ELLIS – PhD position applicant 2022

Work sample as support material

Artur Lourenço

2022-02-8

1 Introduction

This work sample aims to demonstrate the set of skills and prior knowledge that allow the development of a doctoral study under the supervision of researchers who are part of the Department Biogeochemical Integration at the Max Planck Institute for Biogeochemistry (MPI-BG\BGI).

The main point is to show that I have a certain understanding of the type of environmental data used in MPI-BG\BGI. To this end, real data were used (with permission) from an experimental site managed by the institute.

2 Data

Two datasets of real data were used from the experimental site ‘Majadas del Tietar’ in 2018 and 2019. The variables present in the datasets are:

‘daily’ dataset:

date: date of observation in daily scale.
Cam_NDVI_G: NDVI for the herbaceous layer (‘grass’).
Cam_NDVI_T: NDVI for the oak trees.

‘HHly’ dataset:

rDate: half-hourly timestamp of the observation.
daytime: a binary flag 0/1 for ‘night’ and ‘day’.
GPP: Gross primary productivity (micromol co2 m-2 s-1).
ET : Evapotranspiration (mm hour-1).

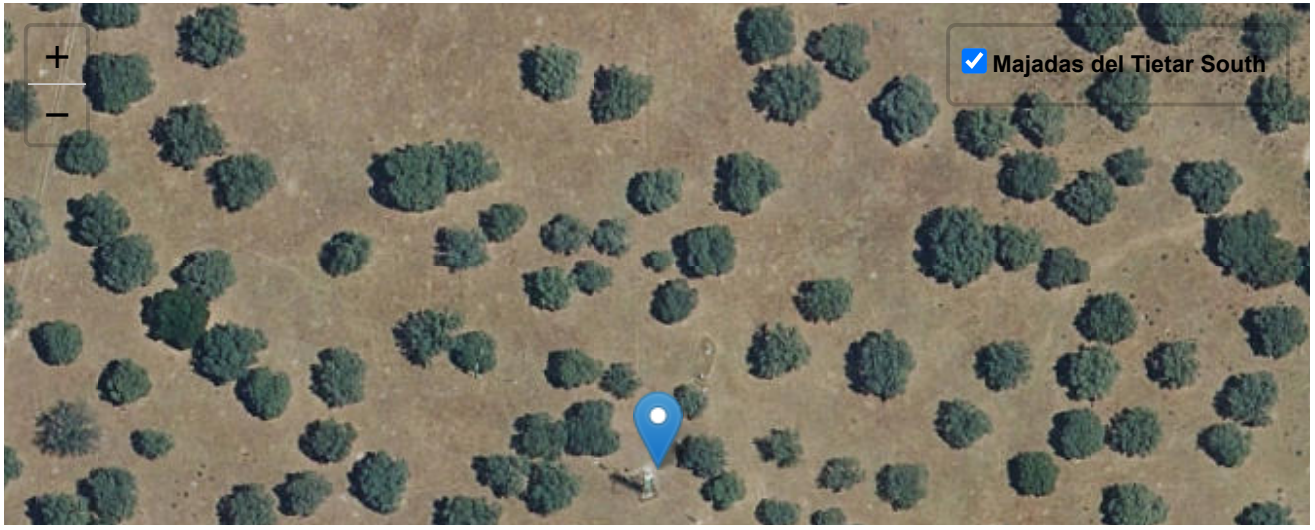
Samples of the datasets:

Samples of the datasets.

rDate	daytime	GPP	ET	date	Cam_NDVI_G	Cam_NDVI_T
2018-01-01 00:00:00	0	0	NA	2015-12-03	0.5727425	0.6011256
2018-01-01 00:30:00	0	0	NA	2015-12-04	0.5727425	0.6018299
2018-01-01 01:00:00	0	0	NA	2015-12-05	0.5829882	0.6083070
2018-01-01 01:30:00	0	0	NA	2015-12-06	0.5839362	0.6089036
2018-01-01 02:00:00	0	0	NA	2015-12-07	0.5856529	0.6140006
2018-01-01 02:30:00	0	0	NA	2015-12-08	0.5592100	0.5863800
2018-01-01 03:00:00	0	0	NA	2015-12-09	0.5545687	0.5832573
2018-01-01 03:30:00	0	0	NA	2015-12-10	0.5639964	0.5852215
2018-01-01 04:00:00	0	0	NA	2015-12-11	0.5680610	0.5839760
2018-01-01 04:30:00	0	0	NA	2015-12-12	0.5698815	0.5856296

3 Study Site

The study site is a mixed tree-grass ‘dehesa’ semi-arid ecosystem in Spain. The area has a low density tree cover. The semiarid climate in this region is characterized by cold winters and long dry summers, with periodic severe droughts [1]. The location of the study site is shown on the interactive map below.



3 Data pre-processing

It was performed the pre-processing and exploration of the datasets to understand its structure and contents.

Number of NA values in the variables

rDate	daytime	GPP	ET	date	Cam_NDVI_G	Cam_NDVI_T
0	2658	0	8635	0	0	0

The NA values of `daytime` and `ET` variables are equivalent to 7.5% and 24.6% of the total, respectively.

Are there any missing dates in the time series?

```
# For the HHly data.frame

length(HHly$rDate) != length(seq(as.POSIXct(first(HHly$rDate)), as.POSIXct(last(HHly$rDate)), by=30*60))
```

```
## [1] FALSE
```

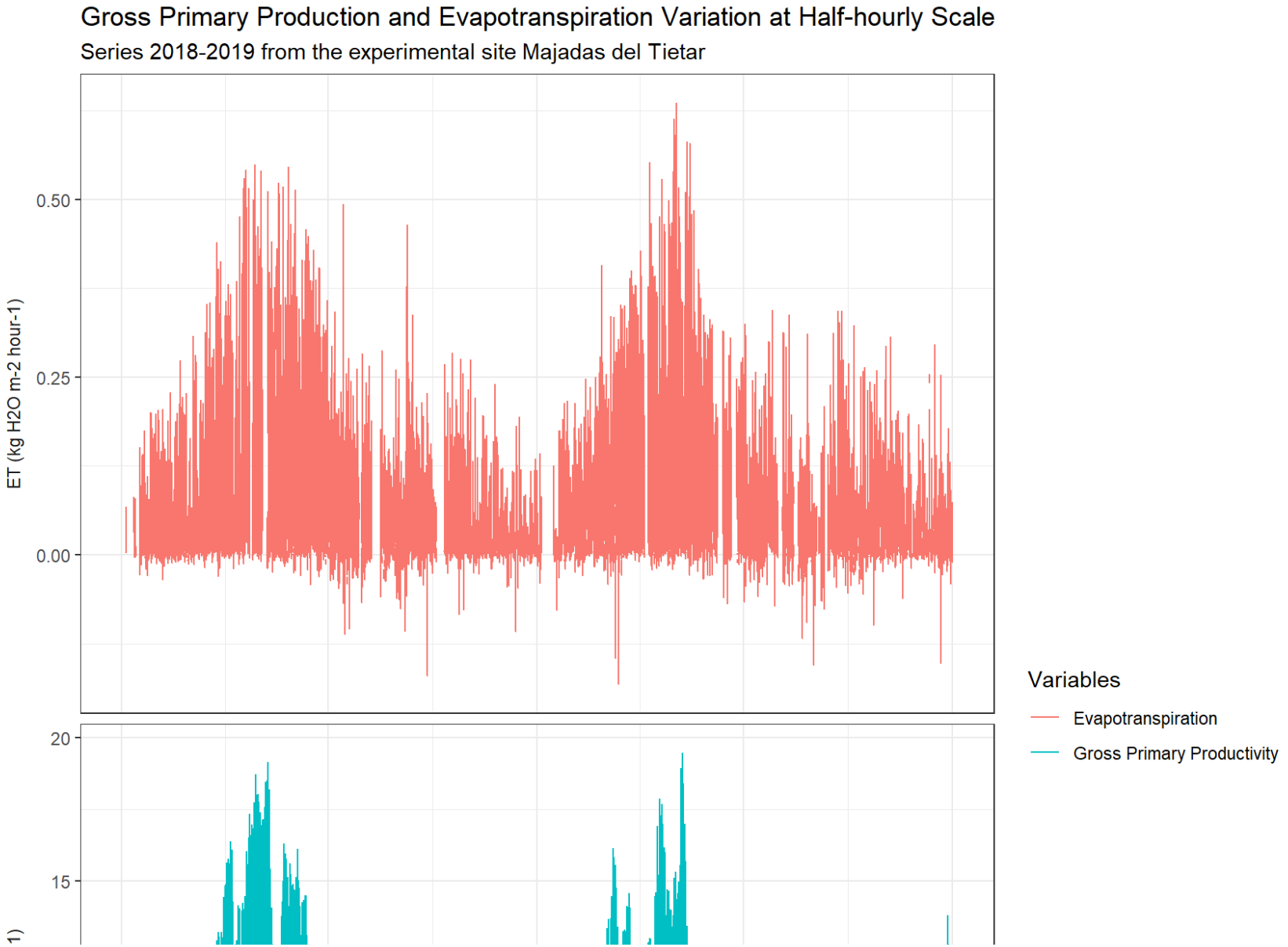
```
# For the daily data.frame

length(daily$date) != length(seq(as.Date(first(daily$date)), as.Date(last(daily$date)), by='day'))
```

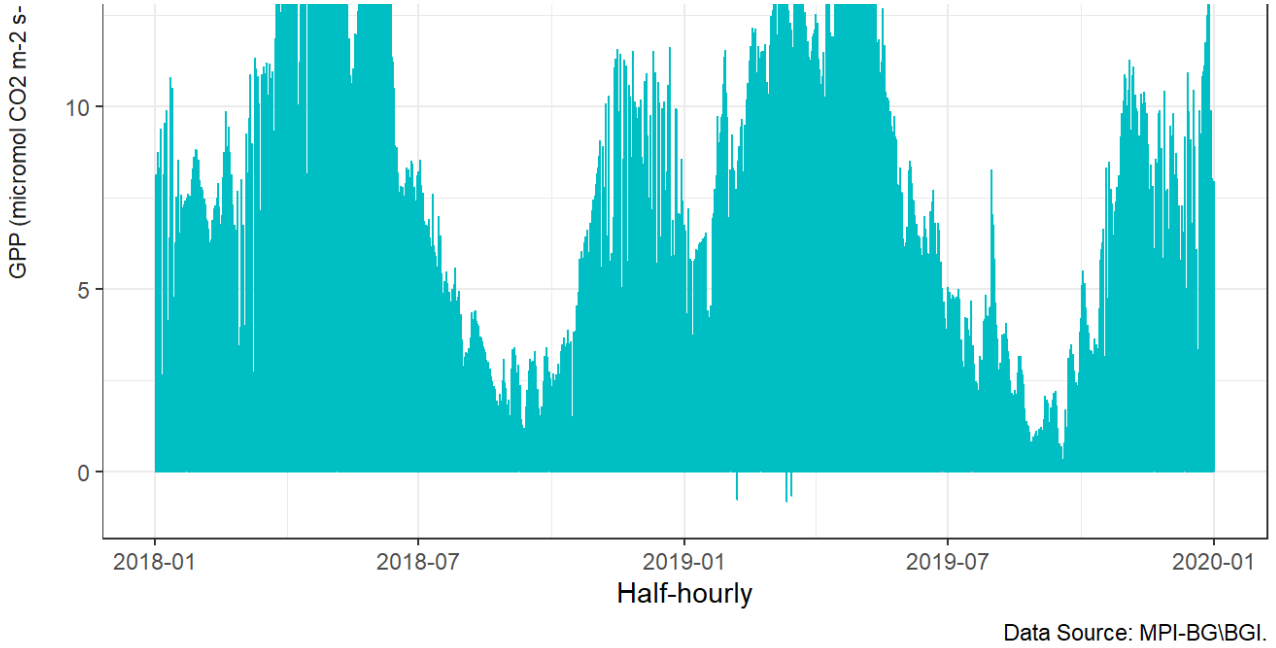
```
## [1] FALSE
```

Note! During the pre-processing of the 'daytime' variable, values that were not consistent with the time of day were found (e.g. 1 for 00:00).

4 Half-hourly GPP and ET Behavior

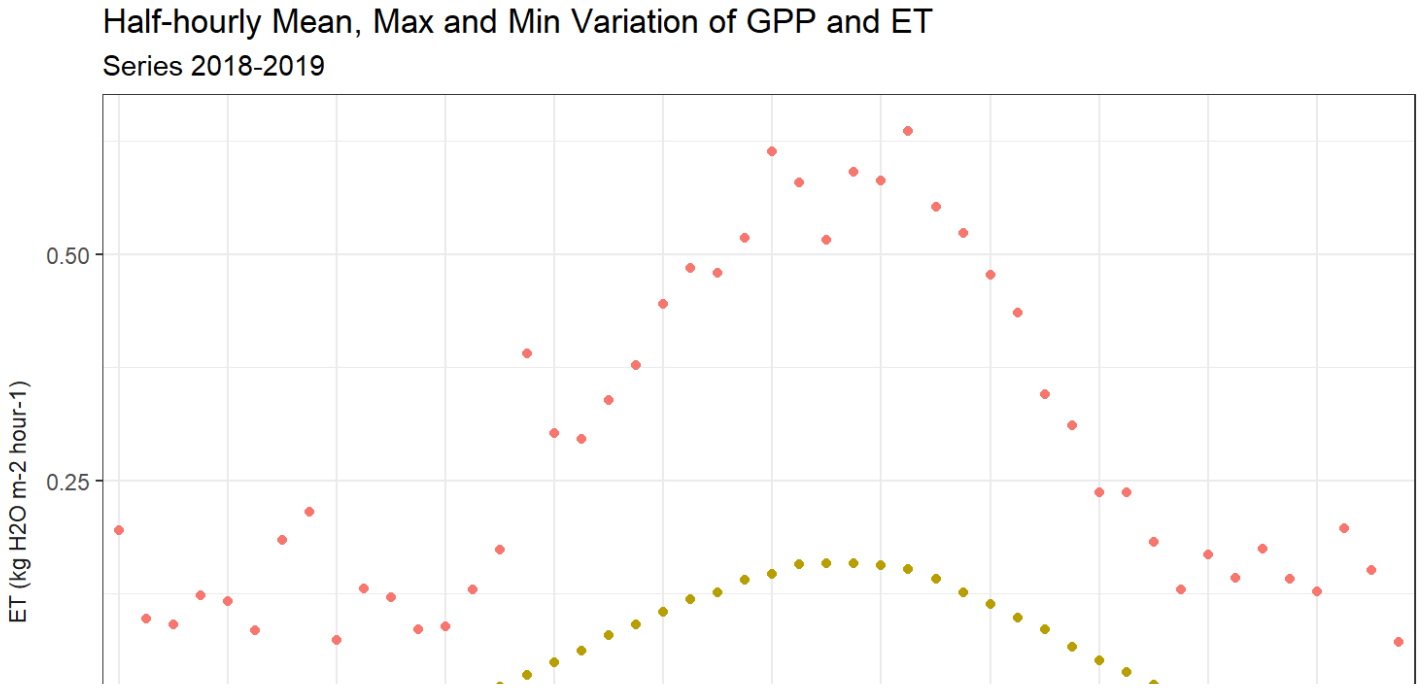
Graphs were produced to visualize the behavior of the GPP and ET variables on the half-hourly scale.

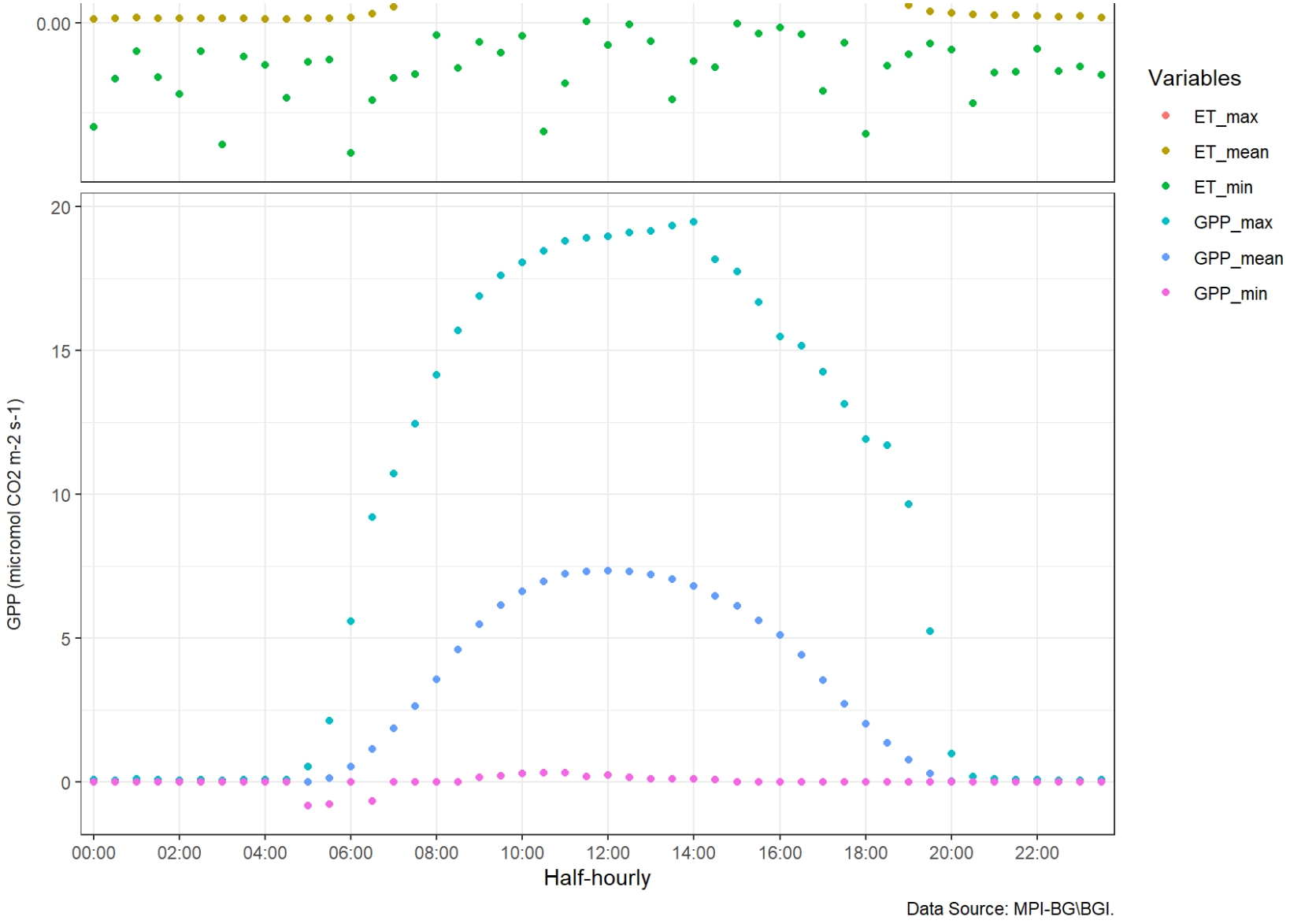




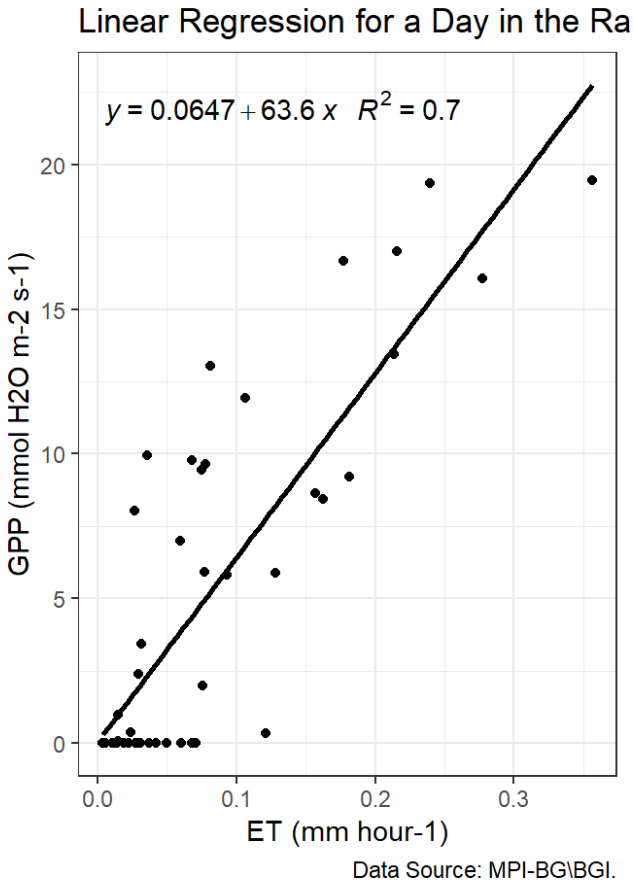
Both ET and GPP were not consistent over the two years of the time series studied, showing high seasonal variation, but not interannual.

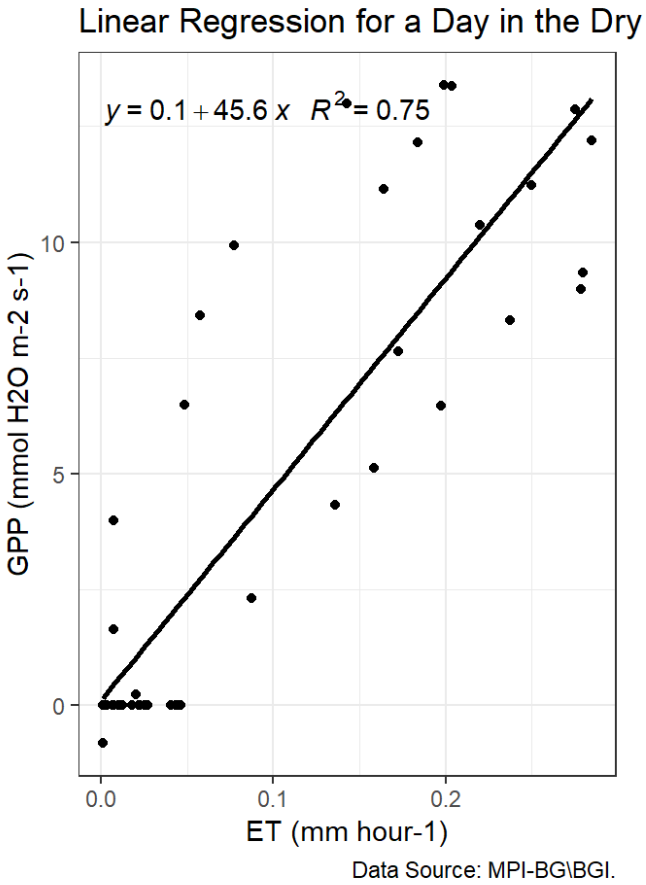
The half-hourly mean, max and min values for GPP and ET variables of the entire serie was calculated.





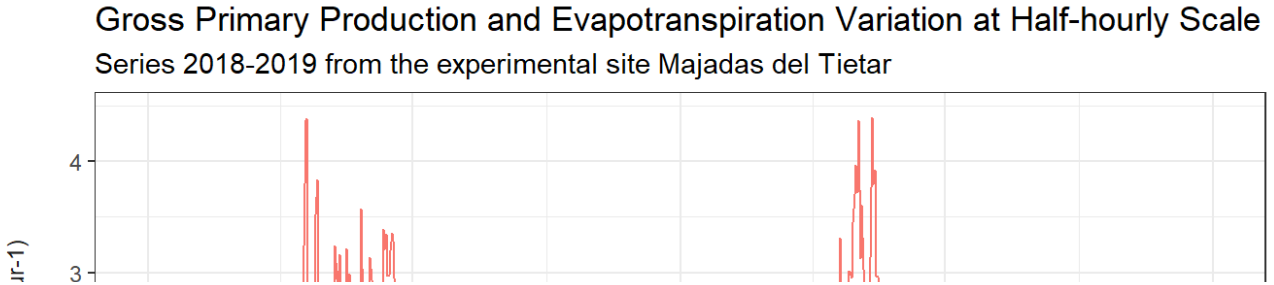
GPP and ET showed similar behaviors but with slightly different peak hours. Considering the calculated means, the highest values of GPP are found at 12:00 as compared to 13:30 for ET. Linear regression showed high correlation between the GPP and ET both on days in the wet and dry season.

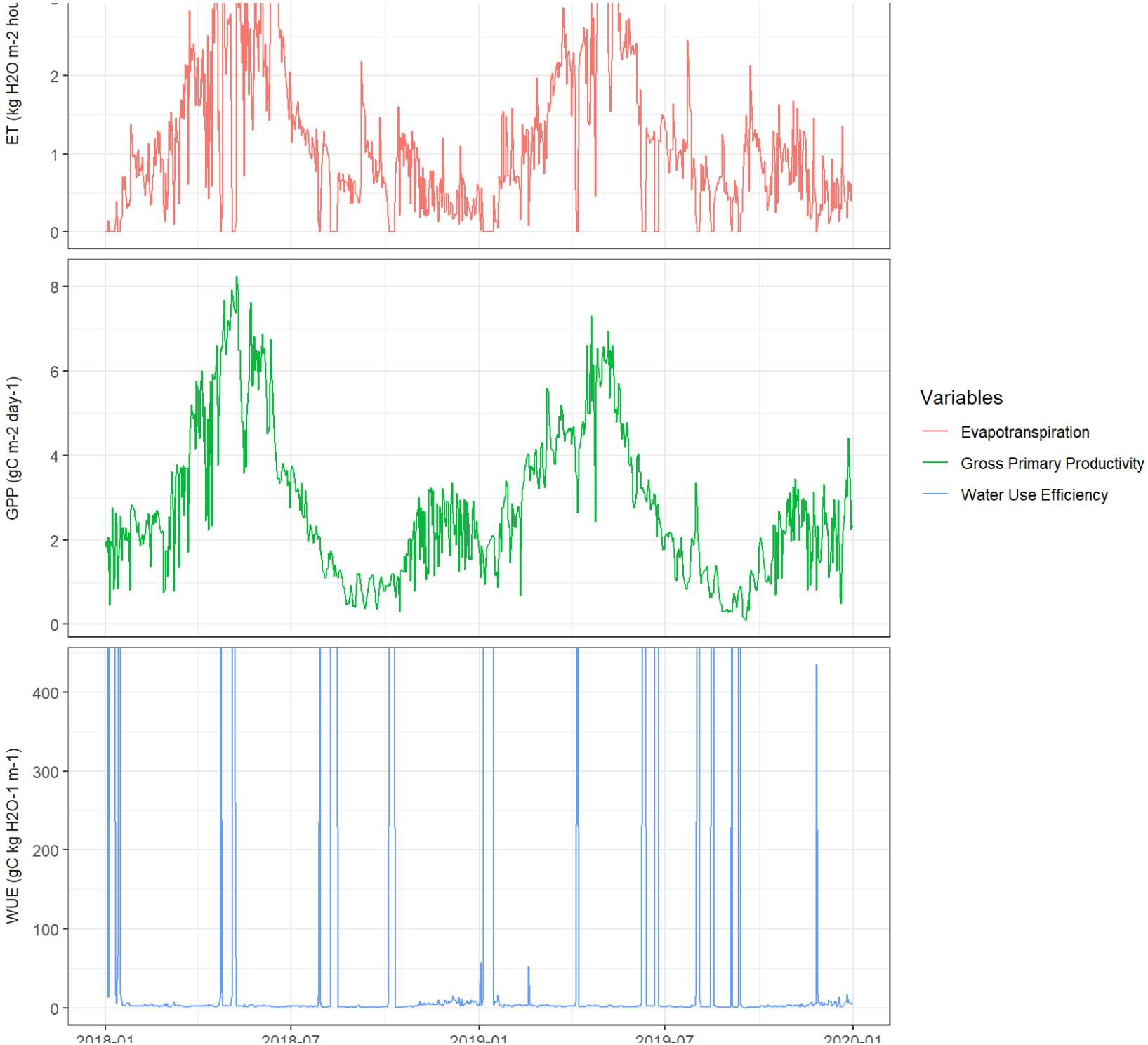


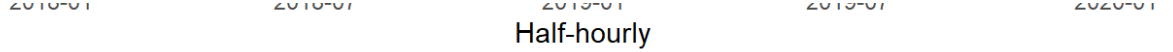


5 Daily GPP and ET Variation and Water-Use Efficiency Calculation

GPP and ET data were converted to daily scale. Functions from the `bigleaf` package were used to perform the conversion to gC m−2 day−1. ET data were also converted to kg H2O m-2 day-1. These operations were performed in order to calculate the Water-Use Efficiency (WUE = GPP/ET) in g C kg H2O−1 time-1.



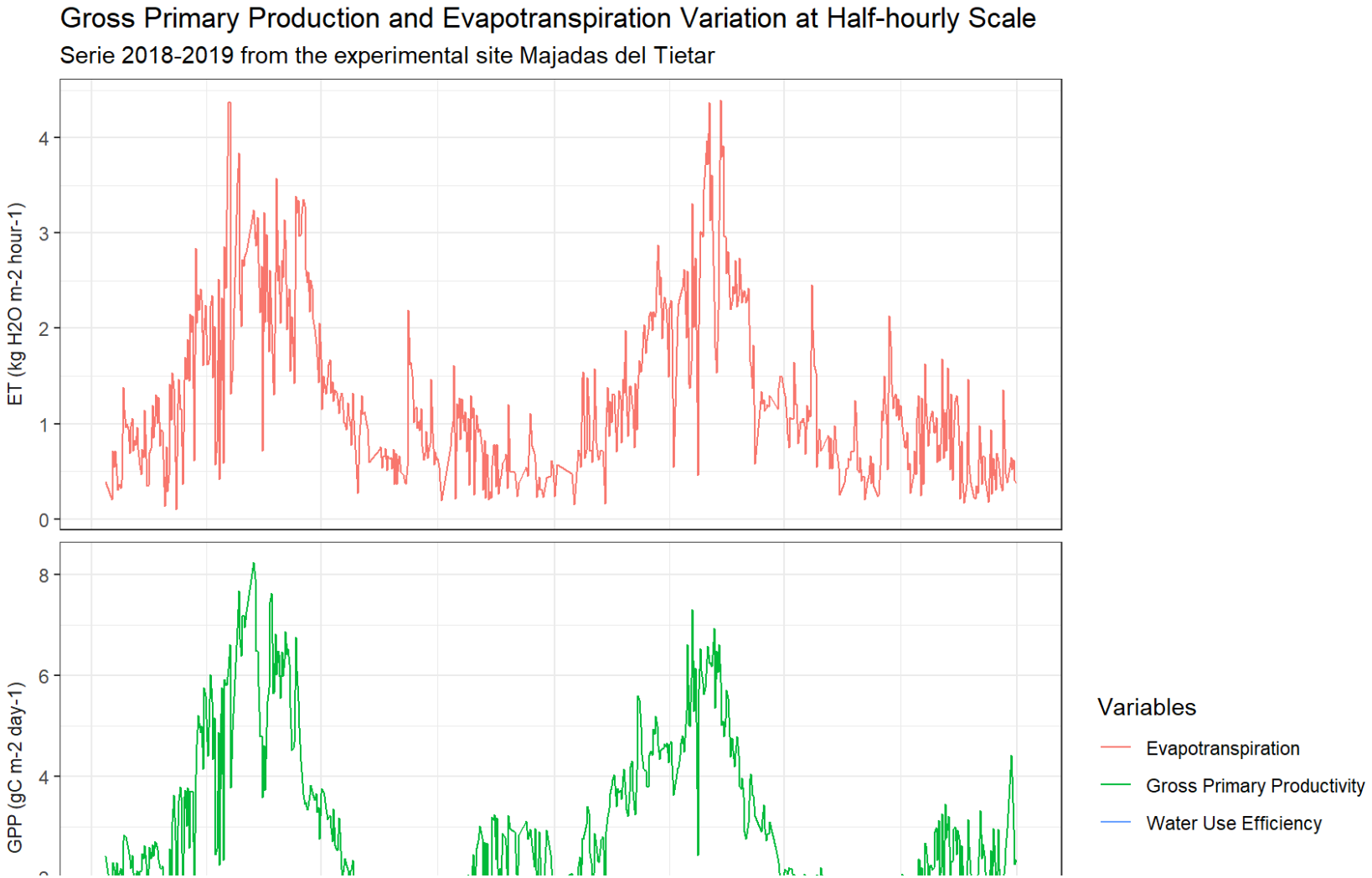

ET (kg H2O m-2 hou
0
1
2
GPP (gC m-2 day-1)
0
2
4
6
8
WUE (gC kg H2O-1 m-1)
0
100
200
300
400
2018-01
2018-07
2019-01
2019-07
2020-01
Variables
Evapotranspiration
Gross Primary Productivity
Water Use Efficiency

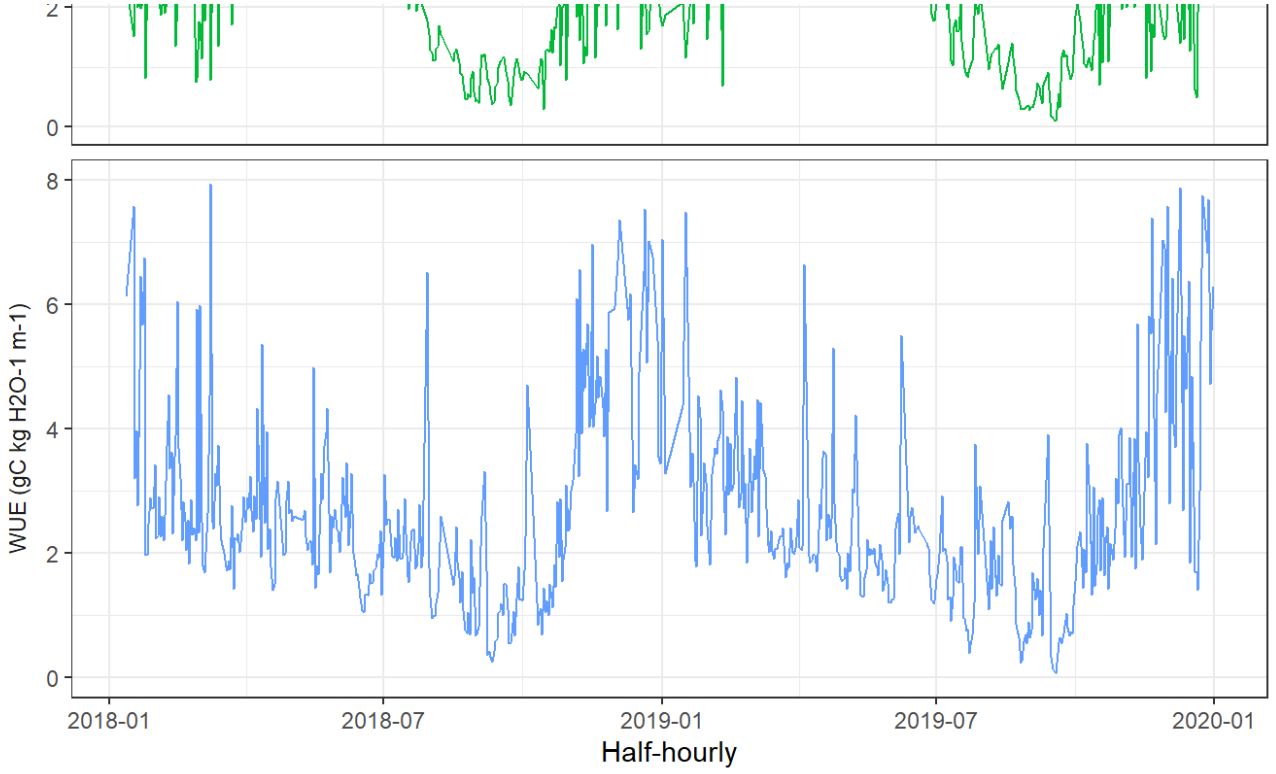


Data Source: MPI-BG\BGI.

Due to the number of flaws present in the ET variable (about 25% of the total observations), the daily values calculated for the WUE present some distortions. Filtering, refining and gap filling techniques must be employed in this situation. A simple removal of data considered as ‘outliers’ was performed considering points below of [Q1-(1.5)IQR] and above of [Q3+(1.5)IQR].

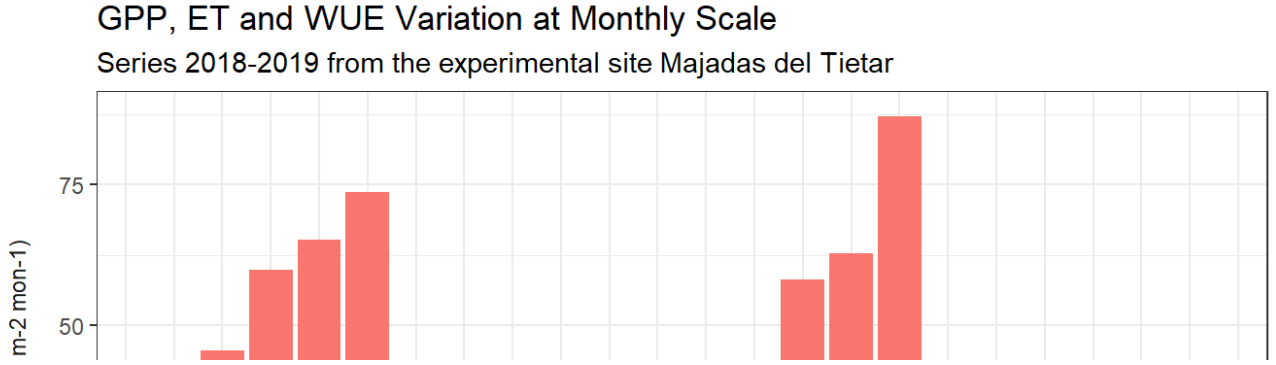
Note! Data considered as ‘ouliers’ were removed only to improve the visualization of the WUE variable in the daily scale. For the other analyses, all data were considered.

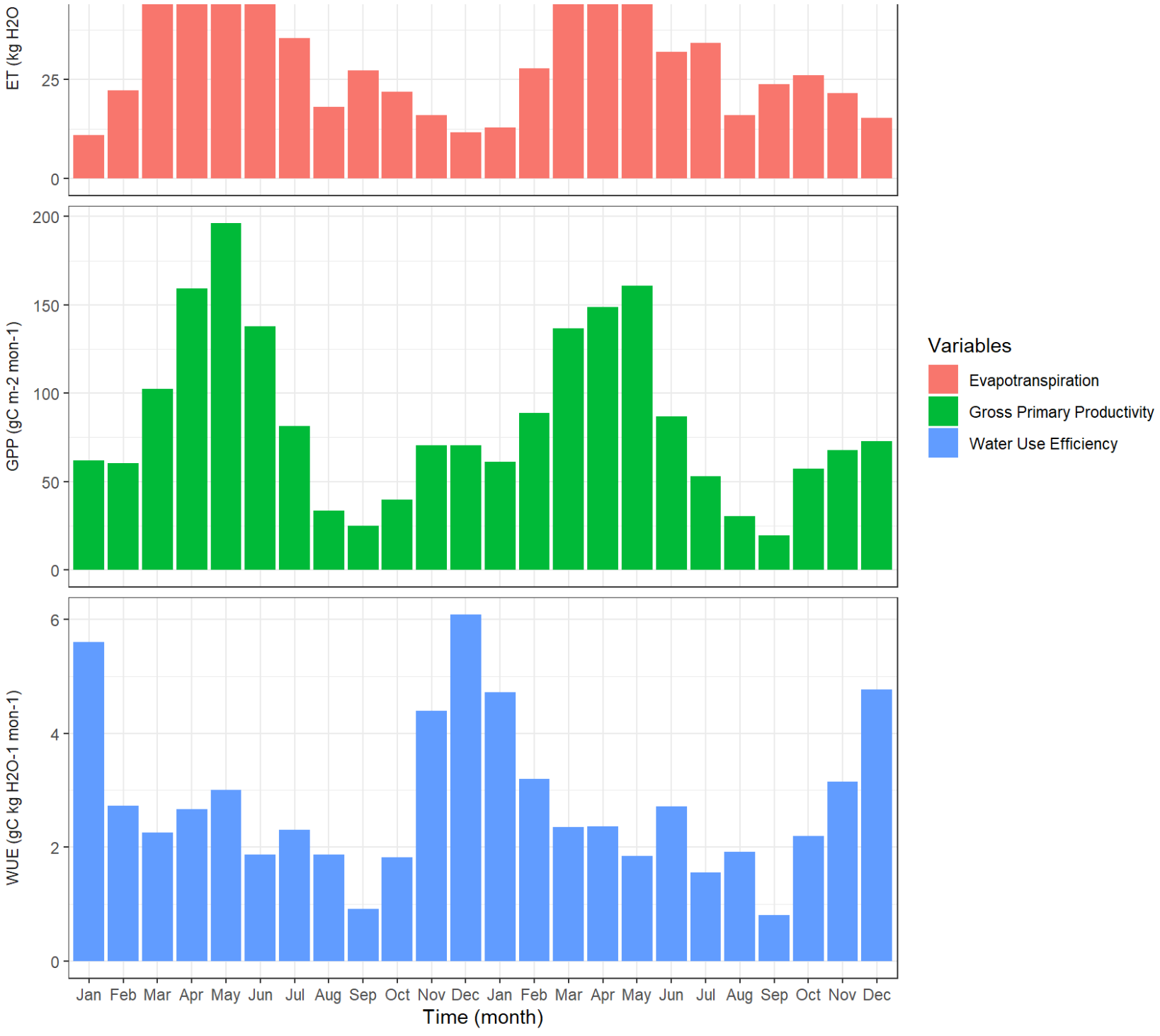




6 Monthly GPP, ET and Water-Use Efficiency Variation

The figure below shows the behavior of the GPP, ET and WUE variables on the monthly scale. It is possible to notice more clearly the seasonal dynamics of the GPP and ET. The WUE presents the highest values in the months where there is a lower influx of carbon into the ecosystem (i.e. low GPP values) and certainly less water avaibility. During this period, vegetation, for example, will use the most efficient water to try to assimilate the same carbon rates.

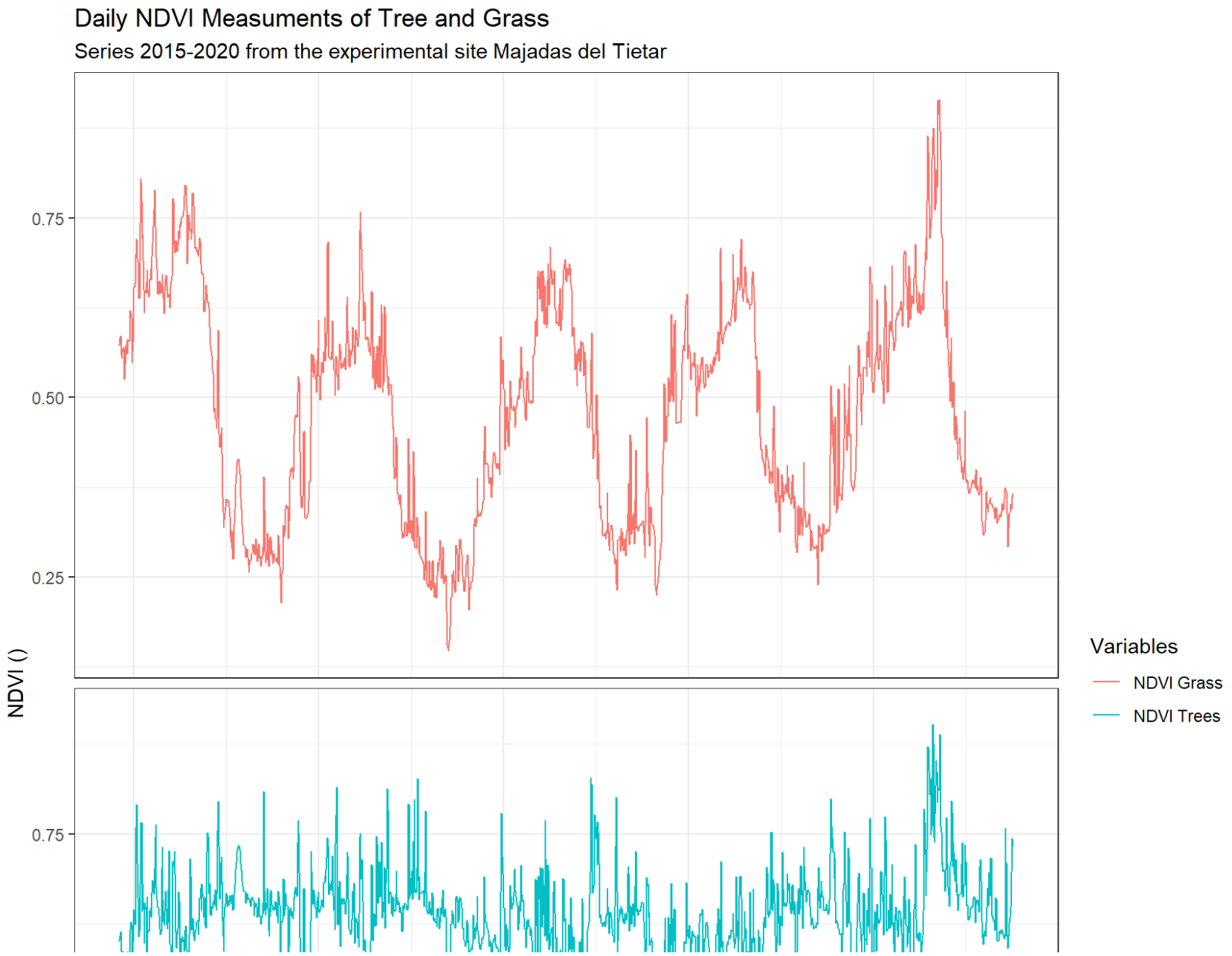


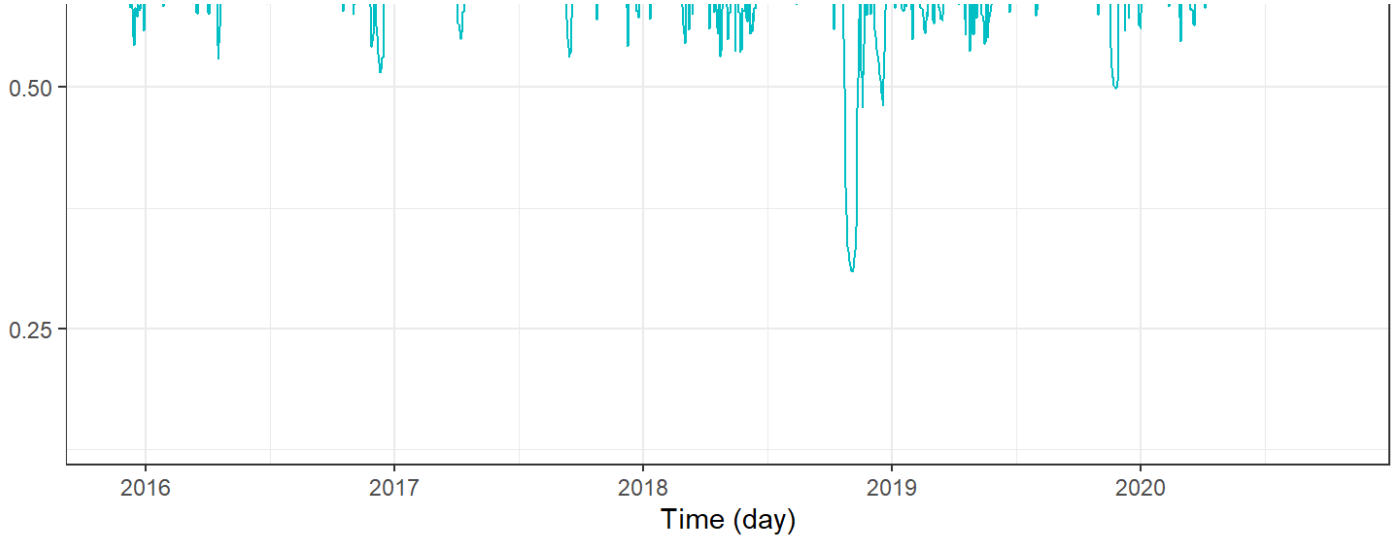


Data Source: MPI-BG\BGI.

7 NDVI links

The figure below shows the behavior of daily NDVI measurements of trees and grass classes on the site.

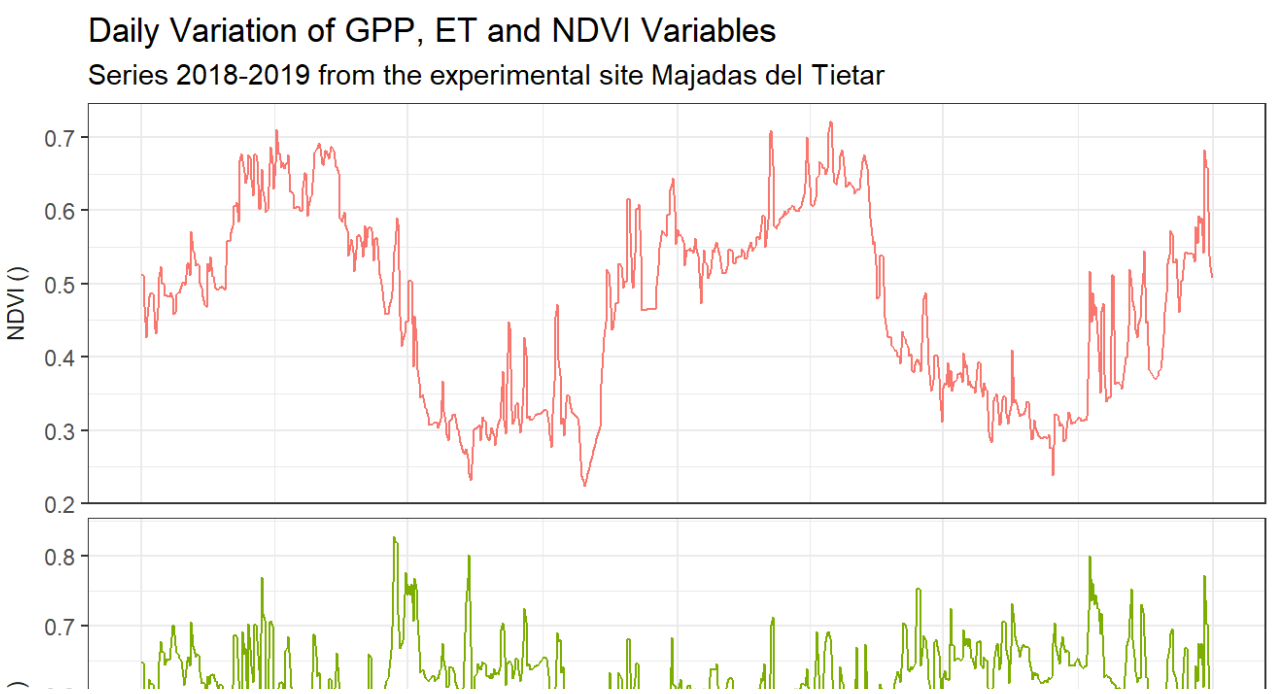


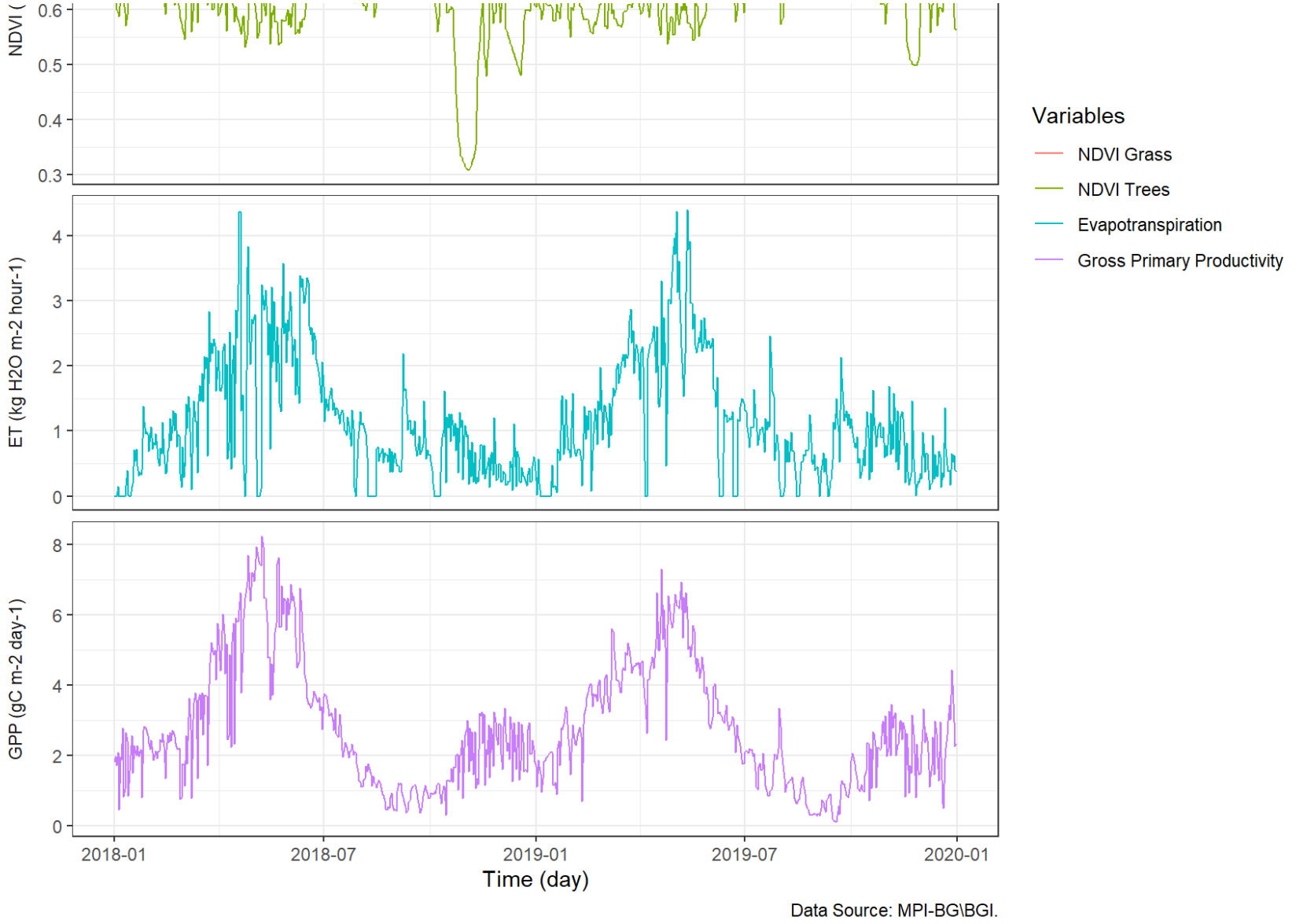


Data Source: MPI-BG\BGI.

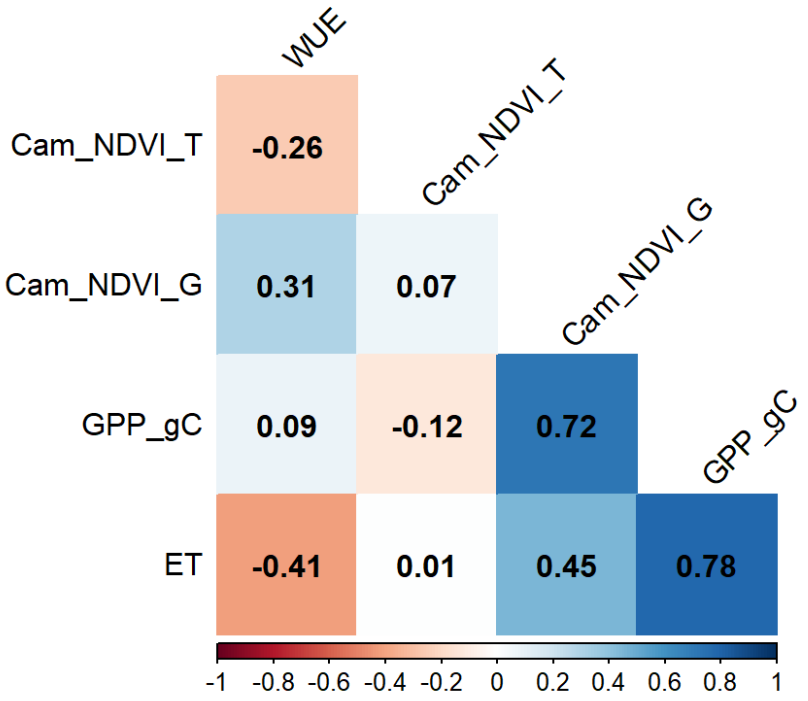
The NDVI values for the tree layer showed little variation seasonally and interannually. This is due to local plant species of the evergreen Mediterranean type that remains green and functional through more than one growing season. On the other hand, the grass area showed high seasonal variation and this behavior is probably mainly a reflection of precipitation variation and other climatological variables.

The dataset of daily NDVI measurements in trees and grass was joined with the GPP and ET data for analysis.





GPP and NDVI grass showed a high correlation throughout the analyzed series. This may mean that the ecosystem's GPP may be more strongly related to the grassland layer. The figure below presents a correlation matrix for the analyzed variables.



8 Annual Fluxes for the Ecosystem

The table below presents the annual values for GPP (g C m−2 year−1), ET (ET kg H2O m−2 year−1) and WUE (g C kg H2O−1) in the two years analyzed. The values presented are consistent with those reported in other works, for example, in the work carried out by El-Madany et. al. (2021) [2].

Annual Fluxes

date	GPP_gC	ET	WUE
2018	1038.7669	407.5151	2.549027
2019	983.5979	417.5322	2.355742

References

[1] Gómez-Giráldez, P.J.; Pérez-Palazón, M.J.; Polo, M.J.; González-Dugo, M.P. Monitoring Grass Phenology and Hydrological Dynamics of an Oak–Grass Savanna Ecosystem Using Sentinel-2 and Terrestrial Photography. Remote Sens. 2020, 12, 600. https://doi.org/10.3390/rs12040600 (https://doi.org/10.3390/rs12040600)

[2] El-Madany, T. S., Reichstein, M., Carrara, A., Martín, M. P., Moreno, G., Gonzalez-Cascon, R., et al. (2021). How nitrogen and phosphorus availability change water use efficiency in a Mediterranean savanna ecosystem. Journal of Geophysical Research: Biogeosciences, 126, e2020JG006005. https://doi.org/10.1029/2020JG006005 (https://doi.org/10.1029/2020JG006005)