https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

# A neural-network based forward operator for the assimilation of microwave satellite observations with LDAS-Monde

Jasmin Vural[1], Pierre Vanderbecken[1], Bertrand Bonan[1], Oscar Rojas-Muñoz[1], and Jean-Christophe Calvet[1]

[1]Météo-France, CNRS, Univ. Toulouse, CNRM, Toulouse, France

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

**Abstract.** The knowledge of land-surface variables (LSVs) is essential for an accurate description of the carbon cycle and, hence, for deriving the fraction of anthropogenic $CO_2$ emissions contributing to global warming. For a better representation of LSVs like the leaf-area index (LAI), we assimilate the observations of the satellite microwave sensors SMAP and AMSR2 with our land-data assimilation system (LDAS). Dedicated observation operators using a neural network (NN) are developed to enable a direct assimilation of the measured brightness temperatures. Direct assimilation is not yet an established method due to the incomplete representation of physical emission processes and the associated computational constraints of physical observation operators. We derive an optimal set of predictors for both instruments resulting in a good match between model equivalent and the observations in the testing period, with correlations up to 0.87/0.93 and a total RMSE of 8.2/3.6 K for SMAP/AMSR2, respectively. The implementation of the derived weights into the LDAS is straightforward and is found to lead to a reasonably good performance of the assimilation system, with a stronger improvement of the departures for SMAP than for AMSR2, which can be attributed to the characteristics of the microwave bands of the observations. For the best assimilation experiments, the verification against LAI observations shows an improvement compared to the open loop on a global scale. For different global cropland regions especially prone to droughts, AMSR2 outperforms SMAP in most cases. Nevertheless, the seasonal and subseasonal variability is still not well represented though due to unsolved issues in our model.

## 1 Introduction

With climate change being one of the most important challenges of our time, accurate climate projections are essential. For those, the knowledge of the current state of the main drivers of the global warming is needed and, hence, not only the increase of atmospheric species contributing to the greenhouse effect but also the fraction of both biogenic and anthropogenic greenhouse gases therein. As a consequence, the EU-funded project CORSO[1], where this study is part of, aims at providing a continuous monitoring of land surface variables, which allows for an improved estimation of the carbon cycle and, eventually, the biogenic $CO_2$ emissions.

With data assimilation (DA), the initial state of the variables of interest can be optimised by ingesting relevant observations. Especially for LSVs, the availability of microwave-satellite instruments observing in near-real time and providing gridded data

[1]https://www.corso-project.eu/

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

for the whole globe offers the possibility to update the model variables to an improved state (Walker and Houser 2001; Reichle 2005; Pauwels et al. 2007 among others).

Frequently, assimilation systems rely on retrieved data products (e.g., Dente et al. 2008; Jarlan et al. 2008; Draper et al. 2012) but with radiative transfer models (RTMs) becoming more accessible and efficient, the direct assimilation of brightness temperatures ($T_\mathrm{b}$) using an RTM-based observation operator is often preferred (Reichle et al., 2021; Vural et al., 2023; Kolassa et al., 2025) as it avoids inconsistencies added by simplifications required by the retrieval method.

As the main difficulty of transforming model variables to observation space to obtain the so-called model equivalent remains, recent approaches include the exploitation of different algorithms of machine learning, especially neural networks (Xue and Forman, 2015; Corchia et al., 2023; De Gélis et al., 2025). These methods are usually resource-intensive in the training part but the actual application of the training results as observation operator is generally straightforward and computationally significantly less expensive than RTMs .

While the assimilation of NN-based retrievals has been established (e.g., Rodríguez-Fernández et al. 2019 on SMOS, Kolassa et al. (2017) on SMAP, and Rodríguez-Fernández et al. 2016 on AMSR-E), SMAP brightness temperatures were also assimilated directly by Reichle et al. (2023) and Kolassa et al. (2025) using a tau-omega RTM with local calibration (De Lannoy et al., 2013), which led to a general benefit on the numerical weather prediction (NWP), including the representation of a tropical cyclone.

Although mostly used for retrieving soil moistures (SMAP; Soil Moisture Active Passive) or atmospheric variables (AMSR2; Advanced Microwave Scanning Radiometer), microwave observations can also be assimilated to improve variables related more closely to vegetation since it is responsible for absorption of the radiation before it reaches the ground. Kumar et al. (2020) improved the estimate of evapotranspiration and GPP by the assimilation of vegetation optical depth (VOD) derived from SMAP and AMSR2 over the United States. Similarly, Mucia et al. (2022) assimilated surface soil moisture, LAI, and VOD derived from different microwave bands and found improvements of varying magnitudes for different LSVs as well. In general, soil moisture itself is often challenging to improve but can benefit from a combined assimilation of active and passive data (Lievens et al., 2017).

Within this context, this study focuses on improving the representation of LAI as one of the important variables within the carbon cycle with the land-surface model (LSM) of Météo-France by the assimilation of microwave satellite observations provided by the two passive sensors SMAP and AMSR2.

Unlike previous studies, we work with low-level data products to avoid the aforementioned issues associated with the assimilation of retrievals. To our knowledge, this is the first time an observation operator based on a neural network has been used for assimilating land-surface variables using SMAP and AMSR2 observations.

The paper is organised as follows: Observations used for training, assimilation, and verification are described in Sect. 2. The LSM, the assimilation system, and their configurations are presented in Sect. 3. Section 4 presents the development of the NN-based observation operator as well as the assessment of the training results, Sect. 5 evaluates the performance of the data assimilation and, eventually, validates the method by comparing the analyses against actual observations. The results are discussed in Sect. 6 and the conclusions are presented in Sect. 7.

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

## 2 Observations

### 2.1 SMAP brightness temperatures

Originally designed for providing soil moisture retrievals at high resolution by combing active and passive remote sensing, the SMAP mission measures brightness temperatures over a variety of vegetation types. We work with the L1C-product (Chan et al., 2020), which contains multi-polarisation observations of $T_\mathrm{b}$ in the L band (1.41 GHz). The observations are gridded on the EASE-Grid 2.0 at 36 km (Brodzik et al., 2012).

The instrument's design is in principle able to account for anthropogenic Radio Frequency Interference (RFI; Chan et al. 2015) with the according quality flags being provided in the product. Our data preprocessing includes flagging with the quality flags given for all polarisations for both fore- and aft-looking data. The data is interpolated linearly onto the model grid. We combine several swaths within one defined time interval, which is constrained by the temporal resolution of the model output of 3 hours in this case. Spatially overlapping points are averaged within one time bin.

### 2.2 AMSR2 brightness temperatures

The AMSR2 instrument measures $T_\mathrm{b}$ in six different bands and two polarisations. The employed gridded L2B-product (Jackson et al., 2016) does not contain C-band due to too strong RFI contamination (Kidd, 2006). The X-band (10.7 GHz) is strongly attenuated by the vegetation canopy and shows therefore a strong sensitivity to the biomass (Das and Paul, 2015). It is therefore a suitable candidate for our study of LAI. The observations are gridded on the EASE-Grid 1.0 at 25 km.

We apply all flags provided in the data product that account for RFI, water, ice, snow, frozen ground, rain, wetland, and urban areas. In addition, we keep only data where more than 95% of the samples of the used footprints were flagged as good.

As for SMAP, the data is interpolated linearly onto the model grid and several swaths are combined within one the time interval of 3 hours by averaging spatially overlapping points.

### 2.3 Leaf area index

We employ two different LAI data sets, one for training and assimilation, and another one for verification. The LAI product for the training has a resolution of 1 km and is derived from the Advanced Very High Resolution Radiometer (AVHRR; Verger 2024) provided by the THEIA data centre. For verification, a 10-daily product of the Copernicus Land Monitoring Service (CLMS; Sentinel-3) with a resolution of $\approx 300$ m is used (Copernicus Land Monitoring Service and Copernicus Land Monitoring Service Helpdesk, 2017a). For matching to the model grid, the nearest neighbour method is applied for both data sets.

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

### 2.4 Gross primary productivity

To also verify against related surface variables, we employ derived observations of the Gross Primary Productivity (GPP) provided by the FLUXCOM-X framework (Nelson et al., 2024), which is a good measure of the carbon uptake of vegetation. The data has a daily temporal sampling and is spatially sampled on the same global $0.25°$ grid as our model data.

### 2.5 Soil-water index

For verifying soil moisture, the global soil-water index (SWI) of MetOp/ASCAT (Copernicus Land Monitoring Service and Copernicus Land Monitoring Service Helpdesk, 2017b; Wagner et al., 2013) at $0.1°$ is used. The instrument is a radar measuring in the C band and the SWI retrievals are based on a change-detection algorithm (Wagner et al., 1999). We apply the provided quality flags and interpolate the data set linearly to the model grid. Since soil moisture observations generally exhibit a bias against the model values (e.g., Reichle et al., 2004), the data are converted into anomalies by subtracting the mean from the actual value for each data set and scaling with the standard deviation.

## 3 The assimilation system

### 3.1 The ISBA model

The SURFEX modelling platform (Masson et al. 2013; version 8.1 here) comprises the description of four different land surfaces by flexible model components in a tile-based approach. For each tile of lake, sea, town, and land, the surface fluxes and the surface state variables are computed with the respective model. For the natural land surface, we employ the ISBA (Interaction-Sol-Biosphère-Atmosphère; Noilhan and Mahfouf 1996) land surface model. The activated multi-layer diffusion scheme (Boone et al., 2000; Decharme et al., 2011) can describe up to 14 soil layers to 12 m depth. Our focus of interest is an improved description of the carbon cycle and, consequently, we use the $CO_2$-responsive version ISBA-A-gs (Jacobs et al., 1996; Calvet et al., 1998) to account for the feedback of $CO_2$ with the vegetation. This feedback is implemented by including the interaction between $CO_2$ and photosynthesis with regard to stomatal respiration and plant growth. The LAI is computed once per day.

The grid cells are composed of a combination of patches that describe the land cover type. Computations on patch-level are aggregated to obtain the grid-cell value eventually. For creating the physiographic database (PGD), which is required for the extraction of land cover and other relevant surface features like altitude or topographic complexity of the terrain, the ECOCLIMAP-SG database is used (Calvet and Champeaux, 2020).

The LSM can be run "offline" and is thereby forced once per hour with 11 atmospheric variables derived from the ERA5 reanalysis providing a resolution of 0.25°. To initialise ISBA with more realistic values, we run a spin up for 20 years. Both open loop (OL) and assimilation experiments are launched from this state.

In our study, we employ ISBA in two distinct roles. First, we extract relevant variables from the open-loop run as predictors for the NN and, second, we employ ISBA as background model within the assimilation system.

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

### 3.2 LDAS-Monde

LDAS-Monde assimilates observations with a simplified Extended Kalman Filter (sEKF), where the background error covariances remain static with time (Mahfouf et al., 2009). An extensive description of LDAS-Monde can be found in Albergel et al. (2017), hence, in the following, we describe the most important components and adaptations only.

The variables of the control vector are evolved over an assimilation window of 24 h. The observation operator is linearised by the computation of finite differences (Hess, 2001), that means each control variable $x_j$ (here LAI and soil moistures of layers 2-7, i.e., 0.01 - 1 m depth) is perturbed by a small amount $\delta x_j = 0.001$ in corresponding units for constructing a Jacobian matrix $H$ with the elements $H_{ij}$ for each model equivalent $\mathcal{H}_i(x)$

$$H_{ij} = \frac{\mathcal{H}_i(\mathbf{x} + \delta x_j) - \mathcal{H}_i(\mathbf{x}))}{\delta x_j} \,, \tag{1}$$

which is then applied to the model variables.

When discussing the departures, we denote the innovations, residuals, and increments as follows:

$$\begin{aligned} \mathbf{d}_{\mathrm{b}}^{\mathrm{o}} &= \mathbf{y}_{\mathrm{o}} - H(\mathbf{x}_{\mathrm{b}}) \\ \mathbf{d}_{\mathrm{a}}^{\mathrm{o}} &= \mathbf{y}_{\mathrm{o}} - H(\mathbf{x}_{\mathrm{a}}) \\ \mathbf{d}_{\mathrm{b}}^{\mathrm{a}} &= \mathbf{x}_{\mathrm{a}} - \mathbf{x}_{\mathrm{b}} \,, \end{aligned} \tag{2}$$

where $\mathbf{y}_{\mathrm{o}}$ is the observation, $\mathbf{x}_{\mathrm{a}}$ the analysis, and $\mathbf{x}_{\mathrm{b}}$ the background vector.

Since previous work showed that the main impact of the assimilation of vegetation-sensitive observations can be found on croplands (Corchia et al., 2023; Shan et al., 2024), we use exclusively grid cells that are dominated by crops. With this restriction, we perform experiments on the global domain with a grid sampling of 0.25°. The selection process is based on the patches extracted from the PGD (Sect. 3.1), whereby only grid cells with a total fraction of $C_3$ and $C_4$ crops of 50% or more are considered (Fig. 1).

The actual impact of the observations on the analysis is controlled by the ratio between background and observation errors. The background errors are prescribed as 0.2 scaled with the LAI or the dynamic range of the soil moisture, respectively.

As a standard approach, we keep the observation errors of the brightness temperatures constant for all grid points. The initial observation errors for each of the observations are derived from the O-B time series obtained after the training of the NN (see Sect. 4).

## 4 A neural network as observation operator

### 4.1 Description of the method

To enable the direct comparison of observations and background required in the assimilation system, the model variables (Sect. 3.1) have to be transformed into observation space. Since the standard approach with a radiative transfer model usually occupies a lot of resources, we develop a forward operator that is based on an artificial neural network. Our strategy is to train a

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

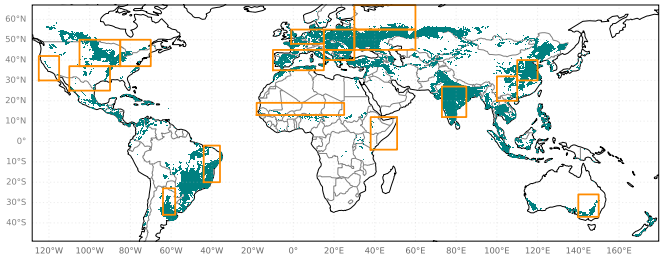


**Figure 1.** Map of the model domain, where the teal points display the croplands as used in the ISBA model. Orange rectangles denote areas that are especially prone to droughts and heatwaves as identified by Albergel et al. (2020). In agreement with the notation of these authors, the regions will be named with CALF, SPLN, MIDW, ENRT, PAMP, NDST, SAHL, WMED, WEUR, EEUR, BALK, WURS, LVOL, EAFR, INDI, SWCH, NRCH, MUDA from left to right, top-down.

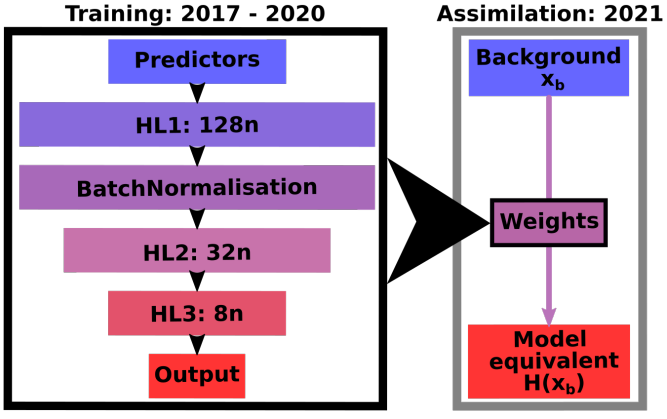


**Figure 2.** Sketch of the functionality of the observation operator. The architecture of the neural network is characterised by the input (predictors), the output, the hidden layers (HL), and the neurons (n).

NN on the $T_b$ observations in a first step and, subsequently, apply the found weights as forward operator to the model variables during the assimilation cycle (Fig. 2). This strategy was applied before (Shan et al., 2022; Corchia et al., 2023; Shan et al., 2024) on ASCAT observations.

To find the best setup, we explore the hyperparameter space of a feedforward NN, that means, different numbers of hidden layers, the number of neurons per layer, and learning rates. Eventually, the training turned out to not be very sensitive to the hyperparameters. This behaviour was already noticed by De Gélis et al. (2025) for the training on snow surfaces. Therefore and in regard of the time consumption of the training, a full brute-force scanning of the parameter space seemed neither necessary nor feasible. Instead, we varied each hyperparameter in several steps, choosing the optimum in terms of evaluated statistics of the obtained model equivalent vs. the observations in the testing period before varying the next hyperparameter.

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

Our data set spans four years (01/2017 − 12/2021) and is divided into training, validation, and test data with a fraction of 60%/20%/20%, respectively, where training and validation data are drawn randomly from the same period, whereas the test period covers the subsequent independent time span (2021).

For the input fields, we employ observed LAI (Sect. 2.3) as well as different static (e.g., coordinates, physiographic data) and variable model fields as predictors and analyse their contribution to the final training performance. An overview of all tested input parameters together with the bias, the root-mean-square error (RMSE), and the Pearson correlation coefficient is given in Table 1 to evaluate the performance of the model equivalent in comparison with the actual observations. To assess all skill metrics, including as well the standard deviation (STD), at the same time, they are displayed in a Taylor diagram (Taylor 2001, Fig. 3), where the best experiments are those closest to the observations. The majority of the experiments is actually not that different from each other on these scales but a trend towards better matching STD, RMSE, and correlation is clearly visible as well as the bad performance of experiments not using enough surface variables as predictors.

Like in previous work (Corchia et al., 2023) we focus on V-polarisation as output variable. When adding the H- to the V-polarisation as output fields, the impact onto the skill metrics is negligible compared to training on V-polarisation only and, additionally, H- always performs worse than V-polarisation when being the only output variable, likely due to the in general lower emissivity (cf. De Gélis et al. 2025).

### 4.2 Performance of the neural network

In addition to the predictors that have already been employed in similar studies, we achieved a significant improvement of the model equivalent when using forcing data. As overfitting by using too many predictors has to be avoided, we systematically omit predictors to test for redundancy. In addition, we compute the feature importance based on a measure developed originally in cooperative game theory, so-called Shapley values, using the SHAP library of python (Lundberg and Lee, Fig. 4). Some static predictors, especially latitude (SMAP) or ZS (altitude, AMSR2), have a high importance but as well SWI2 (soil water index in the soil layer between 0.01 - 0.04 m), Tair (near-surface air temperature), and TG2 (soil temperature, in the soil layer between 0.01 - 0.04 m). The LAI seems of minor importance though, which might be due to the NN extracting similar and additional information from other variable combinations, which cannot be further identified here. Direct shortwave radiation (SWdir) seems to be exploited very differently by the two instruments. This indicates that AMSR2 brightness temperatures are more sensitive to short-range diurnal variations.

Some predictors show only a minor feature importance (Fig. 4), thus, we tested if some, especially, the static predictors might have become redundant, but each omission of one of them degraded the performance (Table 1, Fig. 3). We suspect that the strong benefit seen from adding the forcing variables might mainly originate in the addition of information of the diurnal cycle that is not contained in TG2, especially, since the omission of one of the variables exhibiting a diurnal cycle slightly degrades the performance again.

This evaluation is based on training on the maximum numbers of predictors employed here and the SHAP library only works with a limited sample of the whole data set. Nevertheless, these findings are mostly consistent among the different experiments and runs.

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

As expected, adding variables of deeper soil layers did not bring further impact (not shown here) since the observations cannot penetrate deep into the ground.

**Table 1.** Overview of training experiments (exp) for both SMAP and AMSR2 with the predictors that were varied as well as STD (in K), RMSE (in K), correlation coefficient ($\rho_P$), and bias (in K). The columns describe the longitude (lon), latitude (lat), day of the year (DOY), LAI, soil water index at level 2 (SWI2), soil temperature at level 2 (TG2), altitude (ZS), topographic complexity (ToCo), air temperature (Tair), air humidity (Qair), long wave radiation (LW), direct short wave radiation (SWdir), and scattered short wave radiation (SWsca). The experiments employed later on for data assimilation (Sect. 5) are highlighted with bold font.

| | | | | | | | | | | | | | | SMAP | | | | | AMSR2 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Lon | Lat | DOY | LAI | SWI2 | TG2 | ZS | ToCo | Tair | Qair | LW | SWdir | SWsca | exp | STD | RMSE | $\rho_P$ | bias | exp | STD | RMSE | $\rho_P$ | bias |
| x | x | x | x | - | - | - | - | - | - | - | - | - | sn1 | 14.05 | 11.02 | 0.76 | 0.23 | an1 | 8.47 | 7.53 | 0.74 | -0.23 |
| x | x | x | x | x | - | - | - | - | - | - | - | - | sn2 | 14.10 | 9.20 | 0.83 | 0.04 | an2 | 7.43 | 7.04 | 0.72 | -0.12 |
| x | x | x | x | x | x | - | - | - | - | - | - | - | sn3 | 14.24 | 9.06 | 0.84 | -0.42 | an3 | 9.21 | 4.51 | 0.89 | 0.02 |
| x | x | x | x | - | x | - | - | - | - | - | - | - | sn4 | 13.30 | 11.22 | 0.73 | -0.04 | an4 | 9.11 | 4.76 | 0.88 | -0.23 |
| x | x | x | x | x | x | x | - | - | - | - | - | - | sn5 | 14.53 | 8.66 | 0.85 | 0.32 | an5 | 9.48 | 4.04 | 0.92 | -0.21 |
| x | x | x | x | x | x | x | x | - | - | - | - | - | sn6 | 14.68 | 8.42 | 0.86 | -0.04 | an6 | 9.47 | 4.04 | 0.92 | -0.26 |
| x | x | x | x | x | x | - | x | - | - | - | - | - | sn7 | 14.54 | 8.69 | 0.85 | -0.05 | an7 | 9.41 | 4.19 | 0.91 | -0.29 |
| x | x | x | x | x | x | x | x | x | x | x | x | x | sn8 | 14.91 | 8.16 | 0.87 | -0.18 | an8 | 9.55 | 3.67 | 0.93 | -0.58 |
| **x** | **x** | **x** | **x** | **x** | **x** | **x** | **x** | **x** | **-** | **x** | **x** | **-** | **sn9** | **14.78** | **8.21** | **0.87** | **0.01** | **an9** | **9.58** | **3.62** | **0.93** | **-0.26** |
| x | x | - | x | x | x | x | x | x | - | x | x | - | sn10 | 14.88 | 8.40 | 0.86 | 0.43 | an10 | 9.47 | 3.75 | 0.93 | 0.11 |
| - | x | x | x | x | x | x | x | x | - | x | x | - | sn11 | 14.82 | 8.50 | 0.86 | 0.23 | an11 | 9.44 | 3.87 | 0.92 | -0.17 |
| x | - | x | x | x | x | x | x | x | - | x | x | - | sn12 | 14.48 | 8.58 | 0.85 | 0.16 | an12 | 9.51 | 3.81 | 0.93 | -0.38 |

With respect to all displayed metrics (Table 1, Fig. 3), experiment sn9 for SMAP and experiment an9 for AMSR2 are selected for the following discussion and for assimilation. For the SMAP observations, the 2D-histogram (Fig. 5, left) shows that whereas the vast majority of points is close to the diagonal, a number of especially low-value observations is not matched well. The correlation coefficient reaches 0.87 (Table 1) with better values in the Northern hemisphere but also Australia (Fig. 6, top right). The RMSE is on average around 8.2 K with best values occurring in Europe, Brazil, and Indonesia (Fig. 6, top left).

In general, the AMSR2 observations show less variability and are less subject to flagging than the SMAP observations and are, thus, easier to reproduce by the NN. This is reflected in the performance statistics, where higher correlations (Fig. 6, bottom right) and smaller RMSE (Fig. 6, bottom left) are obtained compared to SMAP. The total correlation reaches 0.93 and exhibits high values all around the globe. The spatially averaged RMSE is 3.6 K, with best values in South America and Europe but also in parts of other regions. Values on the low end of the brightness temperature distribution are better matched than for SMAP and more often over- than underestimated, however, values on the high end are partially underestimated (Fig. 5, right).

Bad performance often arises from a lack of data during the training period and especially affects areas that exhibit values around the thresholds used for masking, like frozen ground, which occurs mainly in the northernmost latitudes. This could

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

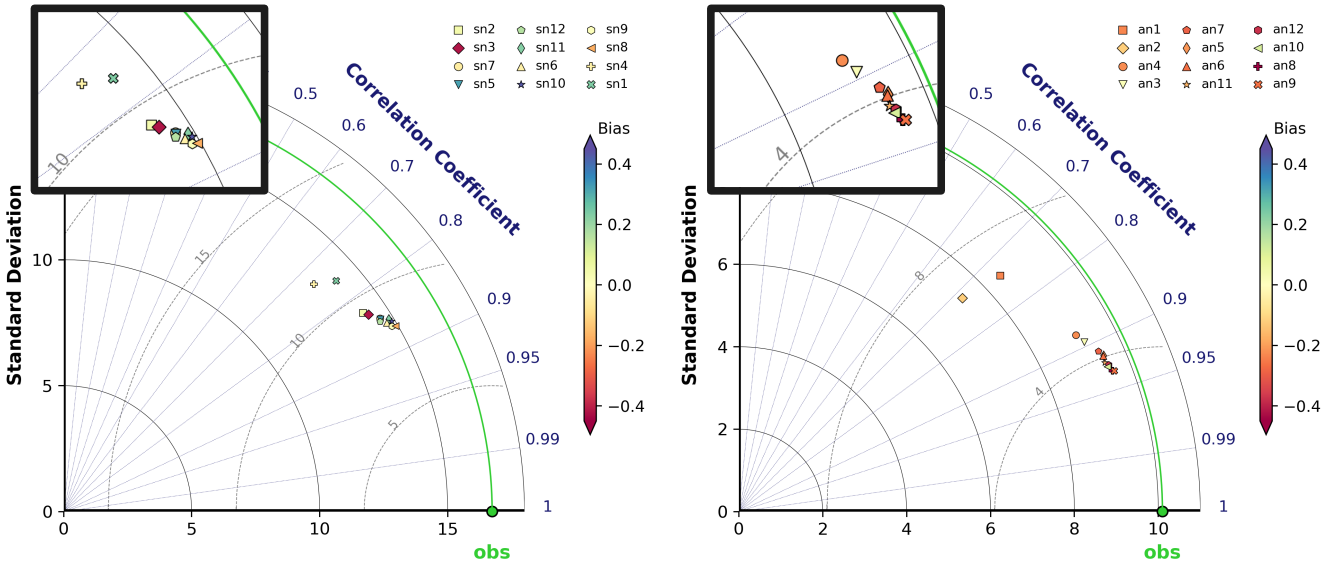


**Figure 3.** Taylor diagrams (adapted from Rochford 2016) of SMAP (left) and AMSR (right) training experiments described in Table 1. The STD is shown on the x- and y-axes, the correlation coefficient corresponds to the polar angle, the RMSE is marked by the distance from the observations (green point), and the bias is indicated by the color of the respective point, zooms on the dense point clouds are inserted within the black rectangles.

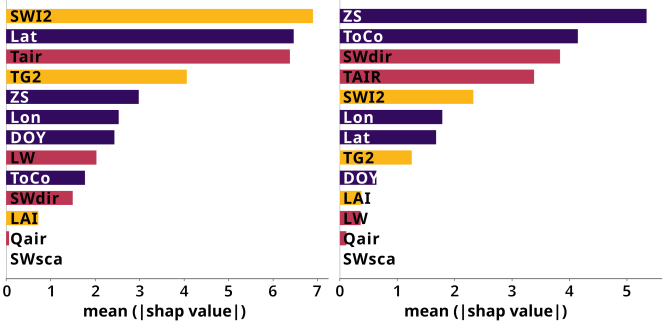


**Figure 4.** Feature importance using Shapley values of training experiments using all tested predictors for SMAP (left) and AMSR2 (right). Static predictors are indicated in dark violet, forcing variables in red, and surface variables in orange. For the explanation of the variable short names, see Table 1.

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

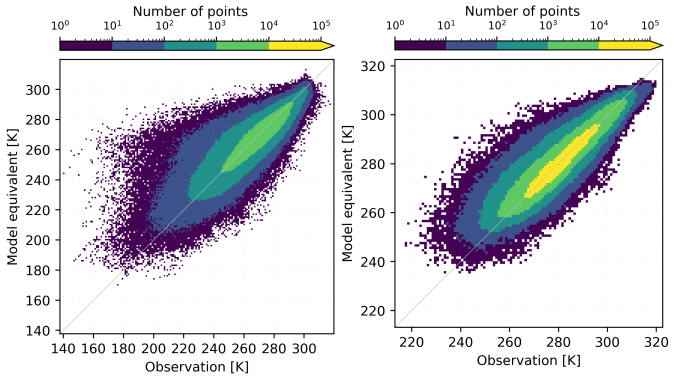


**Figure 5.** 2D histograms comparing model equivalents and observations for SMAP (exp sn9; left) and AMSR2 (exp an9; right).

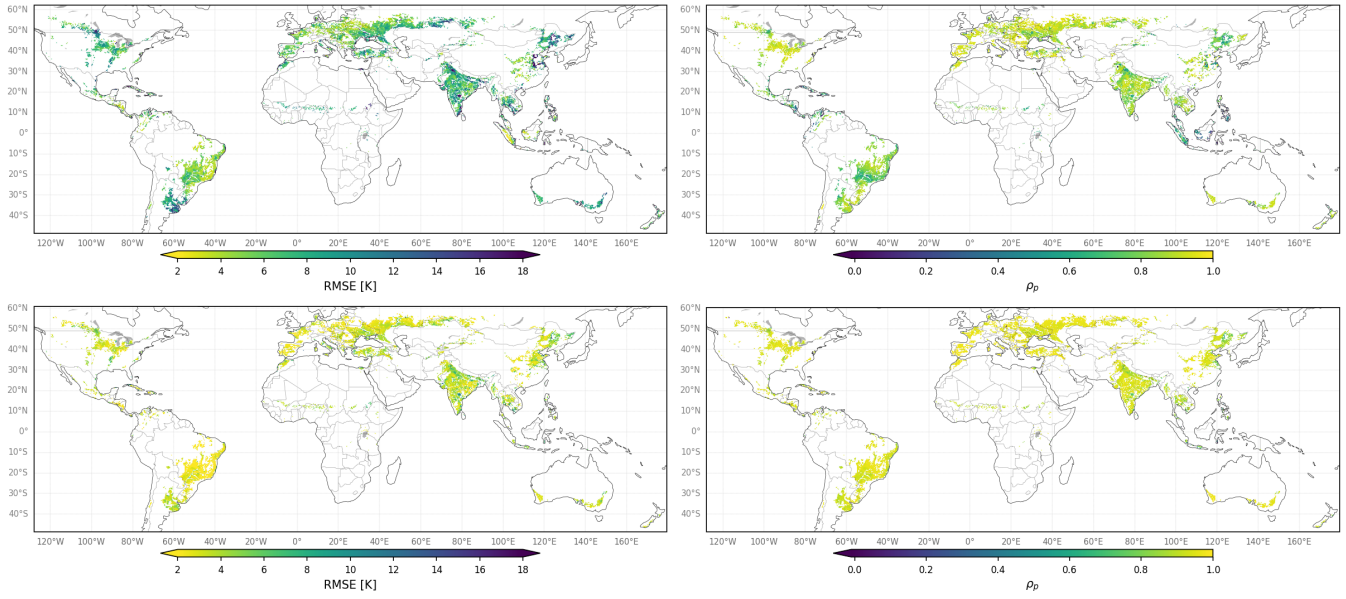


**Figure 6.** RMSE (left) and correlation (right) between observations (top: SMAP, bottom: AMSR2) and model equivalent for the best experiments identified in Sect. 4 (exp sn9 and exp an9, respectively, according to Table 1).

explain a higher RMSE of SMAP in Canada or in the North of Kazakhstan. Areas in China with high RMSE could be the consequence of insufficiently filtered RFI.

Since the penetration depth is larger at L band than at X band, the latter is less sensitive to variations of the soil properties in vegetated areas (Calvet et al., 2011). In return, L band is more affected by agricultural practices. The subsequent larger variability of the L-band signal compared to the X band renders the training process more challenging for the NN, consequently leading to larger errors for the training on SMAP observations.

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

## 5 Data assimilation

### 5.1 Preparation of the assimilation experiments

We implemented the observation operator described in Sect. 4 into our assimilation system (Sect. 3). We ran all assimilation experiments with the same general setup of LDAS-Monde but account for each of the microwave sensors separately, that means that the weights derived by the NN depend on the assimilated instrument with weights from training experiment sn9 for SMAP and training experiment an9 for AMSR2 (cf. Table 1). The assimilation experiments are run over the full year 2021, i.e., the period where the NN was tested and thus no prior knowledge of the observations is contained in the model equivalent.

LDAS-Monde is set up to work on a 24 h assimilation window and the observations are assimilated at a given hour. To account for the diurnal variability of the observations, we ran four separate assimilation cycles using 6-hour time bins around each of 0300UTC, 0900UTC, 1500UTC, and 2100UTC, which correspond to the model output times in our setup. This strategy allows for a balance between a relatively time efficient treatment of the assimilation experiments and a smoothing of the signal over one day due to the averaging. Nevertheless, the time constraints only allow for a limited number of assimilation experiments.

### 5.2 Impact of the observation error

We performed several assimilation experiments for each instrument to investigate the impact of, first, using the forcing variables as predictors and, second, of different magnitudes of observation errors (Table 2).

As mentioned in Sect. 3, we derive the initial observation errors $\sigma_o$ from the O-B time series with 30% of the average O-B difference, where the background B refers to the OL run here. When comparing the specified error with the one diagnosed with the method of Desroziers et al. (2005), it seems that larger values are necessary. Indeed, Muñoz-Sabater et al. (2018) found that the recommended observation error of 2.5-3 K was too small for SMOS observations, which also operate in L band, and found better results with twice that value. In model space, a range of 0.005-0.015 $m^3m^{-3}$ was found for SMAP soil moisture errors for croplands (Vural et al., 2021) dependent on the grid point and for AMSR2, observation errors in a range of 0-0.38 $m^3m^{-3}$ for China (Fan et al., 2022). These values cannot be compared directly but indicate that working with one value for the whole domain is not an optimal solution. Since the grid point-wise approach for the estimation of the observation errors is not feasible here, the different constant error values give only a rough estimate. Here, the results show that the larger errors diminish the impact on the analysis too much on a global scale.

The setups including forcing variables as predictors yield smaller initial observation errors, and, of all tested values, we found the errors around 0.5 K most suitable according to the impact on the assimilation system (Sect. 5.3) as well as on the verification (Sect. 5.4).

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

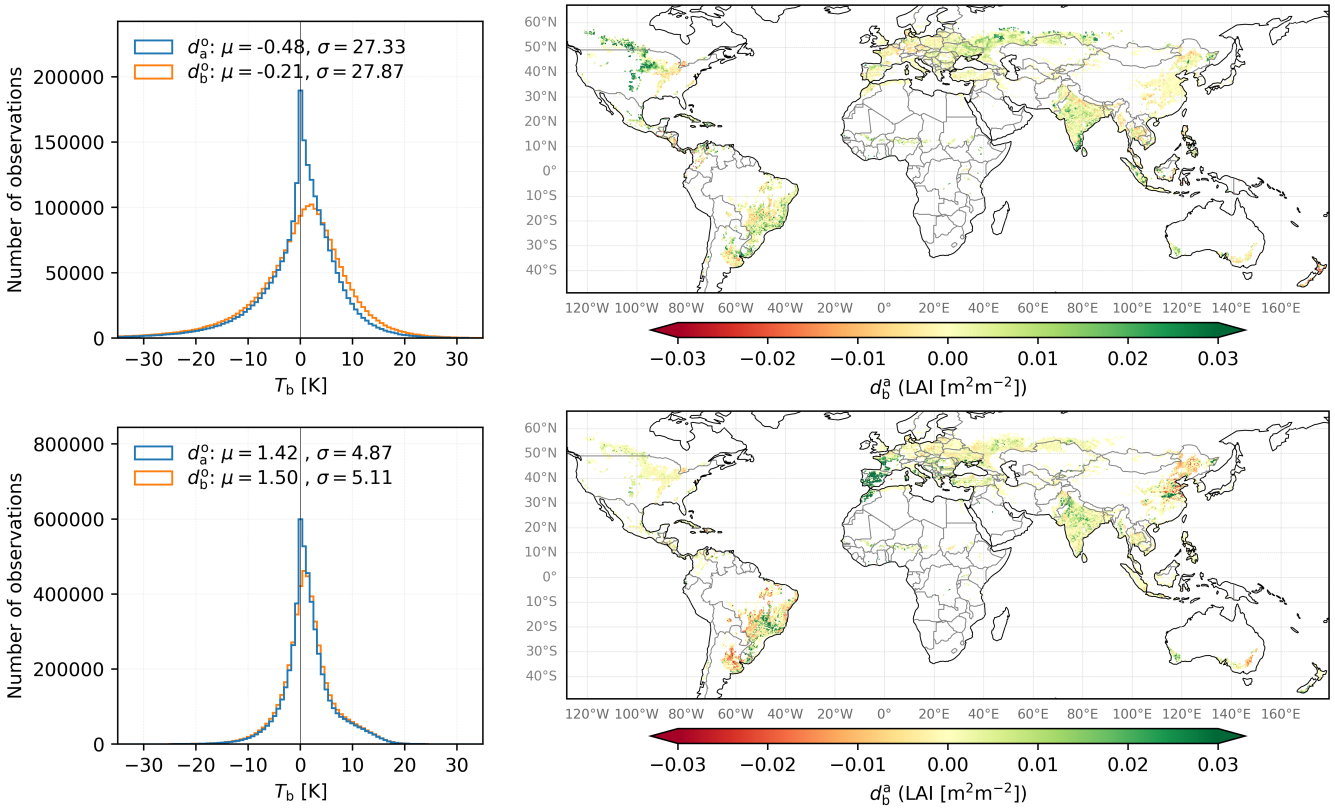


**Figure 7.** Left: Histograms of innovations ($d_b^o$, orange) and residuals ($d_a^o$, blue) for experiment s3 (top) and a3 (bottom). Right: Distribution of accumulated increments $d_b^a$ for experiments s3 (top) and a3 (bottom).

## 5.3 Departure statistics

To validate the assimilation of the introduced observations and the behaviour of the developed observation operator, we compare background and analysis departures ($d_b^o$ and $d_a^o$, respectively) of the respective experiment with each other. The system behaves as expected when the assimilation is able to bring the analysis departures closer to zero on average than the background departures. We confirm that this is the case for both SMAP and AMSR2 (Fig. 7, left, for s3 and a3), although the mean value of the distribution is drawn more to the negative for the SMAP experiment. The distribution of the departures is in general wider for SMAP than for AMSR2, which reflects the behaviour seen already in the evaluation of the NN in the testing period (Sect. 4) and was attributed to the larger variability of the L-band signal compared to the X band as well as the lower number of points remaining for the training after quality control. As a consequence, the assimilation is able to yield a larger improvement for SMAP (difference of mean $\Delta\mu \approx 0.21$ K, difference of STD $\Delta\sigma \approx 0.31$ K). The improvement for AMSR2 is less pronounced ($\Delta\mu \approx 0.04$ K, $\Delta\sigma \approx 0.14$ K) since the model equivalent is already quite close to the observations.

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

**Table 2.** Overview of the assimilation experiments conducted for SMAP and AMSR2. The table indicates the name of the assimilation experiment (DA exp. ), the usage of forcing variables as predictors (forcing), the exact NN experiment as described in Table 1 (NN exp.), and the observation error implemented (obs. err.).

| Instrument | DA exp. | forcing | NN exp. | obs. err. |
|---|---|---|---|---|
| | OL | - | | - |
| SMAP | s1 | - | sn6 | 1.3 |
| | s2 | - | sn6 | 8.0 |
| | s3 | x | sn9 | 0.5 |
| | s4 | x | sn9 | 1.0 |
| AMSR2 | a1 | - | an6 | 1.1 |
| | a2 | - | an6 | 4.0 |
| | a3 | x | an9 | 0.5 |
| | a4 | x | an9 | 1.1 |

The LAI increments range between -1 and 1 due to constraints in the assimilation system with a pronounced maximum around zero indicating as many positive as negative increments. Nevertheless, the spatial distribution (Fig. 7, right) shows that some areas are impacted more by the assimilation on average, namely Southwest Russia, Northern US, and Canada, where LAI is mostly added by the SMAP observations. In contrast, the AMSR2 observations remove LAI dominantly in Southern America and adds LAI in Western Europe, whereas Northeast China is marked by strong increments of both signs.

Soil moisture increments are on average rather small. They decrease more than increase for both instruments for the second and third soil layers, with more symmetric distributions in the deeper layers (not shown here). Apparently, not all removed soil water is distributed to the LAI but might also be subject to increased evaporation.

### 5.4 Verification against LAI observations

For verifying the impact of the assimilation, the analyses are compared against an independent observational data set for LAI (CGLS; Sect. 2.3) to evaluate a potential improvement compared to the OL run.

For both instruments, we obtain an improvement of the temporal evolution of the LAI compared to the OL run for almost all experiments on the global scale (Fig. 8) but, in general, the LAI analyses do not match the temporal evolution of the observations very well. This effect is mostly achieved by an improvement in bias as also noted by Albergel et al. (2020), both unbiased RMSE (ubRMSE) and correlation are not improved in most cases (cf. Fig. 9). Nevertheless, all experiments except s1 and s2 are able to improve the RMSE in most of the investigated regions compared to the OL (Fig. 9).

For experiment a3, the aggregated RMSE maps (Fig. 10, bottom) show a strong relative improvement over Australia, East Asia, and parts of South America, but a degradation of parts of Western Europe and Brazil. The rest of the globe exhibits a smaller mixed impact with a more negative impact in Ukraine and West Russia.

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

The spatial distribution for experiment s3 shows as well an improvement in Australia, in East Asia and Brazil as well as in India but a scattered degradation at high latitudes, especially in Canada. For the rest of the globe, we obtain mostly a mixed to positive impact, albeit with less amplitude than for a3.

Furthermore, we investigated the impact of the assimilation on specific areas especially affected by droughts and heat waves as identified by Albergel et al. (2020, Fig. 1) since detailed knowledge about the ambient conditions is crucial for disaster management.

As for the global time series, the temporal match for the different regions is determined by the OL bias and, thus, not very pronounced in the analyses. With regard to the overall improvement of the RMSE, the score card (Fig. 9a) illustrates the challenge to find one experiment that is suited to improve on all investigated areas. Whereas India (INDI), Northern China (NRCH), the Murray-Darling basin (MUDA), the Pampas (PAMP), and East Africa (EAFR) are better represented by all experiments, the representation of the LAI in West Russia (WRUS) seems to be difficult to improve by the employed observations independently of the experiment. This is in agreement with the lower performance seen in that area on spatial scale (Fig. 10).

Figures 9d,g illustrate that the RMSE improvements are mostly driven by improvements in the bias, which are obtained for almost all cases, whereas the ubRMSE is degraded mostly. The correlation is degraded almost everywhere (Fig. 9j).

Eventually, the shown evaluation is mostly in favour for the implementation of a predictor set making use of forcing variables, underlining the necessity to include more information on the diurnal cycle within the training process. The assessment of the observation errors remains a challenge, nevertheless, our assimilation experiments show that large errors dilute the impact of the observations onto the analysis too much (s2 and a2 in Fig. 9).

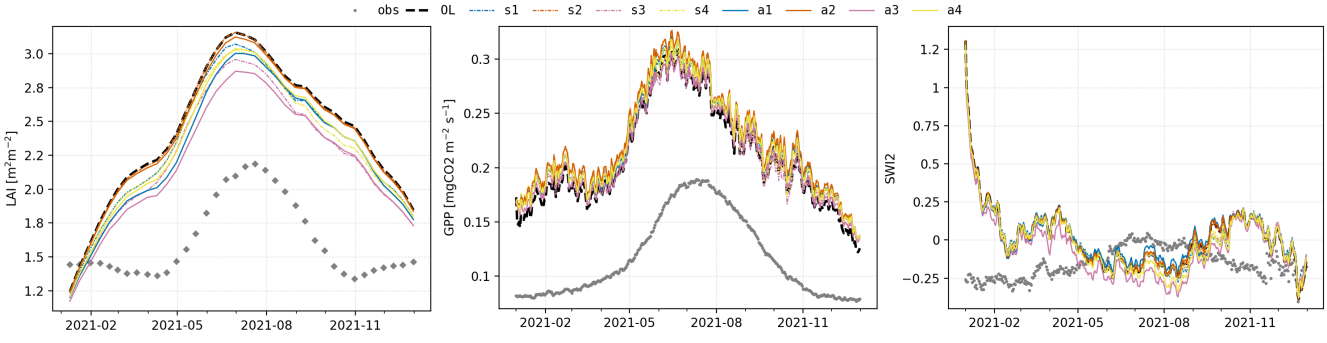


**Figure 8.** Comparison of time series of LAI (left), GPP (centre), and SWI anomalies (right) averaged over global croplands. Observations (gray points), OL (black dashed line), and DA experiments (SMAP: dash-dotted lines, AMSR2: solid lines) are shown in comparison.

### 5.5 Verification against GPP observations

Similar to Section 5.4, we verified the impact of the analyses against the GPP FLUXCOM-X (Sect. 2.4). Globally, the impact is not very pronounced, however, the assimilation moves the analyses further away from the GPP observations (Fig. 8, centre)

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

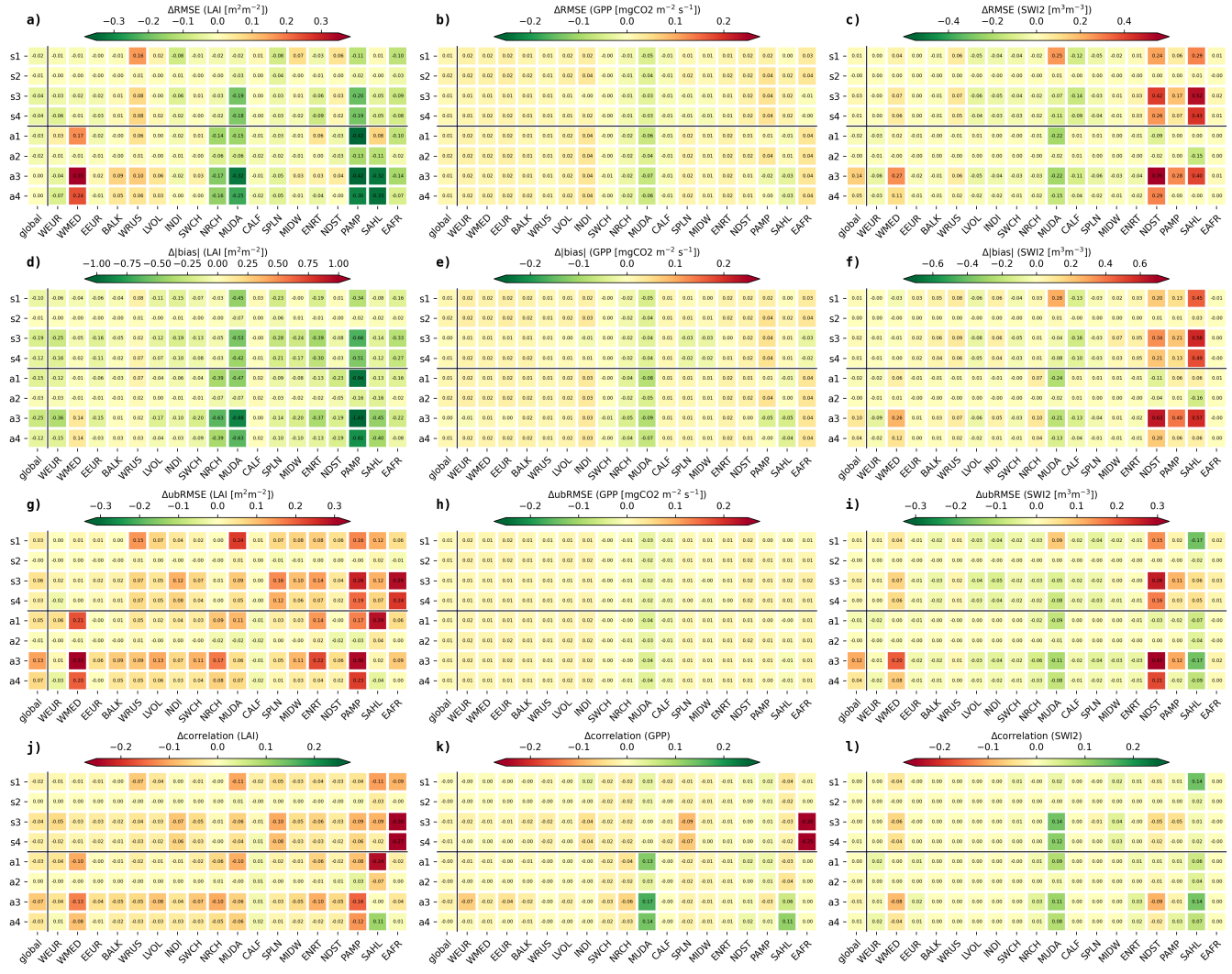


**Figure 9.** Score cards displaying the difference of RMSE (a,b,c), absolute bias (d,e,f), ubRMSE (g,h,i) and correlation (j,k,l) compared to the OL for LAI (a,d,g), GPP (b,e,h), and SWI2 anomaly (c,f,i). The metrics are computed between analysis/OL and observations for all investigated assimilation experiments (Table 2) for each of the regions prone to droughts (Albergel et al. 2020, Fig. 1). Green colours indicate an improvement compared to the OL, red a degradation. The boxes are labelled with the respective values.

for the whole time period of assimilation. The experiments do not exhibit large differences among each other but both s3 and a3 experiments remain closest to the OL run. Nevertheless, a positive impact on the RMSE can be observed for several regions, especially for NRCH and MUDA, but also for SWCH and CALF, for a3 and a4 additionally PAMP and SAHL, for s3 and s4 additionally SPLN and MIDW, and EAFR (Fig. 10). These areas are mostly those where a strong positive impact was noticed already in the RMSE of the LAI. Furthermore, as for LAI, most of the impact results from an improvement in bias (Fig. 9e),

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

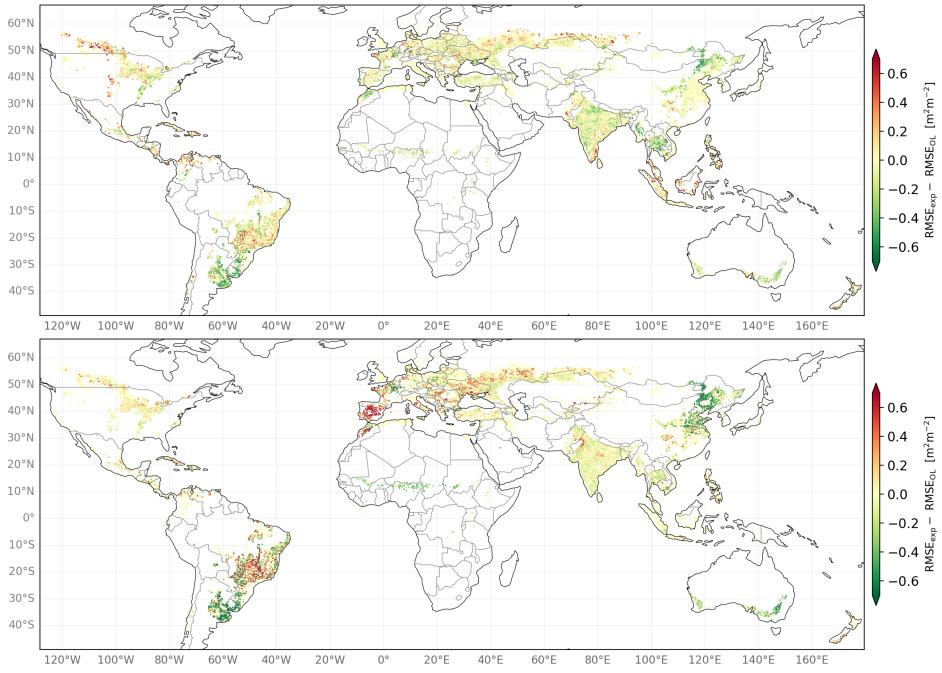


**Figure 10.** Verification against observed LAI. Difference between the RMSE of the LAI of the DA experiment and the OL for experiments s3 (top) and a3 (bottom).

a slight improvement in ubRMSE is obtained though for mainly the NRCH and MUDA regions (Fig. 9h). The impact on the correlation is slightly negative in most cases except for few experiments in the MUDA and SAHL regions (Fig. 9k).

### 5.6 Verification against SWI observations

Measured soil moisture and, hence, SWI cannot be compared accurately to model values due to the inherent representivity mismatch. As a consequence, we compute SWI anomalies (cf. Sect. 2.5) for the soil layer between 0.01 m - 0.04 m and compare those.

The impact on SWI anomalies is mostly negligible for the SMAP experiments, only a3 seems to have a global impact on the whole year (Fig. 8, right). The extreme mismatch of the time series of the OL run is obviously hard to correct. The initial improvement of a3 is not able to maintain over summer in the Northern hemisphere as negative increments are added all over the year. The impact on the single regions is mostly negligible as well with larger improvements for MUDA and a strong degradation for NDST and SAHL (Fig. 9c). As before, the changes in RMSE are mostly driven by the change in bias (Fig. 9f). The correlation is neutral for most regions with a tendency to a positive impact for especially MUDA and SAHL (Fig. 9i).

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

## 6 Discussion

Section 4 showed that an observation operator can be constructed in a rather simple fashion. Improvements could still be achieved by, e.g., combing different architectures of the NN with the different predictor sets or exploring other data sources. Nevertheless, our study showed that a good agreement between the model equivalent created by the NN and the observations can be achieved with a simple architecture. Eventually, the goal of an observation operator is not to perfectly match the observations as no correction would be achieved by data assimilation in that case.

We showed that static predictors are essential for our data sets, even more for AMSR2 than for SMAP (Fig. 4). It would be, however, be interesting to find a configuration with less need to train on information from static predictors to potentially overcome issues at the edges of the distribution. Other studies showed that the approach of a coordinate-independent NN for each grid point is able to achieve reasonable results as well on a small scale (Shan et al., 2022; Corchia et al., 2023).

The small biases found for the model equivalent (Sect. 4, Table 1) justified the assimilation without bias correction of the observations. Nevertheless, the main benefit in model space after assimilation seems the improvement of bias for the LAI as was shown in Sect. 5.4. This impacts the RMSE positively, whereas correlation and ubRMSE do not profit from the assimilation.

This behaviour can partially be attributed to our model run: Since the OL is subject to a large mismatch, which is likely due to a bias in the forcing data, it is consequently difficult for the assimilated observations to fully correct for this.

For both main experiments s3 and a3 for SMAP and AMSR2, respectively, a degradation for high latitudes was noticed. This seems similar to the findings of Corchia (2024) for the assimilation of both ASCAT and SMOS observations. This degradation might be related to the sensitivity of the observations to frozen soil and, possibly, the subsequent lack of training data in case the masking does not affect the whole of the period.

Eventually, the onset of vegetation growth is often not well captured by the model and a lack of it in the beginning of the season is challenging to correct by the observations. Biases introduced by ERA5 forcing were already investigated (e.g., Calvet et al. 2020; Liu et al. 2025) and are generally attributed to systematic errors in the cloud detection (Urraca et al., 2018). Furthermore, brightness temperatures are sensitive to the representation of soil temperature in the ISBA model and soil temperature biases (Calvet et al. 2020; Verger et al. 2025, Rojas-Munoz et al., in prep.).

The separate investigation of areas prone to droughts showed that both sensors have their strengths in different areas. This is expected as the different microwave bands have different characteristics. In terms of RMSE and bias, the assimilation often yields a better match with the LAI observation than the OL, especially AMSR2 is able to achieve a large improvement. On the other hand, large positive increments (Fig. 7) can also lead to stronger degradations as seen in the Mediterranean (WMED). Since the model equivalents (Fig. 2) are unremarkable, we suspect that the LAI added in the beginning of the year leads to an increased depletion of soil water and, therefore, an overestimated decrease of LAI in the dry season if the physical connections between the variables are not mapped properly by the observation operator. This might be as well the cause for the globally minor impact of the assimilation onto the GPP and the SWI anomalies and for the sometimes even opposite effect in correlation. A small impact of the assimilation of an LAI-proxy (vegetation optical depth) retrieved from SMAP and different microwave sensors including AMSR2 on SWI was already found by Kumar et al. (2020).

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

In general, we can assume that the NN-based forward operator is able to act on one or few variables but since no physics-based optimisation was applied, the model equivalent can seem adequate but still interfere within the assimilation process. This can be due to the possible neglect of inter-variable cross-correlations, which makes the information that is spread onto other variables inconsistent (Shan et al., 2024).

## 7 Conclusion

In this study, we probed the benefit of employing a neural network as an observation operator in LDAS-Monde to assimilate microwave brightness temperatures. The best configurations of the NNs were identified for both the SMAP and AMSR2 brightness-temperature observations, using several static predictors, LAI observations, as well as forcing and model variables. The setup of the hyperparameters turned out to be of minor importance, whereas a proper pre-processing of the data sets seems essential. In the end, most training experiments were able to reproduce the observations to a satisfactory degree with high correlations of up to 0.87 for SMAP and up to 0.93 for AMSR2.

The total RMSE of 8.2 K for SMAP exceeds the RMSE of 3.6 K for AMSR2. For this, we identified the following reasons: First, the generally lower variability of the X-band brightness temperatures compared to the L band due to their different sensitivities to canopy and soil surface and, second, the much larger amount of data remaining for AMSR2 after quality control, which can feed the NN with more values on the edges of the distributions. With regard to the employed predictors, surface variables seem more important for SMAP, whereas topography and forcing variables play a larger role for AMSR2.

The found performances of the NNs are suitable to produce a model equivalent for both instruments, leading to the subsequent implementation into our assimilation system LDAS-Monde, where the application of the found weights to the model background acts as the actual observation operator.

The assimilation experiments yielded the expected improvement of the residuals compared to the innovations. The improvement is larger for the SMAP than for AMSR2 data, which is likely a consequence of the better match between model equivalent and observations for AMSR2, giving less room for improvement.

When comparing the analyses against independent LAI observations, we obtain an improvement on global scale with generally stronger improvements for NN configurations using forcing variables as predictors and employing smaller observation errors. In general, AMSR2 matches the temporal evolution of the LAI better, whereas on a spatial scale, SMAP is able to outperform AMSR2 in dependence on the region. With regard to this behaviour, we studied areas especially prone to droughts and identified several regions that benefit strongly from the assimilation, mainly through a reduction in the bias. The verification against GPP and SWI anomaly yields a smaller and sometimes opposite effect on the performance though.

Overall, the results show the potential of assimilating brightness temperatures with a neural network to improve the state estimate of LSVs. Further tuning of the DA system might be required to achieve a more accurate match of the LAI observations, including the solution of model-inherent difficulties with the proper representation of land-surface processes. Eventually, a physics-based training might be able to overcome the observed inconsistencies between different LSVs.

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

*Code and data availability.* The AVHRR-LAI data can be downloaded at https://www.theia-land.fr/en/blog/product/series-of-vegetation-variables-avhrr/. CMLS data is available at https://land.copernicus.eu/en/products/vegetation/leaf-area-index-300m-v1.0 SMAP and AMSR2 data are available at https://nsidc.org/data/explore-data, and FLUXCOM-X at https://data.icos-cp.eu/portal/#{"filterKeywords"%3A["FLUXCOM"]}. The SURFEX code can be obtained following the instructions given at https://www.umr-cnrm.fr/surfex/spip.php?article415.

*Author contributions.* Jasmin Vural: Conceptualization, formal analysis, investigation, methodology, software, validation, visualization, writing (original draft preparation) Pierre Vanderbecken: Methodology, software, validation, writing (review and editing) Bertrand Bonan: Data curation, software, writing (review and editing) Oscar Rojas-Muñoz: Data curation, software, writing (review and editing) Jean-Christophe Calvet: Conceptualization, funding acquisition, project administration, writing (review and editing)

*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* We would like to acknowledge funding for the CORSO project (grant agreement No 101082194) by the European Union. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the Commission. Neither the European Union nor the granting authority can be held responsible for them.

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

## References

Albergel, C., Munier, S., Leroux, D. J., Dewaele, H., Fairbairn, D., Barbu, A. L., Gelati, E., Dorigo, W., Faroux, S., Meurey, C., Le Moigne, P., Decharme, B., Mahfouf, J.-F., and Calvet, J.-C.: Sequential Assimilation of Satellite-Derived Vegetation and Soil Moisture Products Using SURFEX_v8.0: LDAS-Monde Assessment over the Euro-Mediterranean Area, Geoscientific Model Development, 10, 3889–3912, https://doi.org/10.5194/gmd-10-3889-2017, 2017.

Albergel, C., Zheng, Y., Bonan, B., Dutra, E., Rodríguez-Fernández, N., Munier, S., Draper, C., De Rosnay, P., Muñoz-Sabater, J., Balsamo, G., Fairbairn, D., Meurey, C., and Calvet, J.-C.: Data Assimilation for Continuous Global Assessment of Severe Conditions over Terrestrial Surfaces, Hydrology and Earth System Sciences, 24, 4291–4316, https://doi.org/10.5194/hess-24-4291-2020, 2020.

Boone, A., Masson, V., Meyers, T., and Noilhan, J.: The Influence of the Inclusion of Soil Freezing on Simulations by a Soil–Vegetation–Atmosphere Transfer Scheme, Journal of Applied Meteorology, 39, 1544–1569, http://journals.ametsoc.org/doi/abs/10.1175/1520-0450(2000)039%3C1544:TIOTIO%3E2.0.CO%3B2, 2000.

Brodzik, M. J., Billingsley, B., Haran, T., Raup, B., and Savoie, M. H.: EASE-Grid 2.0: Incremental but Significant Improvements for Earth-Gridded Data Sets, ISPRS International Journal of Geo-Information, 1, 32–45, https://doi.org/10.3390/ijgi1010032, 2012.

Calvet, J.-C. and Champeaux, J.-L.: L'apport de La Télédétection Spatiale à La Modélisation Des Surfaces Continentales, La Météorologie, p. 052, https://doi.org/10.37053/lameteorologie-2020-0016, 2020.

Calvet, J.-C., Noilhan, J., Roujean, J.-L., Bessemoulin, P., Cabelguenne, M., Olioso, A., and Wigneron, J.-P.: An Interactive Vegetation SVAT Model Tested against Data from Six Contrasting Sites, Agricultural and Forest Meteorology, 92, 73–95, https://doi.org/10.1016/S0168-1923(98)00091-4, 1998.

Calvet, J.-C., Wigneron, J.-P., Walker, J., Karbou, F., Chanzy, A., and Albergel, C.: Sensitivity of Passive Microwave Observations to Soil Moisture and Vegetation Water Content: L-Band to W-Band, IEEE Transactions on Geoscience and Remote Sensing, 49, 1190–1199, https://doi.org/10.1109/TGRS.2010.2050488, 2011.

Calvet, J.-C., Albergel, C., Bonan, B., Meurey, C., and Zheng, Y.: Copernicus Global Land Operations - Scientific Quality Evaluation Cross-Cutting Consistency - January-December 2019, Tech. rep., CGLOPS Lot 1 consortium, https://meteofrance.hal.science/meteo-04168132, 2020.

Chan, S., Njoku, E., and Colliander, A.: ATBD SMAP L1C Radiometer Data Product, 2015.

Chan, S., Njoku, E. G., and Colliander, A.: SMAP L1C Radiometer Half-Orbit 36 Km EASE-Grid Brightness Temperatures, Version 5, https://doi.org/10.5067/JJ5FL7FRLKJI, 2020.

Copernicus Land Monitoring Service and Copernicus Land Monitoring Service Helpdesk: Leaf Area Index 2014-Present (Raster 300 m), Global, 10-Daily - Version 1, https://doi.org/10.2909/219FDC9F-616B-444B-A495-198F527B4722, 2017a.

Copernicus Land Monitoring Service and Copernicus Land Monitoring Service Helpdesk: Soil Water Index 2007-Present (Raster 12.5 Km), Global, Daily - Version 3, https://doi.org/10.2909/98C5E00E-3580-4BB3-9509-50A572B1E935, 2017b.

Corchia, T.: Apport de l'apprentissage automatique pour l'intégration d'observations satellitaires dans un modèle global du système sol-plante, 2024.

Corchia, T., Bonan, B., Rodríguez-Fernández, N., Colas, G., and Calvet, J.-C.: Assimilation of ASCAT Radar Backscatter Coefficients over Southwestern France, Remote Sensing, 15, 4258, https://doi.org/10.3390/rs15174258, 2023.

Das, K. and Paul, P. K.: Present Status of Soil Moisture Estimation by Microwave Remote Sensing, Cogent Geoscience, 1, 1084 669, https://doi.org/10.1080/23312041.2015.1084669, 2015.

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

De Gélis, I., Prigent, C., Jimenez, C., and Sandells, M.: Forward Modelling of Passive Microwave Emissivities over Snow-Covered Areas at Continental Scale, Remote Sensing of Environment, 328, 114 821, https://doi.org/10.1016/j.rse.2025.114821, 2025.

De Lannoy, G. J. M., Reichle, R. H., and Pauwels, V. R. N.: Global Calibration of the GEOS-5 L-Band Microwave Radiative Transfer Model over Nonfrozen Land Using SMOS Observations, Journal of Hydrometeorology, 14, 765–785, https://doi.org/10.1175/JHM-D-12-092.1, 2013.

Decharme, B., Boone, A., Delire, C., and Noilhan, J.: Local Evaluation of the Interaction between Soil Biosphere Atmosphere Soil Multilayer Diffusion Scheme Using Four Pedotransfer Functions, Journal of Geophysical Research, 116, https://doi.org/10.1029/2011JD016002, 2011.

Dente, L., Satalino, G., Mattia, F., and Rinaldi, M.: Assimilation of Leaf Area Index Derived from ASAR and MERIS Data into CERES-Wheat Model to Map Wheat Yield, Remote Sensing of Environment, 112, 1395–1407, https://doi.org/10.1016/j.rse.2007.05.023, 2008.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of Observation, Background and Analysis-Error Statistics in Observation Space, Quarterly Journal of the Royal Meteorological Society, 131, 3385–3396, https://doi.org/10.1256/qj.05.108, 2005.

Draper, C. S., Reichle, R. H., De Lannoy, G. J. M., and Liu, Q.: Assimilation of Passive and Active Microwave Soil Moisture Retrievals: ASCAT AMSR-E ASSIMILATION, Geophysical Research Letters, 39, n/a–n/a, https://doi.org/10.1029/2011GL050655, 2012.

Fan, J., Luo, M., Han, Q., Liu, F., Huang, W., and Tan, S.: Evaluation of SMOS, SMAP, AMSR2 and FY-3C Soil Moisture Products over China, PLOS ONE, 17, e0266 091, https://doi.org/10.1371/journal.pone.0266091, 2022.

Hess, R.: Assimilation of Screen-Level Observations by Variational Soil Moisture Analysis, Meteorology and Atmospheric Physics, 77, 145–154, http://www.springerlink.com/index/PX3B7NTN0PU74TR9.pdf, 2001.

Jackson, T., Chan, S., Bindlish, R., and Njoku, E.: AMSR-E/AMSR2 Unified L2B Half-Orbit 25 Km EASE-Grid Surface Soil Moisture, Version 1, https://doi.org/10.5067/IKQ0G7ODMLC7, 2016.

Jacobs, C., van den Hurk, B., and de Bruin, H.: Stomatal Behaviour and Photosynthetic Rate of Unstressed Grapevines in Semi-Arid Conditions, Agricultural and Forest Meteorology, 80, 111–134, https://doi.org/10.1016/0168-1923(95)02295-3, 1996.

Jarlan, L., Balsamo, G., Lafont, S., Beljaars, A., Calvet, J. C., and Mougin, E.: Analysis of Leaf Area Index in the ECMWF Land Surface Model and Impact on Latent Heat and Carbon Fluxes: Application to West Africa, Journal of Geophysical Research: Atmospheres, 113, 2007JD009 370, https://doi.org/10.1029/2007JD009370, 2008.

Kidd, C.: Radio Frequency Interference at Passive Microwave Earth Observation Frequencies, International Journal of Remote Sensing, 27, 3853–3865, https://doi.org/10.1080/01431160600702400, 2006.

Kolassa, J., Reichle, R., and Draper, C.: Merging Active and Passive Microwave Observations in Soil Moisture Data Assimilation, Remote Sensing of Environment, 191, 117–130, https://doi.org/10.1016/j.rse.2017.01.015, 2017.

Kolassa, J., Ganeshan, M., McGrath-Spangler, E. L., Reale, O., Reichle, R. H., and Zhang, S. Q.: Impact of Assimilating SMAP Observations over Land on Tropical Cyclone Representation: A Case Study of Tropical Cyclone Idai, 2025.

Kumar, S. V., Holmes, T. R., Bindlish, R., De Jeu, R., and Peters-Lidard, C.: Assimilation of Vegetation Optical Depth Retrievals from Passive Microwave Radiometry, Hydrology and Earth System Sciences, 24, 3431–3450, https://doi.org/10.5194/hess-24-3431-2020, 2020.

Lievens, H., Martens, B., Verhoest, N., Hahn, S., Reichle, R., and Miralles, D.: Assimilation of Global Radar Backscatter and Radiometer Brightness Temperature Observations to Improve Soil Moisture and Land Evaporation Estimates, Remote Sensing of Environment, 189, 194–210, https://doi.org/10.1016/j.rse.2016.11.022, 2017.

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

Liu, E., Zhu, Y., Lü, H., Bonan, B., Munier, S., and Calvet, J.-C.: Impact of Leaf Area Index Assimilation and Gauge-Corrected Precipitation on Land Surface Variables in LDAS-Monde: A Case Study over China, Journal of Hydrology, 659, 133 304, https://doi.org/10.1016/j.jhydrol.2025.133304, 2025.

Lundberg, S. M. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions.

Mahfouf, J.-F., Bergaoui, K., Draper, C., Bouyssel, F., Taillefer, F., and Taseva, L.: A Comparison of Two Off-Line Soil Analysis Schemes for Assimilation of Screen Level Observations, Journal of Geophysical Research (Atmospheres), 114, D08 105, https://doi.org/10.1029/2008JD011077, 2009.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gibelin, A.-L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J.-F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V., and Voldoire, A.: The SURFEXv7.2 Land and Ocean Surface Platform for Coupled or Offline Simulation of Earth Surface Variables and Fluxes, Geoscientific Model Development, 6, 929–960, https://doi.org/10.5194/gmd-6-929-2013, 2013.

Muñoz-Sabater, J., De Rosnay, P., Albergel, C., and Isaksen, L.: Sensitivity of Soil Moisture Analyses to Contrasting Background and Observation Error Scenarios, Water, 10, 890, https://doi.org/10.3390/w10070890, 2018.

Mucia, A., Bonan, B., Albergel, C., Zheng, Y., and Calvet, J.-C.: Assimilation of Passive Microwave Vegetation Optical Depth in LDAS-Monde: A Case Study over the Continental USA, Biogeosciences, 19, 2557–2581, https://doi.org/10.5194/bg-19-2557-2022, 2022.

Nelson, J. A., Walther, S., Gans, F., Kraft, B., Weber, U., Novick, K., Buchmann, N., Migliavacca, M., Wohlfahrt, G., Šigut, L., Ibrom, A., Papale, D., Göckede, M., Duveiller, G., Knohl, A., Hörtnagl, L., Scott, R. L., Dušek, J., Zhang, W., Hamdi, Z. M., Reichstein, M., Aranda-Barranco, S., Ardö, J., Op De Beeck, M., Billesbach, D., Bowling, D., Bracho, R., Brümmer, C., Camps-Valls, G., Chen, S., Cleverly, J. R., Desai, A., Dong, G., El-Madany, T. S., Euskirchen, E. S., Feigenwinter, I., Galvagno, M., Gerosa, G. A., Gielen, B., Goded, I., Goslee, S., Gough, C. M., Heinesch, B., Ichii, K., Jackowicz-Korczynski, M. A., Klosterhalfen, A., Knox, S., Kobayashi, H., Kohonen, K.-M., Korkiakoski, M., Mammarella, I., Gharun, M., Marzuoli, R., Matamala, R., Metzger, S., Montagnani, L., Nicolini, G., O'Halloran, T., Ourcival, J.-M., Peichl, M., Pendall, E., Ruiz Reverter, B., Roland, M., Sabbatini, S., Sachs, T., Schmidt, M., Schwalm, C. R., Shekhar, A., Silberstein, R., Silveira, M. L., Spano, D., Tagesson, T., Tramontana, G., Trotta, C., Turco, F., Vesala, T., Vincke, C., Vitale, D., Vivoni, E. R., Wang, Y., Woodgate, W., Yepez, E. A., Zhang, J., Zona, D., and Jung, M.: X-BASE: The First Terrestrial Carbon and Water Flux Products from an Extended Data-Driven Scaling Framework, FLUXCOM-X, Biogeosciences, 21, 5079–5115, https://doi.org/10.5194/bg-21-5079-2024, 2024.

Noilhan, J. and Mahfouf, J.-F.: The ISBA Land Surface Parameterisation Scheme, Global and planetary Change, 13, 145–159, http://www.sciencedirect.com/science/article/pii/0921818195000437, 1996.

Pauwels, V. R. N., Verhoest, N. E. C., De Lannoy, G. J. M., Guissard, V., Lucau, C., and Defourny, P.: Optimization of a Coupled Hydrology-Crop Growth Model through the Assimilation of Observed Soil Moisture and Leaf Area Index Values Using an Ensemble Kalman Filter: ASSIMILATION OF LAI AND SOIL MOISTURE, Water Resources Research, 43, https://doi.org/10.1029/2006WR004942, 2007.

Reichle, R. H.: Global Assimilation of Satellite Surface Soil Moisture Retrievals into the NASA Catchment Land Surface Model, Geophysical Research Letters, 32, L02 404, https://doi.org/10.1029/2004GL021700, 2005.

Reichle, R. H., Koster, R. D., Dong, J., and Berg, A. A.: Global Soil Moisture from Satellite Observations, Land Surface Models, and Ground Data: Implications for Data Assimilation, Journal of Hydrometeorology, 5, 430–442, https://doi.org/10.1175/1525-7541(2004)005<0430:GSMFSO>2.0.CO;2, 2004.

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

Reichle, R. H., Zhang, S. Q., Liu, Q., Draper, C. S., Kolassa, J., and Todling, R.: Assimilation of SMAP Brightness Temperature Observations in the GEOS Land–Atmosphere Data Assimilation System, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 10 628–10 643, https://doi.org/10.1109/JSTARS.2021.3118595, 2021.

Reichle, R. H., Zhang, S. Q., Kolassa, J., Liu, Q., and Todling, R.: A Weakly Coupled Land Surface Analysis with SMAP Radiance Assimilation Improves GEOS Medium-range Forecasts of Near-surface Air Temperature and Humidity, Quarterly Journal of the Royal Meteorological Society, 149, 1867–1889, https://doi.org/10.1002/qj.4486, 2023.

Rochford, P. A.: SkillMetrics: A Python Package for Calculating the Skill of Model Predictions against Observations, https://github.com/PeterRochford/SkillMetrics, 2016.

Rodríguez-Fernández, N., Kerr, Y., Van Der Schalie, R., Al-Yaari, A., Wigneron, J.-P., De Jeu, R., Richaume, P., Dutra, E., Mialon, A., and Drusch, M.: Long Term Global Surface Soil Moisture Fields Using an SMOS-Trained Neural Network Applied to AMSR-E Data, Remote Sensing, 8, 959, https://doi.org/10.3390/rs8110959, 2016.

Rodríguez-Fernández, N., de Rosnay, P., Albergel, C., Richaume, P., Aires, F., Prigent, C., and Kerr, Y.: SMOS Neural Network Soil Moisture Data Assimilation in a Land Surface Model and Atmospheric Impact, Remote Sensing, 11, 1334, https://doi.org/10.3390/rs11111334, 2019.

Shan, X., Steele-Dunne, S., Huber, M., Hahn, S., Wagner, W., Bonan, B., Albergel, C., Calvet, J.-C., Ku, O., and Georgievska, S.: Towards Constraining Soil and Vegetation Dynamics in Land Surface Models: Modeling ASCAT Backscatter Incidence-Angle Dependence with a Deep Neural Network, Remote Sensing of Environment, 279, 113 116, https://doi.org/10.1016/j.rse.2022.113116, 2022.

Shan, X., Steele-Dunne, S., Hahn, S., Wagner, W., Bonan, B., Albergel, C., Calvet, J.-C., and Ku, O.: Assimilating ASCAT Normalized Backscatter and Slope into the Land Surface Model ISBA-A-gs Using a Deep Neural Network as the Observation Operator: Case Studies at ISMN Stations in Western Europe, Remote Sensing of Environment, 308, 114 167, https://doi.org/10.1016/j.rse.2024.114167, 2024.

Taylor, K. E.: Summarizing Multiple Aspects of Model Performance in a Single Diagram, Journal of Geophysical Research: Atmospheres, 106, 7183–7192, https://doi.org/10.1029/2000JD900719, 2001.

Urraca, R., Huld, T., Gracia-Amillo, A., Martinez-de-Pison, F. J., Kaspar, F., and Sanz-Garcia, A.: Evaluation of Global Horizontal Irradiance Estimates from ERA5 and COSMO-REA6 Reanalyses Using Ground and Satellite-Based Data, Solar Energy, 164, 339–354, https://doi.org/10.1016/j.solener.2018.02.059, 2018.

Verger, A.: Organization Name of Lead Contractor for This Deliverable: CSIC, 2024.

Verger, A., Weiss, M., and Baret, F.: GEOV2-AVHRR: Continuous and Consistent Time Series of Global Leaf Area Index and Fraction Absorbed PAR from 1981 to 2022, Remote Sensing of Environment, 331, 115 029, https://doi.org/10.1016/j.rse.2025.115029, 2025.

Vural, J., Schneider, S., Bauer-Marschallinger, B., and Haslinger, K.: Assimilation of the SCATSAR-SWI with SURFEX: Impact of Local Observation Errors in Austria, Monthly Weather Review, 149, 773–791, https://doi.org/10.1175/MWR-D-20-0160.1, 2021.

Vural, J., Merker, C., Löffler, M., Leuenberger, D., Schraff, C., Stiller, O., Schomburg, A., Knist, C., Haefele, A., and Hervo, M.: Improving the Representation of the Atmospheric Boundary Layer by Direct Assimilation of Ground-based Microwave Radiometer Observations, Quarterly Journal of the Royal Meteorological Society, p. qj.4634, https://doi.org/10.1002/qj.4634, 2023.

Wagner, W., Lemoine, G., Borgeaud, M., and Rott, H.: A Study of Vegetation Cover Effects on ERS Scatterometer Data, IEEE Transactions on Geoscience and Remote Sensing, 37, 938–948, http://ieeexplore.ieee.org/abstract/document/752212/, 1999.

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa-Saldaña, J., de Rosnay, P., Jann, A., Schneider, S., Komma, J., Kubu, G., Brugger, K., Aubrecht, C., Züger, J., Gangkofner, U., Kienberger, S., Brocca, L., Wang, Y., Blöschl, G., Eitzinger, J., and

https://doi.org/10.5194/egusphere-2026-838
Preprint. Discussion started: 25 February 2026
© Author(s) 2026. CC BY 4.0 License.

Steinnocher, K.: The ASCAT Soil Moisture Product: A Review of Its Specifications, Validation Results, and Emerging Applications, Meteorologische Zeitschrift, 22, 5–33, https://doi.org/10.1127/0941-2948/2013/0399, 2013.

Walker, J. P. and Houser, P. R.: A Methodology for Initializing Soil Moisture in a Global Climate Model: Assimilation of near-Surface Soil Moisture Observations, Journal of Geophysical Research: Atmospheres, 106, 11 761–11 774, https://doi.org/10.1029/2001JD900149, 2001.

Xue, Y. and Forman, B. A.: Comparison of Passive Microwave Brightness Temperature Prediction Sensitivities over Snow-Covered Land in North America Using Machine Learning Algorithms and the Advanced Microwave Scanning Radiometer, Remote Sensing of Environment, 170, 153–165, https://doi.org/10.1016/j.rse.2015.09.009, 2015.