## Spatio-Temporal Upscaling of Flux Tower Gross Primary Productivity Measurements

For details of the methodology see Spinosa, A.; Fuentes-Monjaraz, M.A.; El Serafy, G. Assessing the Use of Sentinel-2 Data for Spatio-Temporal Upscaling of Flux Tower Gross Primary Productivity Measurements. Remote Sens. **2023**, 15, 562. https://doi.org/10.3390/rs15030562. Please acknowledge the GPP estimations by citing the above article. Thanks!

### How to use sen2gpp

The workflow is accessible through this GitHub repository and through the VLab environment (<https://vlab.geodab.org/workflows>). The description of inputs and outputs, and steps to processes to run the workflow in VLab environment or locally are explained below.

#### Sen2gpp Input

Sen2gpp workflow requires three main inputs: Net Ecosystem Exchange data, Environmental data, and Experiment setup.

1. Net Ecosystem Exchange data: File in \*.txt or \*.csv format including information of Net Ecosystem Exchange (NEE) and ancillary data derived through an Eddy Covariance System. The format of this file must follow the European Fluxes Database Cluster standards (<http://www.europe-fluxdata.eu/>) and include the next variables.

|  |  |  |
| --- | --- | --- |
| Variable | Europe flux data format | |
| Label | Units |
| CO2 flux or Net Ecosystem Exchange  Air temperature  Relative humidity  Vapour pressure deficit  Friction velocity  Monin-Obukhov length  Wind speed  Maximum wind speed  Wind direction  Global radiation  Standard deviation of the wind component along the v anemometer axis | FC or NEE  TA  RH  VPD  USTAR  MO\_LENGTH  WS  WS\_MAX  WD  SW\_IN  V\_SIGMA | µmol m-2 s-1  °C  %  Pa  m s-1  m  m s-1  m s-1  ° (degrees)  W m-2  m s-1 |

TA, VPD, SW\_IN environmental variables are used for the correction of the data and for the computation of the empirical model. Extra environmental data for the computation of the empirical model can be provided through the Environmental data file.

1. Environmental data file: File in \*.txt or \*.csv format including extra information on meteorological conditions matching the NEE observations such as maximum, minimum temperature, or rainfall. In case rainfall is provided, the data set must include observations one year before the NEE measurements as rainfall data is not used directly but preprocessed to derive rolling averages with delays up to 180-day lags. The data must include daily estimations of the environmental variables with time format YYYY-MM-DD (e.g., 2016-09-01) and column names should not include spaces or special characters (e.g., TA\_MAX).
2. Experiment setup: File in \*.cfg format including information of the workflow parameters, and experiment decisions. An overview of the different parameters is described in Annex A. Some of the configuration file parameters must remain fixed to work correctly in the VLabs environment. The fixed variables can be modified only in case local computation is performed through an Anaconda environment for software development purposes. Annex A includes an extra column describing if the variable is fixed. Variables that can be modified as classified as “site-specific”. The file must be named as “Configuration\_file.cfg” to be read by the workflow. In case of local computation, this file must be saved in the sen2gpp/Configs folder.

In a first section of the configuration file, different components of the workflow can be turned on (when set as TRUE) or turned off (when set as FALSE). The workflow is in series so to compute the last element of the workflow previous steps in the workflow must be computed. Some elements of the workflow are for extra analysis or can be replaced by others inside the workflow and can be turned off.

In a second section, the parameters of the models are defined for the different methodologies included in the workflow. Numerical values and extra experiment decisions are defined in the second part of the configuration file.

Multiple examples of the Net Ecosystem Exchange data, Environmental data, and Experiment set-up files (“Configuration\_file.cfg”) can be found in the GitHub repository within the sen2gpp/Input/NEE sen2gpp/Input/EV and sen2gpp/Configs folders, respectively.

#### Sen2gpp Output

The outputs of the workflow are stored in the sen2gpp/Output folder if the folder structure of the GitHub repository is kept. In the VLab environment sen2gpp/Output folder is zipped and saved as Results.zip which can be downloaded from VLab once the sen2gpp workflow has been run and finalized. The content of the sen2gpp/Output folder is described here.

1. Footprint\_output: Folder containing the climatological footprint results in \*csv files. The files describe the coordinates (lot, lat) of the contour lines for the years of analysis. An extra table is created with the wind direction and velocity information used for the computation of the climatological footprint. This table named as “ID\_online\_data\_year.csv” can be loaded into the Flux Footprint Prediction (FFP) online data processing tool (<https://geography.swansea.ac.uk/nkljun/ffp/www/>)
2. GPP\_output: Results of the flux partitioning computation aggregated into daily values saved as \*csv file.
3. Maps\_output: The folder includes the monthly and 15-day GPP maps for the period of analysis in \*tiff and \*.png format. For visualization of results, interactive maps in \*.html format is also provided. The list of maps is provided in \*.xlsx format with information of the observed and calculated GPP. The upscaling area according to the unsupervised classification is displayed in the \*\_Classification\_map.html map.
4. Model\_output: This file includes multiple tables in \*.xlsx and plots in \*.png with results of the interdependency analysis as described by Spinosa et al. (2023).
5. NEE\_output: Results of processed 30-minute Net Ecosystem Exchange data in \*csv file.
6. VI\_output: Daily values of multiple Vegetation Indexes per day for the full period of analysis, and average values of the Vegetation Indexes per year and region in the climatological footprint.

An example of the sen2gpp/Output directory with example workflow outputs can be found in the GitHub repository.

*Processes to run the workflow in VLab environment.*

1. VLab environment can be accessible through VLab website (https://vlab.geodab.org/) using a google account.

Graphical user interface, text, application

Description automatically generated

1. Once having access into the environment in the left panel under “Workflows” tab Text

   Description automatically generated with medium confidence multiple loaded workflows are displayed. Sen2gpp workflow can be searched using the

“Filter by text” field .

Graphical user interface, text, application, email

Description automatically generated

1. Create a new experiment by selecting “Create New Experiment” tab  in the sen2gpp field. A new window will appear where an experiment name, experiment description and input files in “raw” format can be provided.

The following files can be used for testing purposes. This data and experiment set-up will derive the results published by Spinosa et al. (2023).

NEE: <https://raw.githubusercontent.com/e-shapePP/vlabs/main/data/NEE_data_example.txt>

EV: <https://raw.githubusercontent.com/e-shapePP/vlabs/main/data/Meteo_data_example.txt>

Configuration file: <https://raw.githubusercontent.com/e-shapePP/vlabs/main/data/Configuration_file.cfg>

The files can be stored in other cloud storage containers (e.g., Google Drive, OneDrive, GitHub) to derive the URLs of the files in “raw” format.

Graphical user interface, application

Description automatically generated

1. Once the fields are filled with the required data, we can run the model by pressing next and submit . The status of the workflow will be displayed in My Experiment tab Text

   Description automatically generated. In case the experiment is not displayed re-login into the platform. The workflow will show “Execution” status.

Graphical user interface, text, application, email

Description automatically generated

1. The workflow depending on the amount of data can take ~1-4 hours to run. After completion, the status will change to “Succeded”, and “Results.zip” file will be available for direct download.

Graphical user interface, text

Description automatically generated

*Processes to run the workflow locally.*

sen2gpp is require Python programming language (version 3.6.13) and multiple libraries as described in the next table.

Table 3. 2 Python libraries with version and general description used in the workflow code.

|  |  |  |
| --- | --- | --- |
| Python library | Version | Description |
| time | \* | Count computing time |
| sys | \* | Manipulate files in console |
| configparser | \* | Integrate external configuration file |
| os | \* | Manipulate directories |
| json | \* | Manipulation of json type files |
| datetime | \* | Create date type objects for time series |
| math | \* | Mathematical operation tool |
| numbers | \* | Evaluate if a variable is a numeric object |
| numpy | 1.19.5 | Create and manipulate number objects |
| hesseflux | 3.2 | Correction of eddy covariance data and GPP estimation |
| earthengine-api | 0.1.288 | Manipulation of Google Earth Engine catalogues and algorithms |
| pyproj | 3.0.1 | Reprojection of coordinates of the climatological footprint |
| scipy | 1.5.4 | Mathematical algorithms used in the FFP model |
| pandas | 1.1.5 | Manipulation of data frames |
| statsmodels | 0.12.2 | Compute regression models |
| scikit-learn | 0.24.2 | Divide data in calibration and validation data sets |
| matplotlib | 3.3.4 | Plot regression models |
| altair | 4.1.0 | Plot time series |
| folium | 0.12.1 | Create dynamic maps in html format |

\*Integrated in python 3.16.13

The packages can be installed through conda using the environment described in the sen2gpp/Environment/e-shape\_environment\_linux.yaml.

The workflow can be implemented by saving the input data in the sen2gpp/Input/EV and sen2gpp/Input/NEE, respectively for the environmental data and Net Ecosystem Exchange data files. The Experiment set-up data file should be saved as sen2gpp/Configs/Configuration\_file.cfg. The workflow then can be computed by running the sen2gpp.py/sen2gpp.py code. The outputs will be saved in the sen2gpp/Output different folders.

### Annexe A

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Descriptions** | **VALUE** | **Type of parameter** |
| **ID** | Experiment ID | --- | SS |
| **inputdir** | Input data directory | Input | FX |
| **outputdir** | Output data directory | Output | FX |
| **meteo\_file** | Meteorological data file name | --- | SS |
| **outlier** | Computation of outlier detection | TRUE | FX |
| **ustar** | Computation of velocity friction filter (automatic threshold)  Computation of friction filter (filter values below a threshold as described by the “ustarmin” variable).  Only one of them can be set as TRUE. | TRUE or FALSE | SS |
| **ustar\_non\_annual** | TRUE or FALSE | SS |
| **partition** | Computation of flux partitioning | TRUE | FX |
| **fill** | Computation of gap filling | TRUE | FX |
| **fluxerr** | Estimation of errors in processed data | TRUE | FX |
| **daily\_gpp** | Computation of daily GPP from 30-minute gap filled estimations | TRUE | FX |
| **climatological\_footprint** | Computation of climatological footprint | TRUE | FX |
| **calculated\_ffp** | Use previous calculated climatological footprint. | FALSE | FX |
| **vegetation\_indices** | Calculation of satellite-derived Vegetation Indexes | TRUE | FX |
| **environmental\_variables\_station** | Processing of EV from environmental data file | TRUE | FX |
| **environmental\_variables\_satellite** | Processing of EV from global data | FALSE | FX |
| **tower\_observations** | Processing of EV from net ecosystem exchange data file | TRUE | FX |
| **df\_rainfall\_station\_switch** | Processing rainfall data from environmental data file | TRUE or FALSE | SS |
| **df\_meteo\_station\_switch** | Processing of additional meteorological data from environmental data file | TRUE or FALSE | SS |
| **df\_rainfall\_CHIRPS\_switch** | Processing of rainfall data from CHIRPS dataset | FALSE | FX |
| **df\_temp\_MODIS\_switch** | Processing of temperature data from MODIS | FALSE | FX |
| **df\_meteo\_tower\_switch** | Processing of meteorological data from net ecosystem exchange data file | TRUE | FX |
| **correlation\_analysis** | Computation of correlation analysis | TRUE | FX |
| **correlation\_analysis\_simple** | Computation of correlation analysis (simple) | TRUE | FX |
| **rei\_gpp\_switch** | Flux partitioning method to implement (Reinchstein, Falge or Lasslop) according to hesseflux tool (https://github.com/mcuntz/hesseflux)  Only one of them can be set as TRUE. | TRUE or FALSE | SS |
| **fal\_gpp\_switch** | TRUE or FALSE | SS |
| **las\_gpp\_switch** | TRUE or FALSE | SS |
| **calibration\_validation** | Computation of calibration and validation | TRUE | FX |
| **MODIS\_analysis** | Comparison of results with MODIS GPP product | FALSE | FX |
| **timeseries\_thirty** | Transformation of daily time series into monthly time series | TRUE | FX |
| **timeseries\_fifteen** | Transformation of daily time series into 15-day time series | TRUE | FX |
| **mapping\_GPP** | Computation of maps | TRUE | FX |
| **classification\_maps** | Classification of area and selection of upscaling area | TRUE | FX |
| **maps\_from\_features** | Use previous defined map as upscaling area | FALSE | FX |
| **mapping\_GPP\_thirty** | Computation of monthly maps | TRUE | FX |
| **mapping\_GPP\_fifteen** | Computation of 15-day maps | TRUE | FX |
| **export\_maps\_to\_drive** | Saving maps into a Google Drive storage | FALSE | FX |
| **eufluxfile** | Name of the NEE input files | --- | SS |
| **timeformat** | Time format in the NEE input files | --- | SS |
| **sep** | Separator in the NEE input files | --- | SS |
| **skiprows** | Rows to skip in the NEE input files | --- | SS |
| **undef** | Undefined values in the NEE input files | --- | SS |
| **swthr** | Global radiation threshold to identify day-night hours | --- | SS |
| **outputfile** | File to save the NEE output file | \NEE\_output\ | FX |
| **outputname** | Name of the processed NEE output file | --- | SS |
| **outundef** | Setting null values for flagged data in processed NEE files | FALSE | FX |
| **outflagcols** | Integrate quality flags in processed NEE files | FALSE | FX |
| **carbonflux** | Column name with the NEE data | --- | SS |
| **remove\_SW\_IN** | Remove global radiation data for corrections and model formulation | TRUE or FALSE | SS |
| **nscan** | Data correction parameters according to hesseflux tool (https://github.com/mcuntz/hesseflux) | --- | SS |
| **nfill** | --- | SS |
| **z** | --- | SS |
| **deriv** | --- | SS |
| **ustarmin** | --- | SS |
| **nboot** | --- | SS |
| **plateaucrit** | --- | SS |
| **seasonout** | --- | SS |
| **applyustarflag** | --- | SS |
| **sw\_dev** | --- | SS |
| **ta\_dev** | --- | SS |
| **vpd\_dev** | --- | SS |
| **longgap** | --- | SS |
| **nogppnight** | --- | SS |
| **carbonfluxlimit** | Maximum value of carbon fluxes per 30-minutes | --- | SS |
| **respirationlimit** | Maximum value of ecosystem's respiration per 30 minutes | --- | SS |
| **rolling\_window\_gpp** | Time window for rolling average in GPP daily data to reduce noise in the time series | --- | SS |
| **rolling\_center\_gpp** | Indicate if the rolling average must be centered | --- | SS |
| **rolling\_min\_periods** | Indicate the minimum number of values for the rolling average | --- | SS |
| **altitude** | Climatological footprint parameters according to the Flux Footprint Prediction (FFP) tool (https://geography.swansea.ac.uk/nkljun/ffp/www/) | --- | SS |
| **latitude** | --- | SS |
| **longitude** | --- | SS |
| **canopy\_height** | --- | SS |
| **displacement\_height** | --- | SS |
| **roughness\_lenght** | --- | SS |
| **instrument\_height\_anenometer** | --- | SS |
| **instrument\_height\_gas\_analyzer** | --- | SS |
| **projection\_site\_UTM\_zone** | --- | SS |
| **boundary\_layer\_height** | --- | SS |
| **domaint\_var** | --- |  |
| **nxt\_var** | --- | SS |
| **rst\_var** | --- | SS |
| **bands** | Vegetation Index to analize | --- | SS |
| **max\_cloud\_coverage** | Maximum cloud coverage in satellite images | --- | SS |
| **crs** | Projection in EPSG code | --- | SS |
| **ndviMask** | Minimum NDVI value for VIs time series calculation | --- | SS |
| **mndviMask** | Minimum MNDVI value for VIs time series calculation | --- | SS |
| **rolling\_window\_ev\_meteo** | Time window for rolling average in EV daily data from the meteorological station to reduce noise in the time series | --- | SS |
| **rolling\_window\_ev\_meteo\_sat** | Time window for rolling average in EV daily data satellite data to reduce noise in the time series | --- | SS |
| **rolling\_window\_gpp\_MODIS** | Time window for rolling average in GPP daily data from MOD17 products to reduce noise in the time series | --- | SS |
| **precipitation\_data** | Column names of precipitation data (start with daily precipitation data) | --- | SS |
| **scale\_satellite\_data** | Scale to retrieve EV data from google earth engine | --- | SS |
| **model\_name** | indicate the ranked number of the model to analyze | --- | SS |
| **feature\_collection** | Name of GEE feature collection | --- | SS |
| **ecosystem\_extension** | Extension of exosystem in meters | --- | SS |
| **number\_clusters** | Number of clusters in the classification of the region | --- | SS |
| **training\_scale** | Scale of the pixels used in the classification algorithm | --- | SS |
| **training\_dataset** | Number of data points or pixels for the classification algorithm | --- | SS |
| **scale\_getRegion** | Scale to retrieve results from classification algorithm | --- | SS |
| **vector\_scale** | Scale to transform results from classification algorithm to vectors | --- | SS |

\*FX. - Fixed values in the analysis to worked in VLabs environment.

\*SS.- Site specific parameters that can be modified according to the site characteristics and experiment decisions.