**Weather historical datasets verification portal**

The purpose of this portal is to enable users to conduct personalized analyses to assess the precision of historical weather datasets accessible on the CE Hub for a specific variable and country/region/point. To evaluate the accuracy of each dataset, statistical metrics are utilized to compare its data against weather station data.

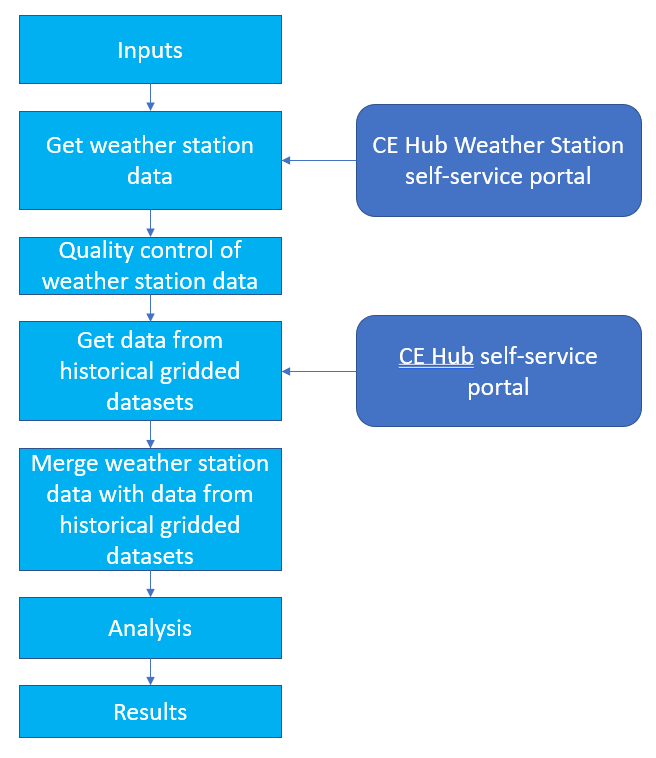


Figure : High-level workflow of the code

The following sections refer to the Jupyter Notebook: Automated\_framework\_historical.ipynb

**Inputs**

**Basic Inputs**

* Variable array (for example ['E\_AIR\_TEMPERATURE','E\_RELATIVE\_HUMIDITY', 'E\_RAINFALL','E\_DEWPOINT','E\_WIND\_SPEED']

*In the code the variable name is the same as for CE Hub weather station self-service portal.*

*In the portal we would like:*

* + *Checklist menu for the variables (multiple selection)*
  + *Display the variable name as: Air temperature, Relative humidity, etc…*
* Aggregation array

For each variable above we should specify the aggregation (daily\_mean, daily\_max, daily\_min, daily\_sum)

*In the portal we would like:*

* *Dropdown menu for each selected variable*
* *Just one selection*
* *For precipitation only daily\_sum*
* Time aggregation: daily, weekly or monthly

*In the portal we would like:*

* + *Daily as default*
* Possibility to run the analysis for one or multiple countries, regions **or** points
  + Country name (as in CE Hub WS self-service portal)
  + Region name (as in CE Hub WS self-service portal)
  + Point: latitude and longitude and radius. The code look for the WS within this radius (not developed yet)
* Start and end date with this format '2021-01-01T00:00:00'

**Historical Gridded Datasets**

* Array of CE Hub gridded datasets

*In the portal we would like to display the table “Datasets\_info.xlsx” and allow the selection of the row of the table (multiple selection)*

**Advanced Section**

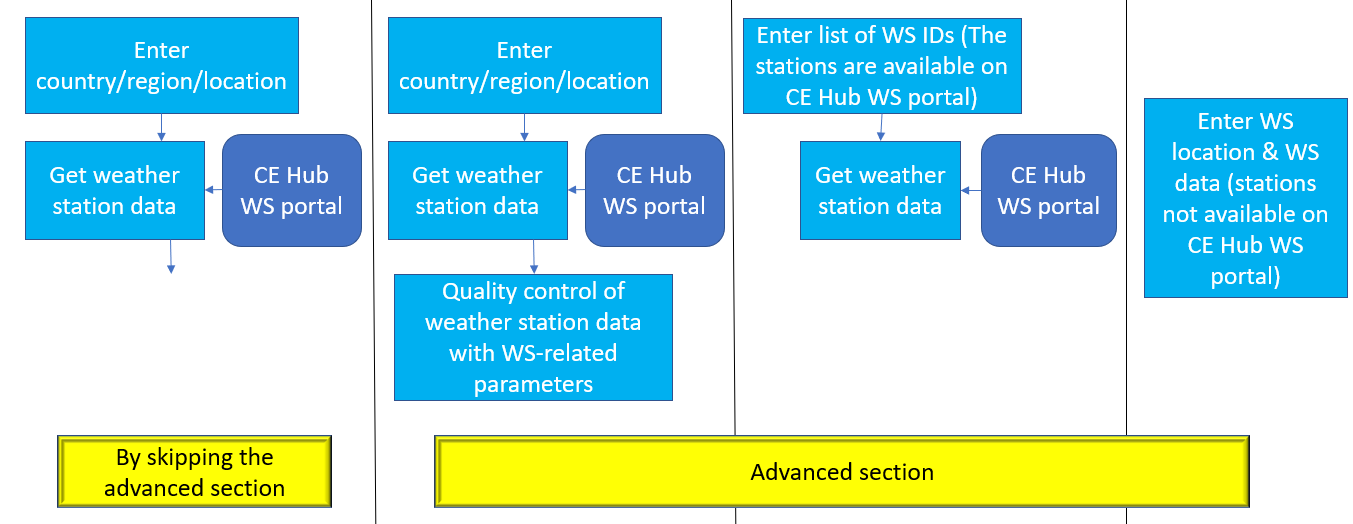
**🡪** It can be skipped and the analysis will be done with all the WS for the region/country/point from CE Hub WS portal

🡪 In the advanced section the user can tune the WS-related parameters (minimum number of observations and minimum completeness) to disregard the WS from CE Hub WS portal with low quality.

*The user should have the flexibility to play with the inputs to obtain the desired WS. Allow the user to see the WS on a map, the time series and WS by country for each variable (WS info 1, 2, 3).*

🡪 In the advanced section the user can enter the list of station ID (stations available on the CE Hub WS portal)

🡪 In the advanced section the user can enter the coordinates of their stations and the weather observations



Inputs for this section:

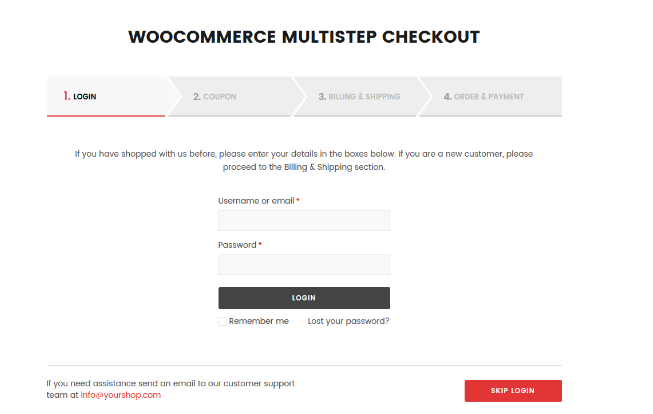
* Weather stations-related parameters
* Suppliers: *default [“Davis”,”Arable”,”Pessl”,”Sencrop”]. Checklist menu for all the possible suplliers*

🡪 In the advanced section the user can enter its data from an external gridded dataset

**Type of Output**

Array of statistical metrics: *Display the table “Metrics\_info.xlsx” and allow a checklist selection of the row of the table*

*The previous inputs should be displayed as a multistep checkout as in the pictures below.*



Additional inputs

* account\_name=Syngenta email to create a bucket on CE Hub S3
* analysis\_name to create a bucket on CE Hub S3

key\_analysis=key + account\_name + '/' + analysis\_name

* AWS Keys and CE Hub’s keys from .env file

**Quick start**

**Scope:** Create the bucket for the analysis on CE Hub S3 and merge the variable and aggregation to simplify the code later

**Identification of WS based on the selection location/country/region**

**Scope:** Identify the station ID of the CE Hub’s weather station within the country/countries, region/regions or nearby the location (function not developed for the point).

This section is skipped if the user enters in the advance section the stationID of CE Hub’s WS or coordinates of external weather stations.

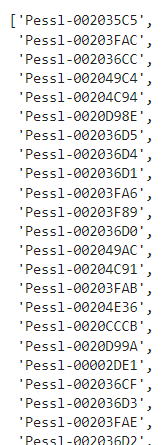


Figure : List of WS IDs

**Observations Data**

**Scope:** Download by means of API WS data from CE Hub WS portal.

This section is skipped if the user enters in the advance section the coordinates of external weather stations and their timeseries.

**Quality Control**

**Scope:** Quality control of CE Hub WS.

Checks:

### Above or equal minimum numbers of observations (input from the user)

### Completeness above or equal min\_completeness (input from the user)

This section is skipped if the user enters in the advance section the stationID of CE Hub’s WS or coordinates of external weather stations.

**Visualization of the retained WS**

**Scope:** Display the **retained weather stations** for each variable

* on the **map** (WS info 1)
* *Possibility to unselect from the map the WS*
* *Possibility to see the time series of a WS by selecting it from the map*
* timeseries (WS info 2)
* number of WS by country (if multiple countries are prescribed as input or a region) (WS info 3)

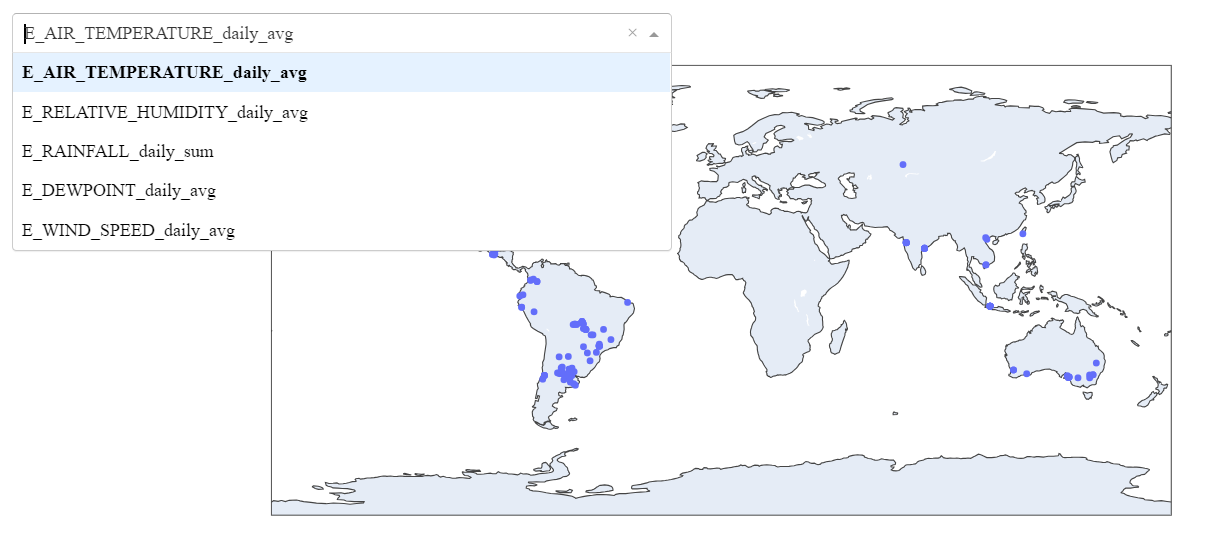


Figure : WS info 1

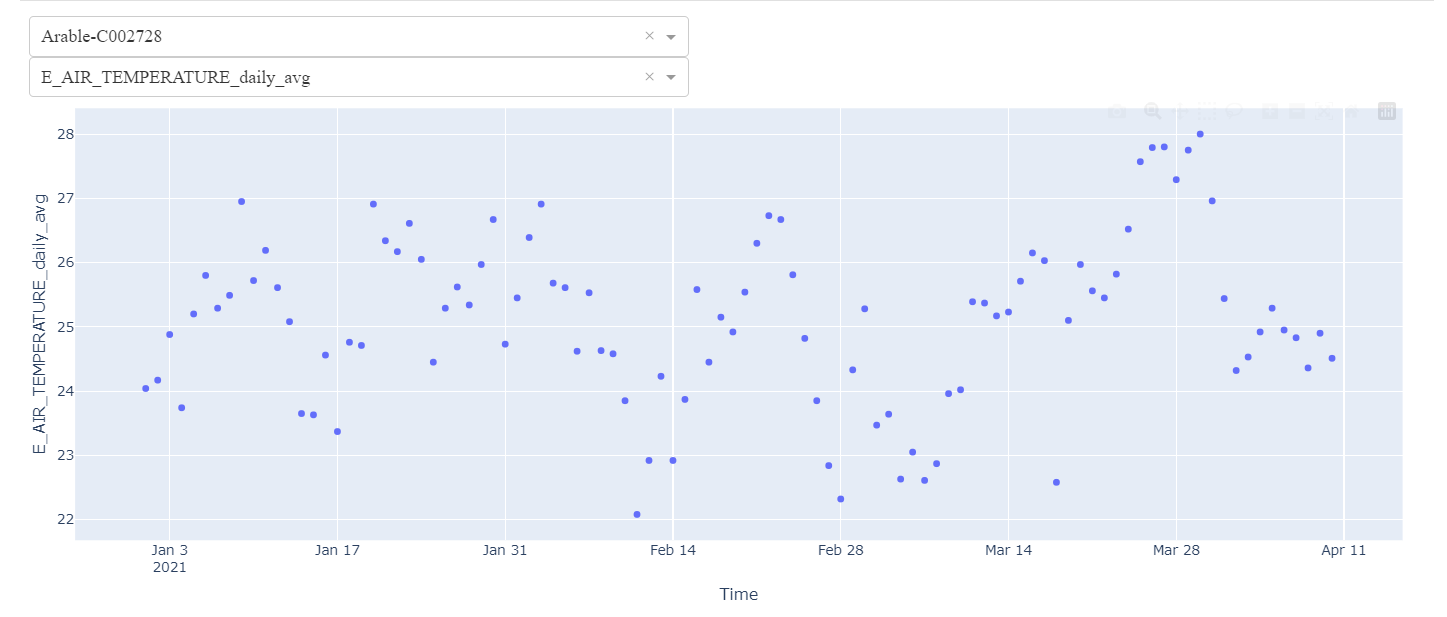


Figure : WS info 2

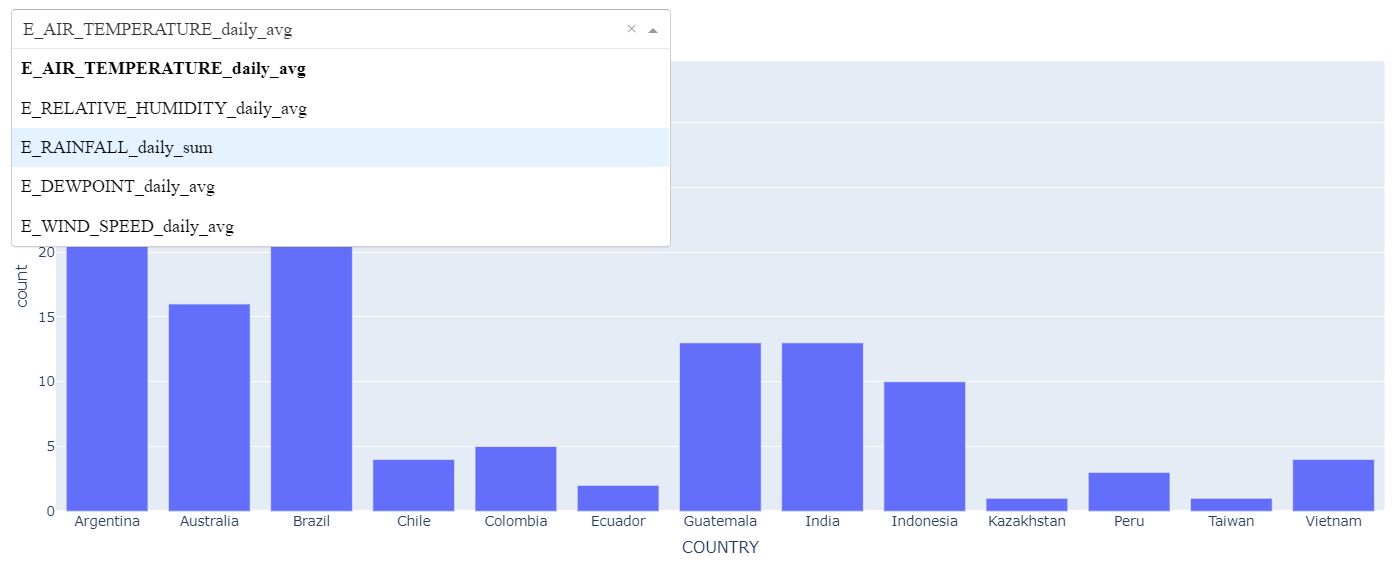
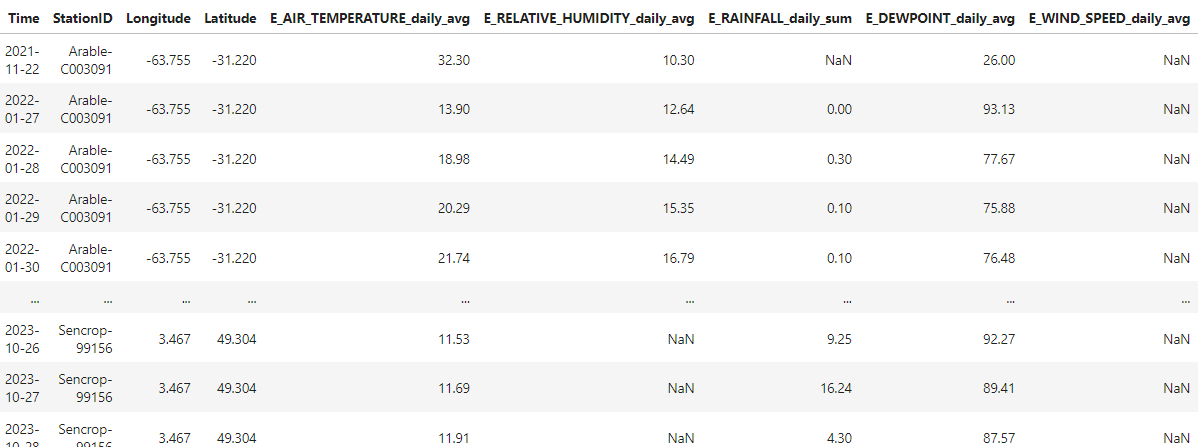


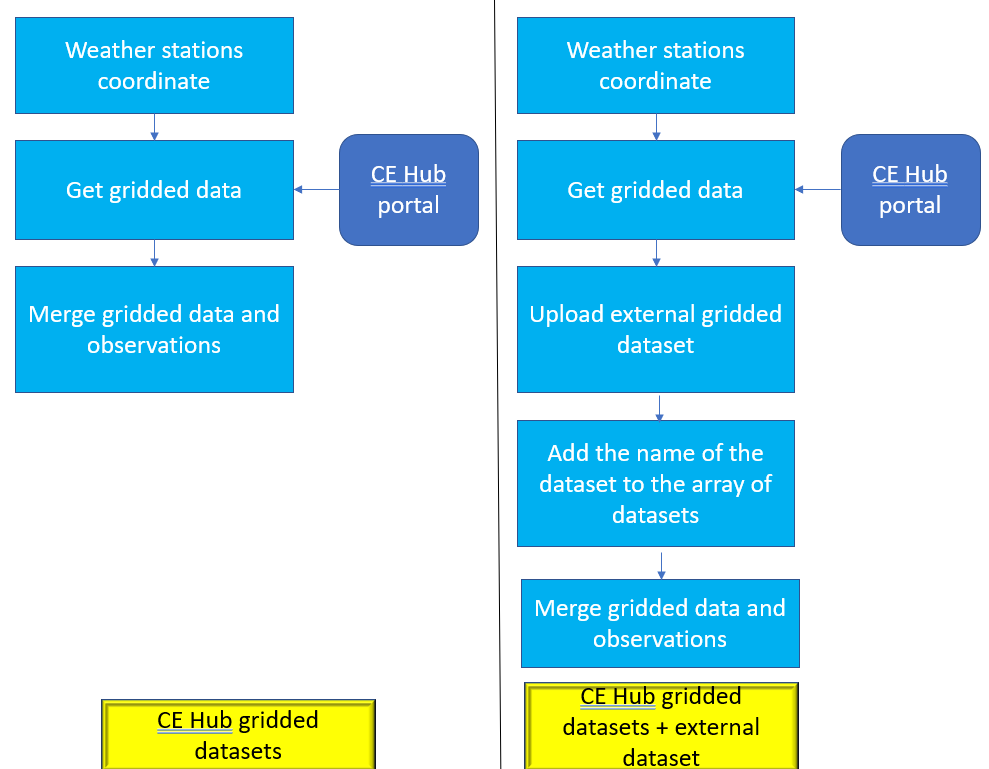
Figure : WS info 3

**Extraction of gridded datasets from CE Hub**

**Scope:** Based on latitude and longitude of the retained WS, the code gets gridded weather data from CE Hub self-service portal

It is also possible to enter a weather data from an external dataset using the advanced section. The dataset should have a daily resolution and this format: Time, StationID, Latitude, Longitude, variable+aggregation





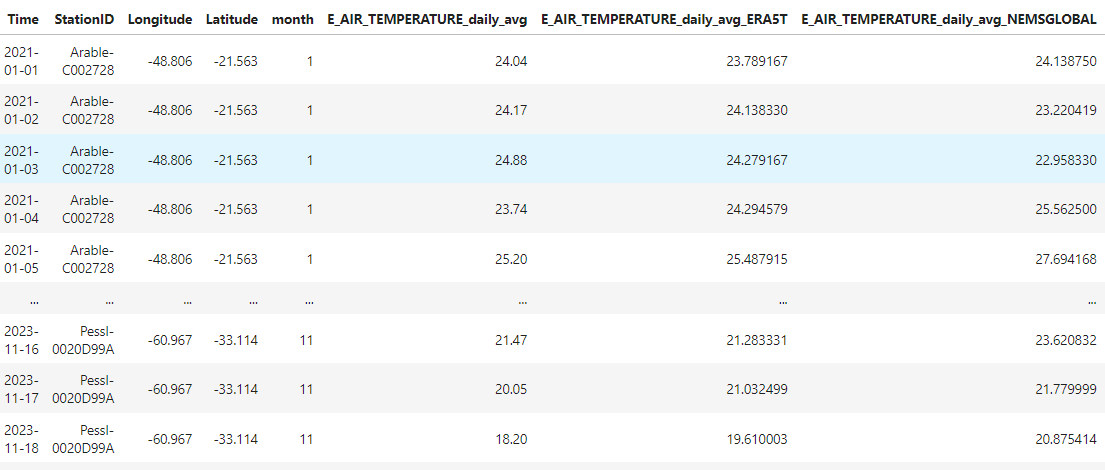
**Merge gridded datasets with WS observations**

**Scope:** Merge gridded weather data with WS observations based on StationID and Time.

**Time aggregation**

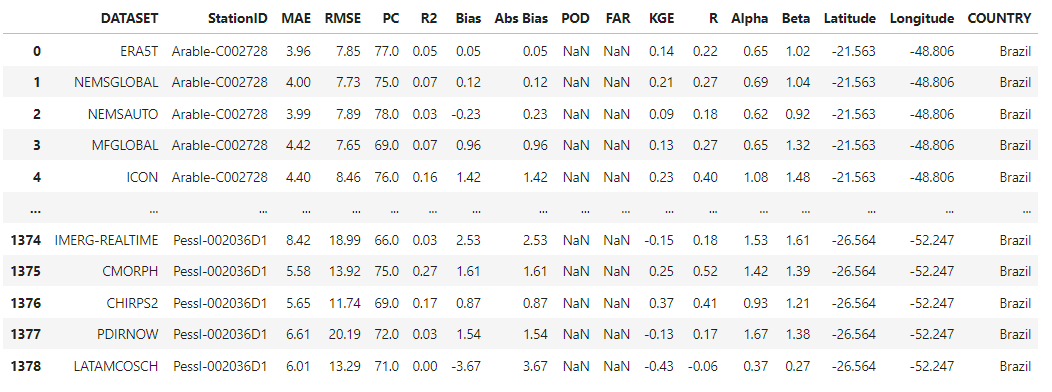
**Scope:** Aggregate data (gridded & observations), if selected by the user, weekly or monthly. Otherwise data remains with daily resolution.

**Creation of a unique dataframe**

**Scope:** For each variable merge observations and gridded data together.

**Statistics**

**Scope:** Compute the statistics for each station and dataset



**Output**

**Scope:** Generate an interactive dashboard and a report(pdf) summarizing the results.

Dashboard

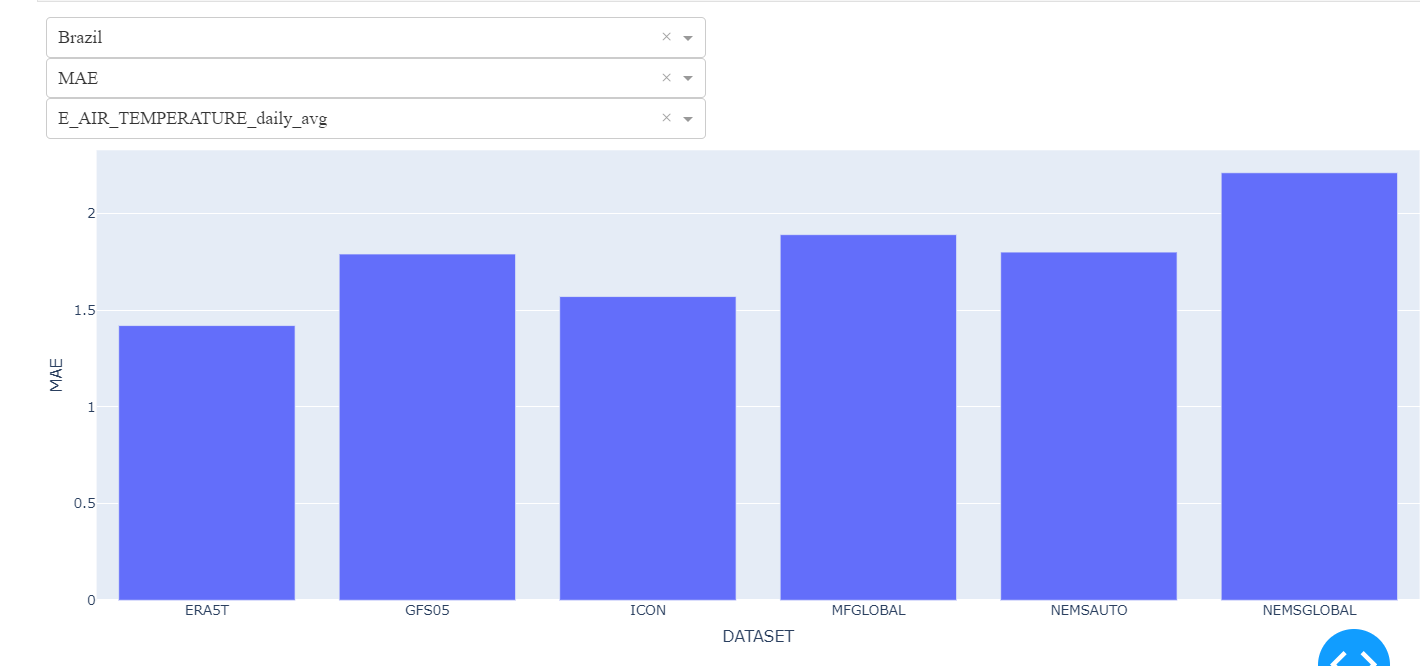
****

Figure : Output 1

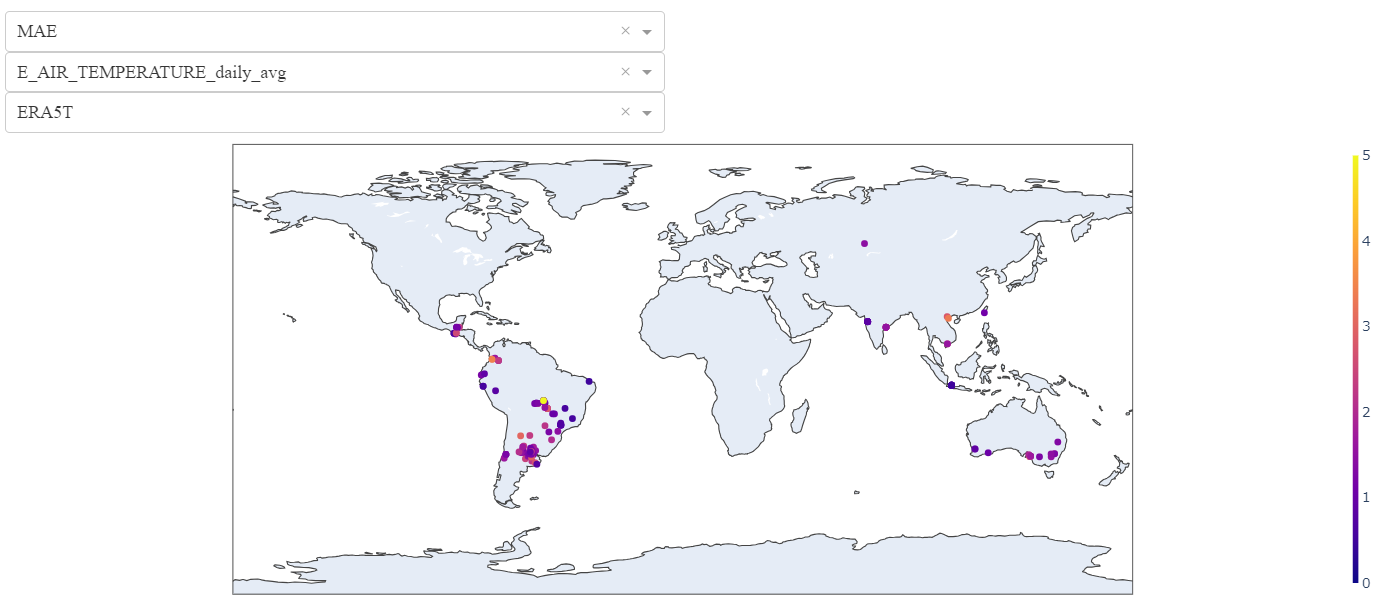
****

Figure : Output 2

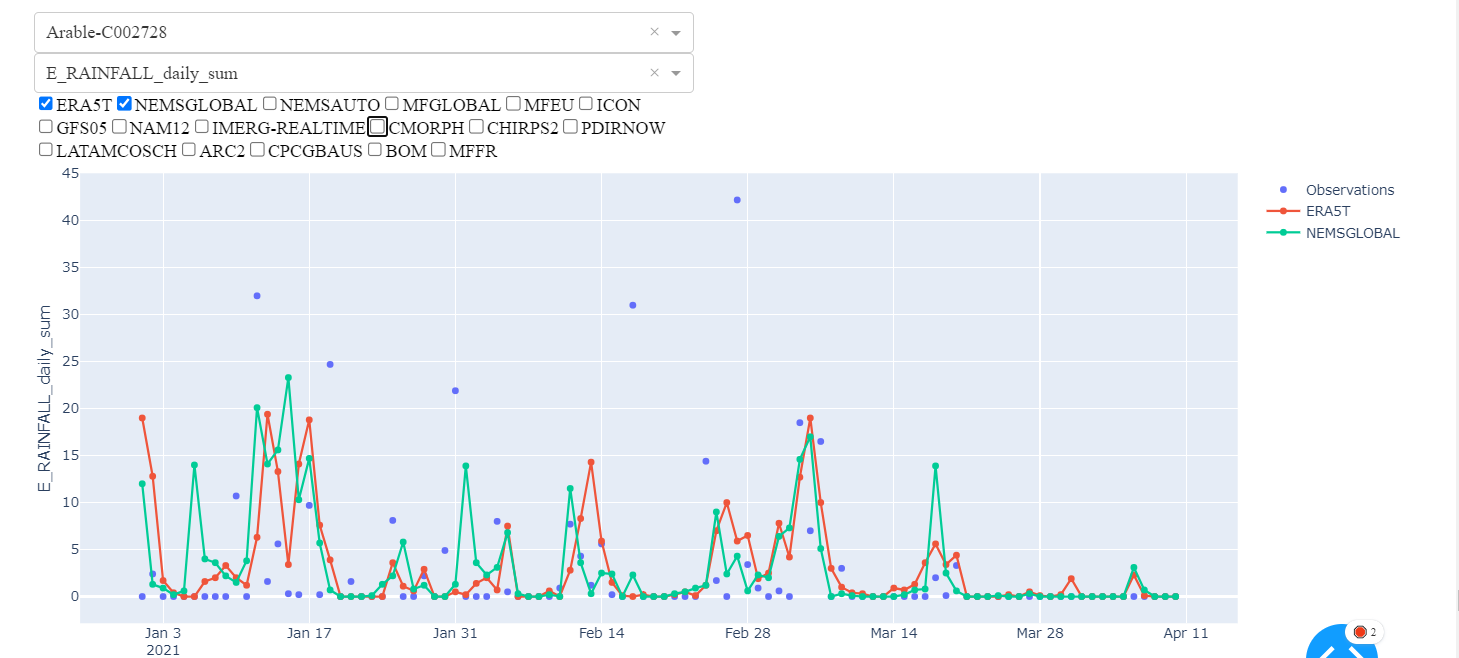
****

Figure : Output 3

*The plots can be downloaded as image or shared via email or get the raw data in excel (plot + inputs)*

Report

For each variable, we create a report containing:

* Inputs
* WS related info (WS info 1, WS info 3)
* Output (Output 1, Output 2)

**Sequence Diagram**

