal resolution, and latency[4]. Some agencies make their data publicly available, and most of the time they provide Application Programming Interfaces (APIs), or other tools to make the downloading process easy to automatize.

The gridded meteorological products shown below were chosen by considering the following features:

- They are widely used in the scientific literature,
- They have daily temporal resolution,
- Their spatial resolution is not less than 0.1° × 0.1° (~11 km),
- A near real time latency (no more than a couple of weeks),
- They provide multiple variables.

Table 3 below shows the different gridded meteorological products selected, and the variables that could be extracted to use in the aforementioned crop models:

Table 3: List of variables provided by each selected gridded meteorological product.

	AgERA5	CHIRPS	IMERG	MSWEP	MSWX	NASA POWER	PERSIANN
Temperature	✓				✓	✓	
Radiation	✓					✓	
Precipitation	✓	✓	✓	✓	✓	✓	✓
Windspeed	✓					✓	
T dew	✓					✓	
RH	✓						
Vapor pressure	✓						

[4] Latency refers to the time a particular product incorporates new data.

As it can be seen from the table above, there are multiple products that provide precipitation only. The reason for including them was because precipitation is among the most difficult variables to accurately measure for a particular location, so the use and assessment of multiple data sources is recommended, to then choose an adequate product for the specific context and use.

Below is a brief description of each of the selected gridded products, containing the temporal and spatial resolution, latency, and the means by which the particular agency generates the variables. The link to their websites will also be provided in case of needing more details.

AgERA5

Website: https://cds.climate.copernicus.eu/cdsapp#!/dataset/sis-agrometeorological-indicators?tab=overview

Spatial resolution: 0.1° × 0.1°

Temporal resolution: Daily

Latency: 1 month

Description: "This dataset provides daily surface meteorological data for the period from 1979 to present as input for agriculture and agro-ecological studies. This dataset is based on the hourly ECMWF ERA5 data at surface level and is referred to as AgERA5. Acquisition and pre-processing of the original ERA5 data is a complex and specialized job. By providing the AgERA5 dataset, users are freed from this work and can directly start with meaningful input for their analyses and modeling. To this end, the variables provided in this dataset match the input needs of most agriculture and agro-ecological models. Data were aggregated to daily time steps at the local time zone and corrected towards a finer topography at a 0.1° spatial resolution. The correction to the 0.1° grid was realized by applying grid and variable-specific regression equations to the ERA5 dataset interpolated at 0.1° grid. The equations were trained on ECMWF's operational high-resolution atmospheric model (HRES) at a 0.1° resolution. This way the data is tuned to the finer topography, finer land use pattern and finer land-sea delineation of the ECMWF HRES model."

CHIRPS

Website: https://www.chc.ucsb.edu/data/chirps

Spatial resolution: 0.05° × 0.05°

Temporal resolution: Daily

Latency: 3 weeks

Description: "Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) is a 35+ year quasi-global rainfall data set. Spanning 50°S-50°N (and all longitudes) and ranging from 1981 to near-present, CHIRPS incorporates our in-house climatology, CHPclim, 0.05° resolution satellite imagery, and in-situ station data to create gridded rainfall time series for trend analysis and seasonal drought monitoring."

IMERG

Website: https://gpm.nasa.gov/data/imerg

Spatial resolution: 0.1° × 0.1°

Temporal resolution: 30 min, 3-hourly, daily

Latency: Up to 12 hours - 3.5 months for the more processed version

Description: "The Integrated Multi-satellitE Retrievals for GPM (IMERG) algorithm combines information from the GPM satellite constellation to estimate precipitation over the majority of the Earth's surface. This algorithm is particularly valuable over the majority of the Earth's surface that lacks precipitation-measuring instruments on the ground. Now in the latest Version 06 release of IMERG the algorithm fuses the early precipitation estimates collected during the operation of the TRMM satellite (2000 - 2015) with more recent precipitation estimates collected during operation of the GPM satellite (2014 - present). The longer the record, the more valuable it is, as researchers and application developers will attest. By being able to compare and contrast past and present data, researchers are better informed to make climate and weather models more accurate, better understand normal and extreme rain and snowfall around the world, and strengthen applications for current and future disasters, disease, resource management, energy production and food security."

MSWEP

Website: https://www.gloh2o.org/mswep

Spatial resolution: 0.1° × 0.1°

Temporal resolution: Daily

Latency: 3 hours

Description: "MSWEP incorporates daily gauge observations and accounts for gauge reporting times to reduce temporal mismatches between satellite-reanalysis estimates and gauge observations. Near real-time estimates are available with a latency of ~3 hours. MSWEP is compatible with GloH2O's operational Multi-Source Weather (MSWX); MSWX forecasts can thus be used to extend MSWEP into the future. MSWEP tends to exhibit better performance than other precipitation products in both densely gauged and ungauged regions (see the Performance section on this page, Beck et al., 2017, and Beck et al., 2019)."

MSWX

Website: https://www.gloh2o.org/mswx/

Spatial resolution: 0.1° × 0.1°

Temporal resolution: Daily

Latency: Around 4.5 hours the NTR-version - 5 days the more processed version

Description: "Other meteorological products, such as ERA5, HydroGFD, PGF, and WFDE5, are not available in near real-time, lack freely available forecasts, and have a coarse spatial resolution (≥0.25°). MSWX combines the best data sources for each time-scale and eliminates systematic biases to provide an effective and readily available solution for use in operational modeling applications. In addition, MSWX is compatible with GloH2O's Multi-Source Weighted-Ensemble Precipitation (MSWEP) product, which merges gauge, satellite, and reanalysis data to obtain the highest quality precipitation estimates."

NASA POWER

Website: https://power.larc.nasa.gov/docs/methodology/

Spatial resolution: 1° x 1° for radiation, and 0.5° x 0.625° for other meteorological data

Temporal resolution: Daily

Latency: From a couple of weeks to a couple of hours (depending on data set)

Description: "National Aeronautics and Space Administration (NASA), through its Earth Science research program, has long supported satellite systems and research providing data important to the study of climate and climate processes. These data include long-term climatologically averaged estimates of meteorological quantities and surface solar energy fluxes. Additionally, mean daily values of the base meteorological and solar data are provided in time series format. These satellite and model-based products have been shown to be sufficiently accurate to provide reliable solar and meteorological resource data over regions where surface measurements are sparse or nonexistent. The products offer two unique features: the data is global and generally contiguous in time."

PERSIANN [5]

Website: https://chrsdata.eng.uci.edu/

Spatial resolution: 0.25° × 0.25°

Temporal resolution: Daily

Latency: 2 days

Description: "The current operational PERSIANN (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks) system developed by the Center for Hydrometeorology and Remote Sensing (CHRS) at the University of California, Irvine (UCI) uses neural network function classification/approximation procedures to compute an estimate of rainfall rate at each 0.25° x 0.25° pixel of the infrared brightness temperature image provided by geostationary satellites. An adaptive training feature facilitates updating of the network parameters whenever independent estimates of rainfall are available. The PERSIANN system was based on geostationary infrared imagery and later extended to include the use of both infrared and daytime visible imagery. The PERSIANN algorithm used here is based on the geostationary longwave infrared imagery to generate global rainfall. Rainfall product covers 60°S to 60°N globally

[5] There exists a version of Persiann with higher spatial resolution (0.04°), but the data available is from 1983 to 2020.

Configuring the environment for Clim2Agri

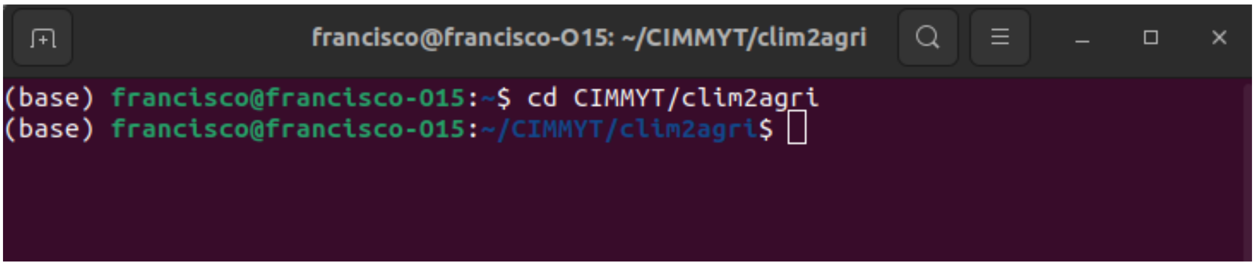
In this section, we show how to configure a Python virtual environment in order to run the code correctly[6]. In order to carry on the following steps, the used will need to have installed Python version 3[7] and the pip package[8]. The list of instructions is as follows:

Downloading:

1. Download and then decompress the folder.

Create Virtual Environment:

2. Open the terminal and access the folder you have just decompressed. This can be accomplished with the 'cd' command followed by the path to the folder. The picture below shows an example.



3. In order to create the virtual environment, you should write the following command, where 'name_of_env' is an arbitrary name for the environment you are creating.

>> python3 -m venv name_of_env

Create Jupyter Kernel:

[6] The creation of the environment is not necessary but highly recommended. If you wish you could skip this step and just install the needed packages in step 5.

[7] https://www.python.org/

[8] https://pypi.org/project/pip/

4. The Jupyter Kernel can be created as follows

```
>> source name_of_env/bin/activate
>> ipython kernel install --name name_of_env --user
```

Installing necessary packages:

5. Now we need to install all the needed packages for the code to run, which can be done with the following command.

```
>> pip install -r requirements.txt
```

Now that the virtual environment is configured, you can open the Jupyter notebook and start using it.

Access to gridded meteorological data

The Jupyter notebook can already be used to download variables from any of the selected meteorological gridded products and to create input files for crop models or other applications. However, some additional steps are necessary to download data from the gridded products , for which we encourage to follow the instructions from each product website.

IMERG

In this case you need to create an account on the EarthData website[9]. Later, you should run the code 'imerg.py', which is located in the compressed file Clim2Agri. The code will ask you to provide your username and password. After running this code, some folders will be created in your directory in order to access IMERG data.

MSWEP and MSWX

Both MSWEP and MSWX require for you to create an account on their website[10] so that they can share a drive folder with the data. Be aware that you need to follow the instructions in their website to have access through Python and download the data through the Jupyter notebook we provide.

What is the most efficient way to download data from the Google Drive?

Downloading data from shared Google Drive folders is relatively easy using cloud store manager software such as rclone. The following instructions explain how to set up rclone to download MSWEP:

1. Download and install rclone.
2. Link rclone to your Google account by following the steps in this video.
3. Access the MSWEP shared folder by visiting it via your browser. Check if the shared folder is listed under "Shared with me" on your Google Drive page.
4. Confirm that rclone can find the shared folder:

```
$ rclone lsd --drive-shared-with-me GoogleDrive:
          -1 2021-02-03 10:14:35        -1 MSWEP_V280
```

5. Download daily and monthly MSWEP data from the shared folder to your local drive:

```
$ rclone sync -v --exclude 3hourly/ --drive-shared-with-me GoogleDrive:/MSWEP_V280 c:/temp/MSWEP_V280
          2021/02/03 11:09:02 INFO  : Past/Monthly/202007.nc: Copied (new)
          2021/02/03 11:09:02 INFO  : Past/Monthly/202002.nc: Copied (new)
          2021/02/03 11:09:02 INFO  : Past/Monthly/202005.nc: Copied (new)
          ...
```

The video referenced in the instructions is the following: https://www.youtube.com/watch?v=vPs9K_VC-lg

[9] https://urs.earthdata.nasa.gov/oauth/authorize?response_type=code&redirect_uri=http%3A%2F%2Fdisc.gsfc.nasa.gov%2Flogin%2Fcallback&client_id=C_kKX7TXHiCUqzt352ZwTQ

[10] https://www.gloh2o.org/mswep/ and https://www.gloh2o.org/mswx/

AgERA5

In order to download data from AgERA5, first you need to create an account on https://cds.climate.copernicus.eu/cdsapp#!/home and then create an API key so that you can have access through their API. All the instructions can be found on their website.

How to…?

The Jupyter notebook named `main_codes.ipynb` was made with the objective of being self-explanatory and without the need of running any other additional codes. Therefore, we consider the best way to learn how to use it is by following the steps provided in the Jupyter notebook itself. Nevertheless, here we summarize the steps you need to accomplish in order to download data and/or create input files.

Creating the weather_variables.csv file

This is a crucial first step. Without the `weather_variables.csv` file, you will not be able to run any of the codes below. In order to create this file, there is a built-in function which is:

>> create_empty_csv()

Once you run the line above, a new and empty `weather_variables.csv` file will be created in your directory. Be aware that if you already have such a file and you run this function, then all the data that you might have will be deleted.

The motivation for creating this file is twofold. The first one is to have a place to save and visualize the data that is being downloaded. In addition, a .csv file is relatively easy to work with and perform additional analysis. The second motivation is that you might want to add data from other sources, e.g. local meteorological stations, data from other reanalysis, etc. This can be done by simply pasting the data in the `weather_variables.csv` file. The only caveat is that the units of the variables should match the units shown in the corresponding column.

Downloading Data

To download data, you just need to run the following line:

>> download_variable(ini_date,fin_date,lat,lon)

Where `ini_date`, and `fin_date` correspond to the initial and final date of the period you want to download, and `lat` and `lon` are the coordinates of the point of interest.

After running the line above, a message will be deployed on your screen asking you to choose the variable you want to download. Once you select the variable by typing the corresponding number, another message will appear asking you from which meteorological product you want to extract the particular variable you chose. Recall the data will be saved in the `weather_variables.csv` file. You can download as many variables and from as many gridded meteorological products as you want, you just need to repeat the above-explained sequence.

Creating Input Files

In this step, you just need to run one of the following lines depending on the input file you want to generate.

>> build_APSIM_input_file(lat,lon)

>> build_AQUACROP_input_file(lat,lon)

>> build_CROPSYST_input_file(lat,lon)

>> build_DSSAT_input_file(lat,lon)

>> build_EPIC_input_file(lat,lon)

```
>> build_ORIZA_input_file(lat,lon)
>> build_STICS_input_file(lat,lon)
>> build_WOFOST_input_file(lat,lon)
```

The file generated will be a .txt extension file containing the name of the crop model. Therefore, you will need to change the name of the file and the extension before using it in the corresponding crop model.

References

Abatzoglou, J.T., S.Z. Dobrowski, S.A. Parks, and K.C. Hegewisch. 2018. Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015. *Scientific Data*. 5, 170191.

Brocca, L.L., C.M. Ciabatta, T. Moramarco, S. Hahn, S. Hasenauer, R. Kidd, W. Dorigo, W. Wagner, and V. Levizzani. 2014. Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data. *Journal of Geophysical Research: Atmospheres*. 119, 5128-5141.

Buontempo, C., and Coauthors. 2020. Fostering the development of climate services through Copernicus Climate Change Service (C3S) for agriculture applications. *Weather and Climate Extremes*. 27,100226.

Copernicus Climate Change Service (C3S). 2019. Data Stream 2: AgERA5 historic and near real time forcing data. Product User Guide and Specification. Copernicus Climate Change Service Climate Data Store (CDS). Online: http://datastore.copernicus-climate.eu/documents/sis-global-agriculture/C3S422Lot1.WEnR.DS2_ProductUserGuideSpecification_v2.2.pdf

Funk, C., and Coauthors. The climate hazards infrared precipitation with stations - A new environmental record for monitoring extremes. *Scientific Data*. 2, 150066.

Pinzon, J.E., and C.J. Tucker. 2014. A non-stationary 1981–2012 AVHRR NDVI3g time series. *Remote Sensing*. 6, 6929-6960.

Rienecker, M., and Coauthors. 2011. MERRA: NASA's Modern-Era retrospective analysis for research and applications. *Journal of Climate*. 24, 3624–3648.

Vicente-Serrano, S.M., S. Beguería, and J.I. López-Moreno. 2010. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of Climate*. 23, 1696-1718.

Zhu, Z., J. Bi, Y. Pan, S. Ganguly, A. Anav, L. Xu, A. Samanta, S. Piao, R. R. Nemani, and R. B. Myneni. 2013. Global data sets of vegetation Leaf Area Index (LAI)3g and Fraction of Photosynthetically Active Radiation (FPAR)3g derived from Global Inventory Modeling and Mapping Studies (GIMMS) Normalized Difference Vegetation Index (NDVI3g) for the period 1981 to 2011. *Remote Sensing*. 5, 927-948.

Annex A: Detailed input files

From Fig. A1 to A8 we show the typical input file structure for each of the selected crop models. Most of the files are self-explanatory, but when needed there is a short description to better understand the corresponding file.

APSIM

apsim.txt
~/CIMMYT/clim2agri

```
site = somewhere
latitude =  #e.g. -30.1
longitude =
tav =
amp =
 year day radn    maxt mint rain
 ()   ()  (MJ/m2) (oC) (oC) (mm)
 1996 65  20      29   20.5 0
 1996 66  20      29   20   0
 1996 67  20      29   21   0
 1996 68  20      22.5 21   5
 1996 69  20      27.5 18.5 2
 ...  ... ...     ...  ...  ...
```

Figure A.1: APSIM input file.

Aquacrop

Aquacrop.txt
~/

```
      Day      Month     Year     Tmin(C) Tmax(C) Prcp(mm)    Eto(mm)
      1        1         2023     3.0     10.6    5.3     0.3
      2        1         2023     4.1     11.8    2.7     0.5
      3        1         2023     6.1     7.2     0.2     0.9
      4        1         2023     1.6     6.5     2.6     0.5
      5        1         2023     6.6     9.4     3.4     0.2
      ...      ...       ...      ...     ...     ...     ...
```

Figure A.2: Aquacrop input file.

CropSyst

The input file for CropSyst contains a column for the day of the year, and then for each meteorological variable a new column is added to the right. In Fig. A.3, the sequence of input variables are precipitation, maximum and minimum temperature, and solar radiation. The file could contain precipitation only, and minimum and maximum temperature. If needed, the user can add solar radiation, maximum relative humidity, and wind speed.

CropSyst.txt
~/

```
1 0 7.3 4.4 2
2 0.2 6 4 2.1
3 0 8 3 5
4 0 9 2 5
5 0.3 7 4 1.9
...
```

Figure A.3: CropSyst input file. The rows represent: Day of year, precipitation, maximum temperature, minimum temperature, and solar radiation.

DSSAT

DSSAT.txt

```
*WEATHER DATA: NAME OF STATION

@   INSI     LAT      LONG    ELEV     TAV     AMP  REFHT    WNDHT
    XXXX  16.780  120.950      210    25.0     3.0    2.0      2.0

@DATE    SRAD     TMAX     TMIN     RAIN
13001    17.7     28.2     12.8      0.0
13002    18.2     32.5     15.3      0.1
13003    18.6     32.8     16.7      0.5
13004    17.2     31.8     15.9      1.0
13005    18.8     32.1     15.3      0.0
...       ...      ...      ...      ...
```

Figure A.4: DSSAT input file

EPIC

In this input file one must add the name of the weather file in the first line, then in the second line the numbers of years followed by the beginning year. In the third line, one should add the following rows: year, month, day, solar radiation, maximum temperature, minimum temperature, precipitation, relative humidity and wind speed. Fig. A.5 depicts an example of an input file for this model.

EPIC.txt

```
## Weather file name
## Number of years followed by the beggining year e.g. if the data starts
## year 2023 and there are 10 years of data then: 102023
  2023   1   1   2.5 18.5    9.6   1.6   0.8   5.1
  2023   1   2   2.8 25.1   10.4   0.0   0.7     2
  2023   1   3   2.2 18.6    8.1   1.0   0.3   3.3
  2023   1   4   2.0 17.5    7.3   1.0   0.3   1.2
  2023   1   5   2.3 15.3    8.0   0.0   0.2   0.0
   ...
```

Figure A.5: EPIC input file

Oryzav3

Oryza3v.txt

```
*-----------------------------------------------------------

* Station Name :

* Author : Climate Unit, IRRI                -99.: nil value

* Source :

*

* Comments  : This file is extracted from CLICOM database.

* Longitude :             Latitude:         Altitude:

*

* Column Daily Value

* 1  Station number

* 2  Year

* 3  Day

* 4  irradiance            KJ m-2 d-1

* 5  min temperature              oC

* 6  max temperature              oC

* 7  vapor pressure              kPa

* 8  mean wind speed            m s-1

* 9  precipitation             mm d-1

*-----------------------------------------------------------

121.25, 14.18, 21.00,0.000,0.000

1,1992, 1,  16379.00, 21.60, 29.00, 2.38, 1.90,  0.00

1,1992, 2,  15911.00, 22.20, 29.20, 2.51, 2.30,  0.00

1,1992, 3,  16631.00, 21.50, 29.00, 2.50, 2.50,  0.00

1,1992, 4,  18251.00, 22.10, 29.00, 2.52, 2.40,  0.00

1,1992, 5,  11519.00, 21.60, 29.00, 2.61, 1.10,  0.90

...
```

Figure A.6: Example of Oryzav3 input file

STICS

The data in the input file for the STICS model is organized per column in the following order: name of the station, year, month, day of the month, day of the year, minimum temperature, maximum temperature, radiation, potential evapotranspiration, precipitation, wind speed, relative humidity, and CO_2 concentration. In Fig. A.7, an example of an input file for the STICS model is presented.

*STICS.txt

```
...
NAME 1994  1 29  29    0.5    9.5    3.9    0.7    0.0 -999.9 -999.9  330.0
NAME 1994  1 30  30    6.0    9.7    2.7    1.2    0.5 -999.9 -999.9  330.0
NAME 1994  1 31  31   -1.7    7.8    5.9    0.3    0.0 -999.9 -999.9  330.0
NAME 1994  2  1  32    2.7    8.9    2.2    1.0    0.0 -999.9 -999.9  330.0
NAME 1994  2  2  33    4.3    8.5    2.3    0.5   12.0 -999.9 -999.9  330.0
...
```

Figure A.7: STICS input file

WOFOST

In this crop model the first couple of lines there is general information about the location of interest. The first line after the last star '*' corresponds to a list of: latitude, longitude, elevation, and A and B coefficients of the Angstrom formula. Afterwards, the columns represent: number of the station, year, day of the year, minimum temperature, maximum temperature, radiation, vapor pressure, wind speed, and precipitation.

WOFOST.txt
~/CIMMYT/clim2agri

```
*--------------------------------------------------------------------------
*
* Station name: Wageningen (Haarweg), Netherlands
* Year: 1987
* Author: Willem Stol en Vitale Nuzzo
* Source: Natuur- en Weerkunde - data invoer: Hennie de Ruiter
(…)
*
** WCCDESCRIPTION=Netherlands, Wageningen
** WCCFORMAT=2
** WCCYEARNR=1987
*--------------------------------------------------------------------------
*
5.67 51.97 7. -0.18 -0.55

1 1987 1 470. 3.0 7.9 0.770 2.8 13.0
1 1987 2 620. -3.9 7.3 0.660 5.4 2.7
1 1987 3 2420. -6.4 4.0 0.500 1.9 0.1
(…)
1 1987 168 12750. 5.3 13.6 1.060 3.0 3.5
1 1987 169 18130. 6.9 15.9 1.020 3.2 0.0
1 1987 170 7950. 6.9 14.4 1.260 2.1 7.0
(…)
1 1987 363 710. 8.3 11.9 1.010 5.2 2.9
1 1987 364 1400. 7.8 11.3 1.100 3.5 5.0
1 1987 365 2000. 7.4 10.0 0.980 3.8 0.0
```

Figure A.8: Example of a WOFOST input file.

[LOGO]
[LOGO]
[LOGO]
[LOGO]
[LOGO]

Name, Title, emailaddress@cgiar.org

Name, Title, emailaddress@cgiar.org

We would like to thank all funders who support this research through their contributions to the CGIAR Trust Fund: www.cgiar.org/funders.

To learn more about this Initiative, please visit this webpage.

To learn more about this and other Initiatives in the CGIAR Research Portfolio, please visit www.cgiar.org/cgiar-portfolio

CGIAR

INITIATIVE ON
Foresight