Supplement of Biogeosciences, 23, 6687–6703, 2026
https://doi.org/10.5194/bg-23-6687-2026-supplement


Biogeosciences
Open Access

*Supplement of*

# Using deep learning to assimilate sun-induced fluorescence satellite observations in the ISBA land surface model

**Pierre Vanderbecken et al.**

*Correspondence to:* Jean-Christophe Calvet (jean-christophe.calvet@meteo.fr)

## S1 Details on the Neural Network design

### S1.1 Choice of the output

The neural network tends to perform better with normalised data and with distribution close to a Gaussian. After filtering, the distribution of SIF measurements becomes log-normal-like with a mode close to 0 as shown in Fig. S1(a). To make the distribution more Gaussian-like, and thus ease the training of the NN, we applied the function of $SIF_l$ described in the article. The distribution of the transformed values $SIF_l$ is displayed in Fig. S1(b).

### S1.2 Choice of the inputs

The SIF signal is produced under specific conditions during the photosynthesis process (Rossini et al., 2015). Its magnitude is correlated with environmental conditions such as the hydrological stress experienced by vegetation (Sun et al., 2015). On a larger scale, this results is an almost linear relationship between the SIF signal, the LAI, and the GPP. However, the slope of this relationship varies according to the vegetation type, the regional climate, and the season (Leroux et al., 2018). The following predictors were considered: the day of the year (DOY), the latitude (LAT) and the longitude (LON), the surface elevation (ZSE), the surface soil moisture (SSM), the land surface temperature (LST), the net radiative flux (RNF), the evapotranspiration (ET), the sensible heat flux (H), the GPP, and the LAI. These variables were obtained through a reanalysis using the LDAS, which was computed in advance. The reanalysis was computed over the European domain, extending from 26°W to 46°E and from 28°N to 72°N with a resolution of 0.1. As it was done in (Hamer et al., 2025), the reanalysis results from the assimilation of LAI-V1 (Baret et al., 2013) in ISBA-A-gs, and use ECOCLIMAP-II as the land cover map. The reanalysis used ECMWF high-resolution analysis as atmospheric forcing. In addition to these simulated variables, the 10-day average LAI-V1 product was also considered as an input for the NN. Therefore, with the exception of LAI-V1 data, all the physical variables were taken from a model reanalysis defined over the European domain. To gain insight into the relationship between these variables and the TROPOMI SIF data, we computed Pearsons correlation coefficient for each variable over a one-year period (Fig. S2).

The candidate inputs belong to one of these three categories:

- control variables of ISBA (i.e. LAI, SSM)
- derived variables of ISBA (i.e. GPP, ET, RNF, LST, H)
- metadata (i.e. LAT, LON, DOY, ZSE)

Control variables are prognostic variables of the ISBA model, meaning they can be updated through the data assimilation process. Derived variables are physical quantities computed from the prognostic variables or the atmospheric forcings. They cannot be changed directly by data assimilation. Metadata are specific to the current grid and domain in use and cannot be modified during assimilation. According to Fig. S2, LAI-V1 observation data shows the strongest correlation with TROPOMI SIF. Therefore, LAI-V1 was chosen as the LAI input for training the NN rather than LAI-mdl , since LAI-V1 is more highly correlated with TROPOMI SIF. Of the reanalysis variables, only H and SSM are uncorrelated with the SIF estimates. We have chosen to eliminate the sensible heat flux (H) from the candidate inputs, since it is uncorrelated with TROPOMI SIF and is a derived variable. The LST and the RNF are correlated, but are removed from the candidate inputs, since they are computed from the atmospheric forcings. We have chosen not to introduce make the NN dependent on the atmospheric forcings used. According to their respective correlations with TROPOMI SIF and the other variables, the metadata should be excluded from the inputs. However, we found that the results of the training without the DOY, LAT and LON produced a worse NN model in terms of correlation and mean square error. These metadata provide information on the expected seasonal cycles and the spatial variability. Nevertheless, it is questionable whether they will provide information that cannot be accommodated in our assimilation setup (i.e. ISBA is a one-dimensional model), which could prevent the model from generalising well on a different grid scale or domain. To evaluate the information brought by one the inputs with respect to the other inputs, we compute the shapley importance expected values (Lundberg and Lee, 2017) assuming independency between the inputs and a Gaussian kernel. This method of feature importance originates from the gaming industry and can be interpreted as a qualitative sensitivity analysis of the resulting SIF estimate with respect to the different inputs. The average expected shapley value over a sample of

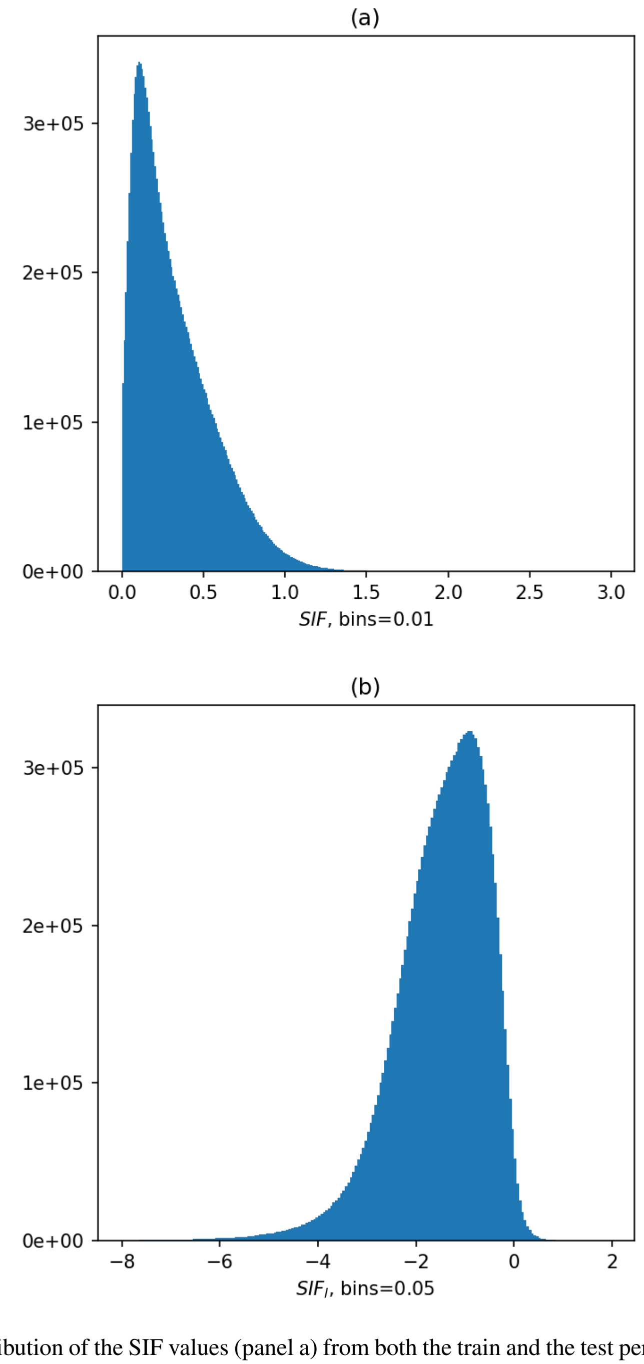


**Figure S1.** Comparison of the distribution of the SIF values (panel a) from both the train and the test period and the analog $SIF_l$ values (panel b) used for the neural network.

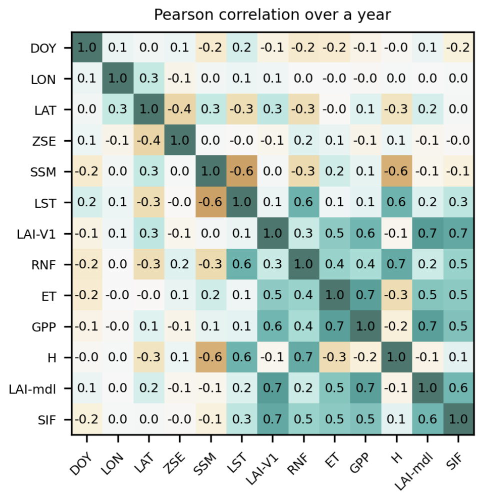


**Figure S2.** Pearson's correlation between candidate inputs and output.

**Table S1.** Mean of the absolute Shapley values of the NN's inputs.

| Inputs | Mean(\|shapley value\|) |
|---|---|
| DOY | 0.464 |
| LAI-V1 | 0.282 |
| LAT | 0.246 |
| LON | 0.058 |
| ZSE | 0.049 |
| SSM | 0.008 |
| GPP | 0.006 |
| ET | 0.006 |

1000 of the train dataset are displayed in Tab. **??**, and it appears that the GPP and the ET are the least important inputs, while the LAI, LAT and the DOY have the greatest impact on the estimation. Based on the results of this training test, it has been decided that the ET, GPP, SSM and ZSE inputs can be removed. Unexpectedly, the GPP had almost no impact. The hypothesis of assuming that these inputs are uncorrelated may interfere in these results, but in our case we can assume that the LAI likely provides most of the necessary information to emulate accurately enough TROPOMI SIF. We tried the configuration only using LAI-V1, LAT, LON and DOY. Fig. S3 shows a two-dimensional histogram that compares the SIF estimates from the neural network with the targeted observed TROPOMI SIF values for the test period using the Neural network described in the article (ie using LAI-V1 observation, LAT, LON and DOY) and Fig. S4 is the same but using the model trained using the 8 inputs listed in Tab. **??**. The RMSE and the correlation are similar in the two configurations. Both NNs have difficulties reproducing the extremes of TROPOMI SIF values. This is understandable, given that we used a Huber loss function focused on minimising the RMSE.

The NN used in the paper presents outliers (around 400) in Fig. S3, but the objective here is not to reproduce the TROPOMI SIF signal perfectly, but rather to design an operator that strikes a balance between the accuracy of the estimated TROPOMI SIF observation, ease the use within our data assimilation framework, and has sufficient impact on the analysis. It is worth to note that including derived variables complicate the assimilation scheme since it requires to either add them to the control vector or to find a way to update them post assimilation. Involving prognostic variables and diagnostic variables from the model

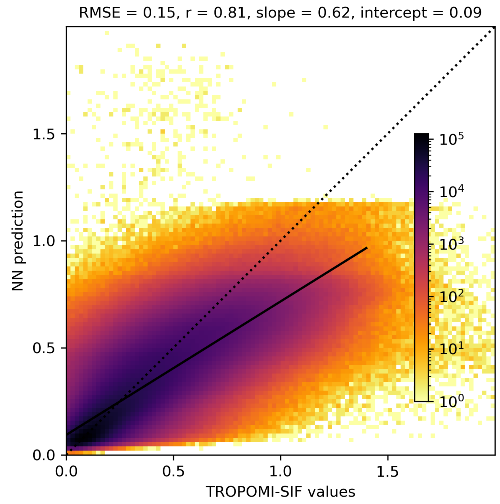


**Figure S3.** Comparison between the neural-network SIF simulation (using 4 inputs) and the values of TROPOMI SIF for the test dataset in mW.m$^{-2}$.sr$^{-1}$.nm$^{-1}$. The dashed line is the $y = x$ curve. The straight line is the result of the linear regression of the NN predictions in dependency of the TROPOMI SIF values. The color scale shows the number of observations for a given bin.

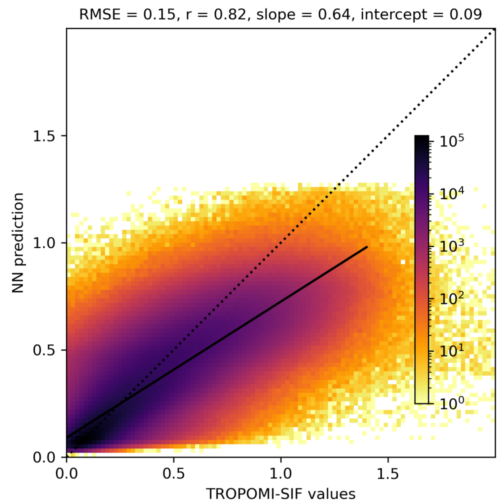


**Figure S4.** Comparison between the neural-network SIF simulation (using 8 inputs) and the values of TROPOMI SIF for the test dataset in mW.m$^{-2}$.sr$^{-1}$.nm$^{-1}$. The dashed line is the $y = x$ curve. The straight line is the result of the linear regression of the NN predictions in dependency of the TROPOMI SIF values. The color scale shows the number of observations for a given bin.

also bring the question of the model biases that are likely to be learnt by the neural network. This leads us to discard all these predictors to keep only the observed LAI-V1, LAT, LON and DOY for the training to make sure that the observation operator would not be specific to our current version of the land surface model.

### S1.3 Learning process details

Due to regularisation, the validation loss tends to be lower than the training loss. However, as Fig. S5 shows, both converge. We compare the learning curves in terms of both the Huber loss applied to $\mathrm{SIF}_l$ (left panel) and the mean square error on the SIF estimation reconstructed from $\mathrm{SIF}_l$ (right panel). The reconstructed mean squared error appears to be more variable, but remains low. A learning scheduler was used for the training, the learning rate $\lambda$ dropped every 20 epochs. If $n$ is the number of epochs then the learning rate follows the equations Eq. S1.

$$m = floor(\frac{n}{20}) \tag{S1a}$$

$$\lambda(n) = 0.001 * \prod_{i=0}^{m}(0.5)^i \tag{S1b}$$

This ranges from 0.001 for the first twenty epochs to about $1\times10^{-6}$ for the last epoch. This results in the large steps in the learning curves displayed in Fig. S5. Different splits of the training datasets were tested from 40% for the training and 60% for the validation to 80% for the training and 20% for the validation. Increasing the training part slightly reduces the loss values, but due to the size of the training dataset, it significantly increases the training time. The training dataset comprises more than 14 million of input/output pairs, and we performed data augmentation to double the initial size of the dataset. This large amount of data is likely why splitting the data so that there is less for training than validation achieves similar loss values. In the article, we used a split of 40% for training and 60% for validation.

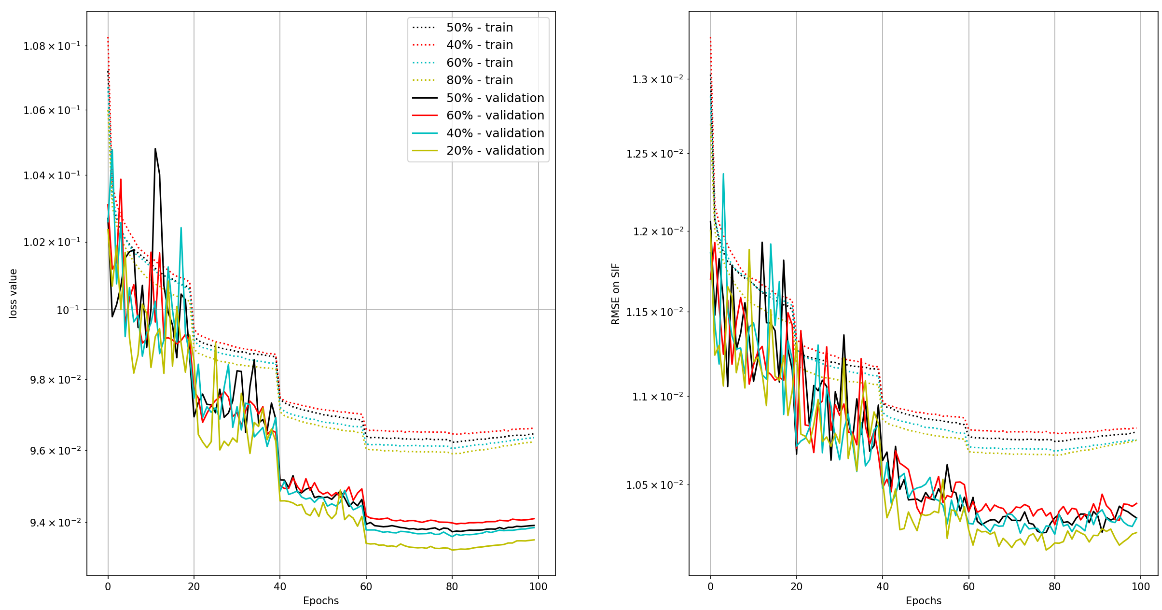


**Figure S5.** Learning curves of the neural-network over the SEEDS domain.

## S2 Qualitative comparison with the Hyplant airborne SIF measurement campaign over the Ebro basin

In July 2021, an airborne SIF measurement campaign was conducted over the irrigated fields of the Ebro basin (Rascher et al., 2015), which lies within the grid cell used for the previous time series. The measurement resolution is approximately 10 m and the observed area is primarily covered by C3 crops and bare soil. The airborne measurement of the 27th of July are displayed in Fig. S6(a).

The operator was initially trained to mimic TROPOMI SIF retrieval using the LDAS control variables. Here, we discuss whether the neural network can provide relevant first guess of other SIF observations. We compare the daily SIF values retrieved from the open-loop experiment (OL) and the $SIF_{20}$ experiment from the article (Table 1) for the selected grid cell located with the 10-m resolution airborne measurements from the HyPlant campaign and with the TROPOMI SIF values. As the in situ observations and the TROPOMI SIF are centred on different wavelengths, we will focus more on the temporal variations than on the actual values of the simulated SIF. The significant difference in spatial resolution means that we must compare all in-situ SIF measurements on a given day diplayed in the boxplots of Fig. S6(b) with the simulated SIF value resulting from the analysis for the two simulations shown in Fig. S6(c). Comparing the results with those of the OL experiment reveals that when TROPOMI SIF is assimilated, the values are higher overall and closer to the range of airborne measurement value, though they do not follow the airborne daily average variations. As expected, the simulated SIF values are closer to the TROPOMI SIF values in terms of both magnitude and temporal variation than the airborne measurement average, since the neural network was trained to match the TROPOMI SIF retrieval. This is an example of the limitations of using a neural network and show that it is specific to the TROPOMI SIF daily product. The neural network was trained using observations, which means it is independent from the LSM used for assimilation, This does not make it a substitute for a SIF physical parameterisation.

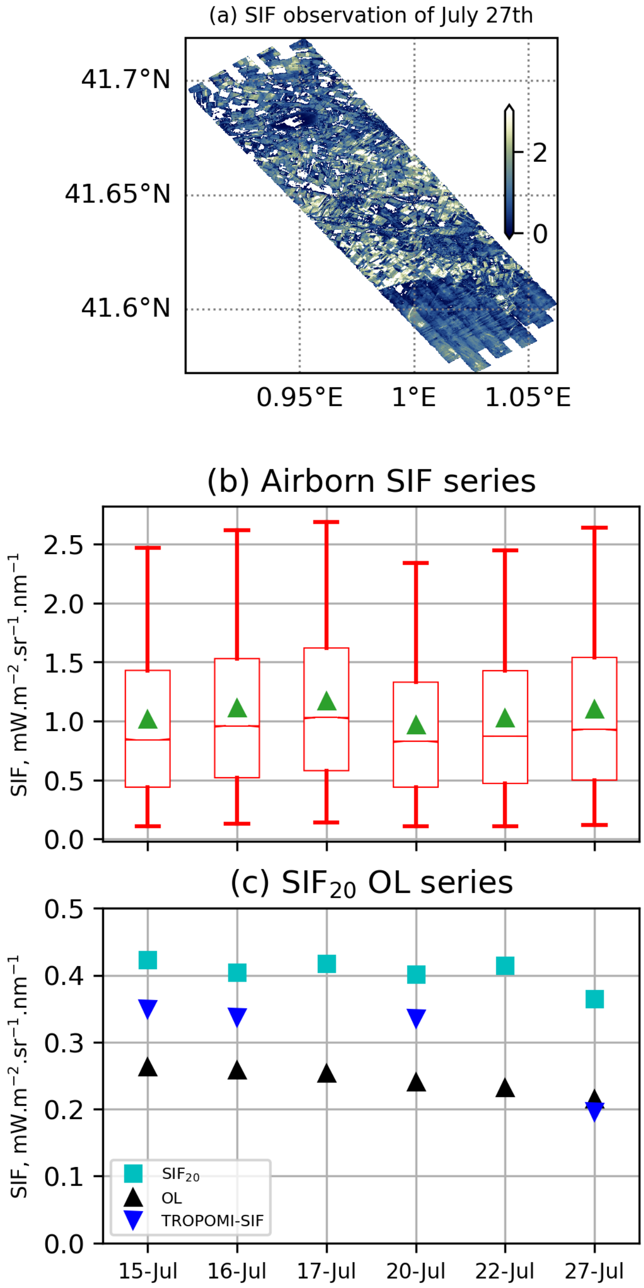


**Figure S6.** Comparison of the OL and $SIF_{20}$ experiments with the SIF airborne measurements of the HyPlant instrument within the LIAISE campaign. Boxes represents the 25th and the 75th percentiles, the whiskers the 5th and 95th percentiles of the airborne SIF observations. The central horizontal red lines represent the 50th percentile and the green triangles the mean of the airborne SIF observations.

## S3 FLUXCOM-X GPP over the Ebro basin from 2018 to 2021

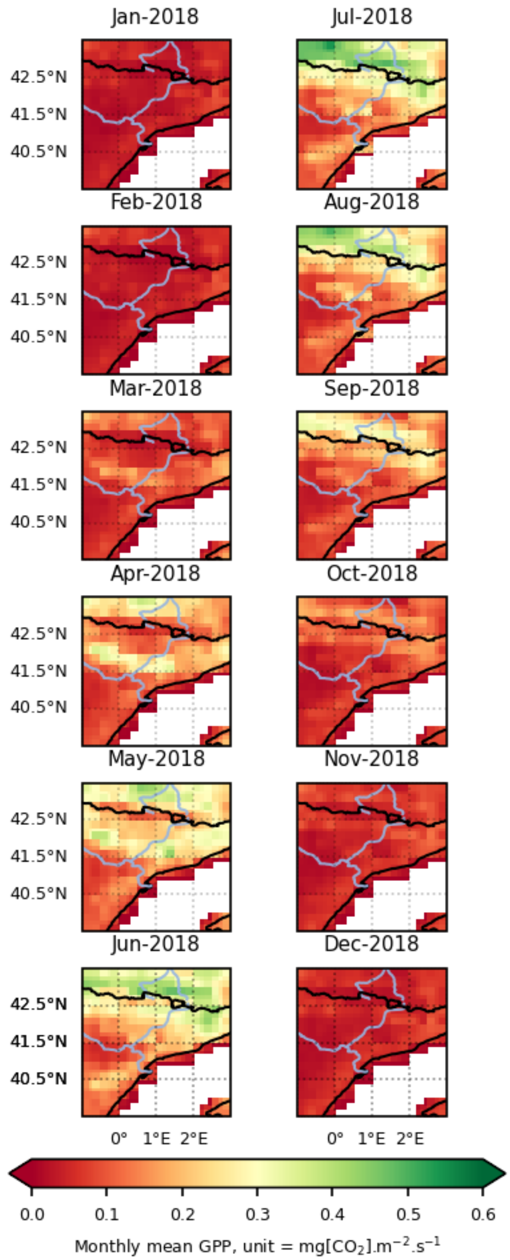


**Figure S7.** Monthly GPP mean from FLUXCOM-X over the Ebro basin for the year 2018

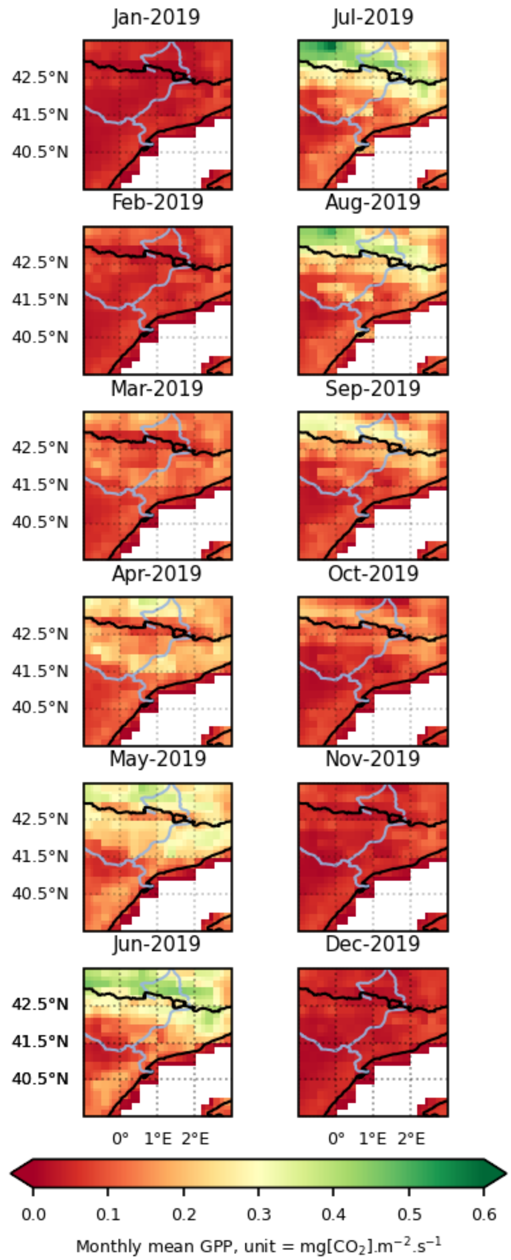


**Figure S8.** Monthly GPP mean from FLUXCOM-X over the Ebro basin for the year 2019

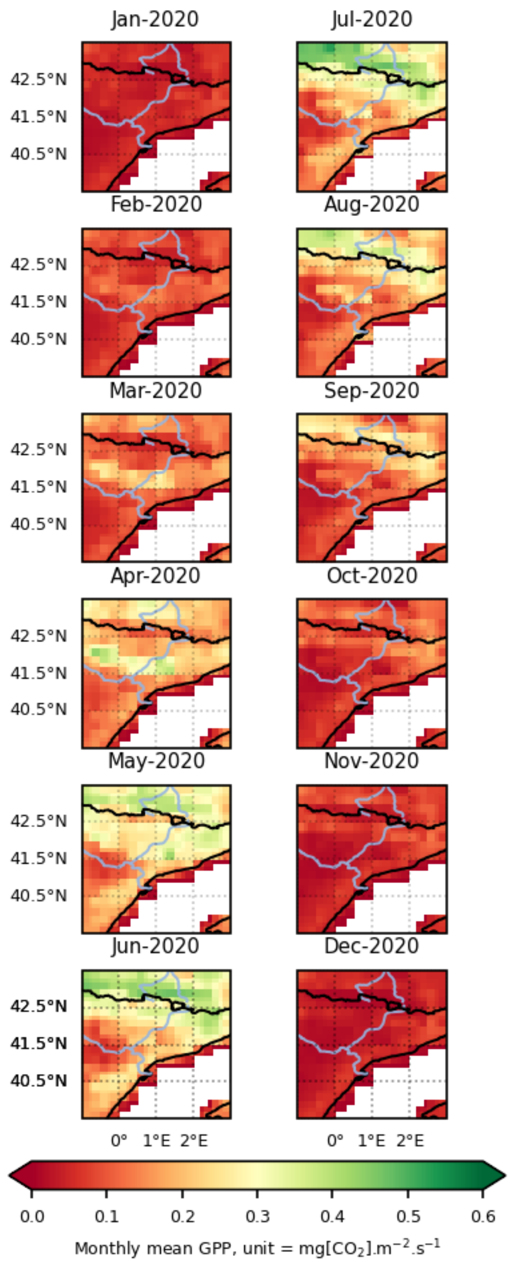


**Figure S9.** Monthly GPP mean from FLUXCOM-X over the Ebro basin for the year 2020

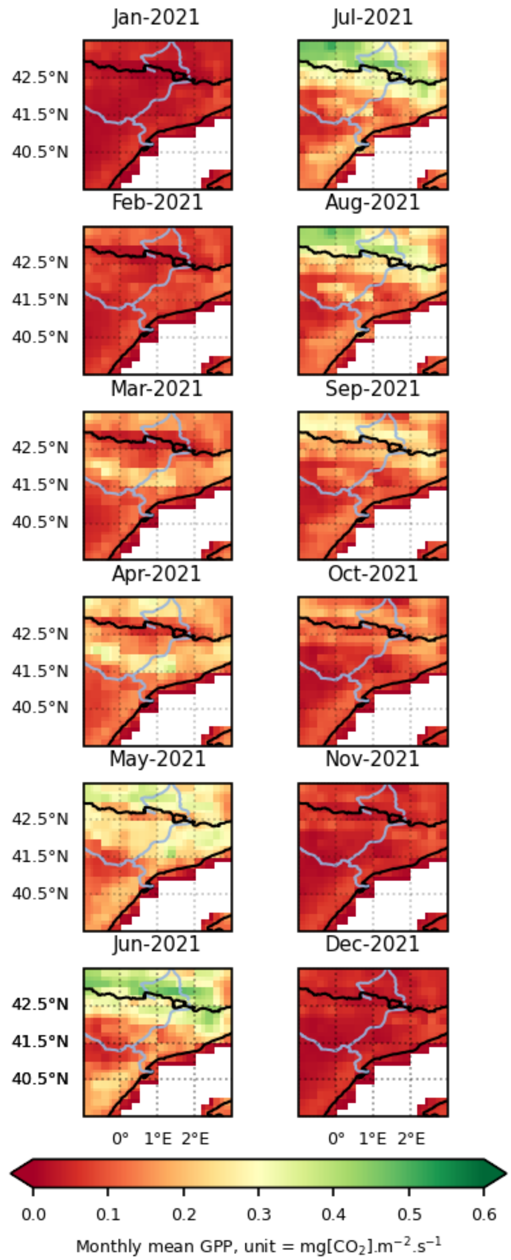


**Figure S10.** Monthly GPP mean from FLUXCOM-X over the Ebro basin for the year 2021

## References

Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmara, H., Pacholcyzk, P., and Smets, B.: GEOV1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part1: Principles of development and production, Remote Sensing of Environment, 137, 299–309, https://doi.org/https://doi.org/10.1016/j.rse.2012.12.027, 2013.

Hamer, P. D., Markelj, M., Rojas-Munoz, O., Bonan, B., Calvet, J.-C., Marécal, V., Guenther, A., Trimmel, H., Vallejo, I., Eckhardt, S., Sousa Santos, G., Sindelarova, K., Simpson, D., Schmidbauer, N., and Tarrasón, L.: Two Biogenic Volatile Organic Compound Emission Datasets over Europe Based on Land Surface Modelling and Satellite Data Assimilation, Earth System Science Data Discussions, 2025, 1–48, https://doi.org/10.5194/essd-2025-442, 2025.

Leroux, D. J., Calvet, J.-C., Munier, S., and Albergel, C.: Using Satellite-Derived Vegetation Products to Evaluate LDAS-Monde over the Euro-Mediterranean Area, Remote Sensing, 10, https://doi.org/10.3390/rs10081199, 2018.

Lundberg, S. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, https://arxiv.org/abs/1705.07874, 2017.

Rascher, U., Alonso, L., Burkart, A., Cilia, C., Cogliati, S., Colombo, R., Damm, A., Drusch, M., Guanter, L., Hanus, J., et al.: Sun-induced fluorescence–a new probe of photosynthesis: First maps from the imaging spectrometer HyPlant, Global change biology, 21, 4673–4684, 2015.

Rossini, M., Nedbal, L., Guanter, L., Ač, A., Alonso, L., Burkart, A., Cogliati, S., Colombo, R., Damm, A., Drusch, M., et al.: Red and far red Sun-induced chlorophyll fluorescence as a measure of plant photosynthesis, Geophysical research letters, 42, 1632–1639, 2015.

Sun, Y., Fu, R., Dickinson, R., Joiner, J., Frankenberg, C., Gu, L., Xia, Y., and Fernando, N.: Drought onset mechanisms revealed by satellite solar-induced chlorophyll fluorescence: Insights from two contrasting extreme events, Journal of Geophysical Research: Biogeosciences, 120, 2427–2440, 2015.