CO2MVS RESEARCH ON SUPPLEMENTARY OBSERVATIONS

D4.4: Report on multivariate coupled land data assimilation impact on water and carbon cycle representation in the IFS

Due date of deliverable	December 2025
Submission date	November 2025
File Name	CORSO-D4.4-V1.1
Work Package /Task	WP4
Organisation Responsible of Deliverable	ECMWF
Author name(s)	Garrigues, S., P. de Rosnay, P. Weston, C. Rüdiger, D. Fairbairn, E. Pinnington, S. Boussetta, A. Agusti-Panareda, R. Engelen, J.-C. Calvet
Revision number	V2.02
Status	Final
Dissemination Level / location	PUBLIC www.corso-project.eu

The CORSO project (grant agreement No 101082194) is funded by the European Union.
Views and opinions expressed are however those of the author(s) 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.

1 Executive Summary

The objective of this work is to assimilate active microwave and Solar Induced Fluorescence (SIF) satellite observations in the Integrated Forecasting System (IFS) to update Leaf Area Index (LAI) and assess the impact on the IFS predictions of the Gross Primary Production (GPP), soil moisture and low-level meteorological fields (2m temperature and humidity). This work relies on the development of machine learning (ML)-based observation operators for (1) the normalized backscatter at 40° from the ASCAT (Advanced Scatterometer) instrument onboard METOP-B and C satellites and (2) SIF derived from TROPOMI (TROPOspheric Monitoring Instrument) onboard Copernicus Sentinel-5p satellite, which were described in the report D4-2 (December 2024). For SIF, the data assimilation was performed in the offline Land Data Assimilation (LDAS) system. The updated LAI was used to initialise IFS coupled model forecasts and the impacts on the IFS forecast of GPP, soil moisture, and low-level meteorological fields were evaluated. For ASCAT backscatter, the assimilation was directly conducted in the coupled LDAS in the IFS and the impacts of assimilating ASCAT backscatter on the IFS forecast performances were assessed against the operational configuration which relies on the assimilation of ASCAT soil moisture retrieval.

Table of Contents

2 Introduction

2.1 Background

To enable the European Union (EU) to move towards a low-carbon economy and implement its commitments under the Paris Agreement, a binding target was set to cut emissions in the EU by at least 40% below 1990 levels by 2030. European Commission (EC) President von der Leyen committed to deepen this target to at least 55% reduction by 2030. This was further consolidated with the release of the Commission's European Green Deal on the 11th of December 2019, setting the targets for the European environment, economy, and society to reach zero net emissions of greenhouse gases in 2050, outlining all needed technological and societal transformations that are aiming at combining prosperity and sustainability. To support EU countries in achieving the targets, the EU and European Commission (EC) recognised the need for an objective way to monitor anthropogenic CO_2 emissions and their evolution over time.

Such a monitoring capacity will deliver consistent and reliable information to support informed policy- and decision-making processes, both at national and European level. To maintain independence in this domain, it is seen as critical that the EU establishes an observation-based operational anthropogenic CO_2 emissions Monitoring and Verification Support (MVS) (CO2MVS) capacity as part of its Copernicus Earth Observation programme.

The CORSO research and innovation project will build on and complement the work of previous projects such as CHE (the CO2 Human Emissions), and CoCO2 (Copernicus CO2 service) projects, both led by ECMWF. These projects have already started the ramping-up of the CO2MVS prototype systems, so it can be implemented within the Copernicus Atmosphere Monitoring Service (CAMS) with the aim to be operational by 2026. The CORSO project will further support establishing the new CO2MVS addressing specific research & development questions.

The main objectives of CORSO are to deliver further research activities and outcomes with a focus on the use of supplementary observations, i.e., of co-emitted species as well as the use of auxiliary observations to better separate fossil fuel emissions from the other sources of atmospheric CO_2. CORSO will deliver improved estimates of emission factors/ratios and their uncertainties as well as the capabilities at global and local scale to optimally use observations of co-emitted species to better estimate anthropogenic CO_2 emissions. CORSO will also provide clear recommendations to CAMS, ICOS, and WMO about the potential added-value of high-temporal resolution $^{14}CO_2$ and APO observations as tracers for anthropogenic emissions in both global and regional scale inversions and develop coupled land-atmosphere data assimilation in the global CO2MVS system constraining carbon cycle variables with satellite observations of soil moisture, LAI, SIF, and Biomass. Finally, CORSO will provide specific recommendations for the topics above for the operational implementation of the CO2MVS within the Copernicus programme.

2.2 Scope of this deliverable

2.2.1 Objectives of this deliverable

The objective of this work is to evaluate the impact of the assimilation of active microwave and Solar Induced Fluorescence (SIF) satellite observations on the Integrated Forecast System (IFS) prediction of the carbon and water fluxes and low-level meteorological fields.

2.2.2 Work performed in this deliverable

We implemented the assimilation of (1) the normalized backscatter at 40° from the ASCAT (Advanced Scatterometer) instrument onboard METOP-B and C satellites in the coupled land data assimilation system (LDAS) in the IFS and (2) Solar Induced Fluorescence (SIF) derived from TROPOMI (TROPOspheric Monitoring Instrument) onboard Copernicus Sentinel-5p satellite, in the offline LDAS. This work relies on the development of machine learning (ML)-based observation operators for both ASCAT backscatter and SIF which were described in the report D4-2 (December 2024). The SIF data assimilation developments were first presented in the report D4-3 (December 2024) as well as in Garrigues et al. 2025a, 2025b. This report presents the implementation of both SIF and ASCAT data assimilation in the offline and coupled, respectively, LDAS. For the SIF data assimilation configuration, the updated LAI was used to initialise IFS coupled model forecasts and the impacts on the IFS forecast of Gross Primary Production (GPP), soil moisture, and low-level meteorological fields (2 m temperature and humidity) were evaluated. For ASCAT, the impacts of assimilating ASCAT backscatter on the IFS forecast performances were assessed against the operational configuration which relies on the assimilation of ASCAT soil moisture retrieval.

The SIF data assimilation methodology and the associated results are also presented in Garrigues et al. (2025a,b).

2.2.3 Deviations and counter measures

None reported.

2.3 Project partners:

Partners	
EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS	ECMWF
AKADEMIA GORNICZO-HUTNICZA IM. STANISLAWA STASZICA W KRAKOWIE	AGH
BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION	BSC
COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES	CEA
KAMINSKI THOMAS HERBERT	iLab
METEO-FRANCE	MF
NEDERLANDSE ORGANISATIE VOOR TOEGEPAST NATUURWETENSCHAPPELIJK ONDERZOEK TNO	TNO
RIJKSUNIVERSITEIT GRONINGEN	RUG
RUPRECHT-KARLS-UNIVERSITAET HEIDELBERG	UHEI
LUNDS UNIVERSITET	ULUND
UNIVERSITE PAUL SABATIER TOULOUSE III	UT3-CNRS
WAGENINGEN UNIVERSITY	WU
EIDGENOSSISCHE MATERIALPRUFUNGS- UND FORSCHUNGSANSTALT	EMPA
EIDGENOESSISCHE TECHNISCHE HOCHSCHULE ZUERICH	ETHZ
UNIVERSITY OF BRISTOL	UNIVBRIS
THE UNIVERSITY OF EDINBURGH	UEDIN

3 Data

3.1 ASCAT backscatter dataset

The Advanced Scatterometer (ASCAT) data consist of C-band radar backscatters (sigma0). In this work, we use the ASCAT sigma0 normalised at an incidence angle of 40° which is available from the EUMETSAT HSAF service. Digital Object Identifier (DOI) is: https://doi.org/10.15770/EUM_SAF_H_0009. Four years of ASCAT covering the period 2016 to 2019 was used in this work to develop a ML-based observation operator for assimilating sigma0 in the IFS. Frozen soil, water bodies, snow areas and mountain areas were excluded from the observation database.

3.2 TROPOSIF dataset

The TROPOSIF dataset (Guanter et al, 2021), which consists of global SIF observations from TROPOMI (TROPOspheric Monitoring Instrument) on board Sentinel-5P satellite, was used for the period from 2019 to 2022. The data with view and solar zenith angles above 60° and 70°, respectively, were filtered out. We used here SIF retrievals gridded at 0.1° spatial resolution and temporally averaged over 8 days to enhance the signal-to-noise ratio of SIF and to minimize the uncertainties caused by geometric registration errors.

3.3 LAI satellite-based dataset

The Copernicus Land Monitoring Service (CLMS) LAI dataset was used to evaluate the updated LAI from the land data assimilation system (CLMS, 2025). The CLMS LAI was estimated from the surface reflectances derived from the VEGETATION instrument on board PROBA-V before August 2020 and from the Ocean and Land Colour Instrument (OLCI) on board Sentinel-3 since September 2020 (CLMS, 2025). We used the RT1 version of LAI CLMS. For this study, the dataset was resampled at 0.1° spatial resolution and 8-day temporal frequency.

3.4 GPP satellite-based dataset

The IFS forecasts of GPP were assessed against the FluxSat GPP dataset, version 2.0 (Joiner and Yoshida 2021). FluxSat GPP is retrieved from MODIS reflectances using a neural network trained on GPP FLUXNET measurements. In this work, the FluxSat daily product was resampled at 8-day temporal frequency and 0.1° resolution spatial resolution.

3.5 In situ soil moisture measurements

The evaluation of the soil moisture analysis and forecast was performed using various in situ measurements from distinct measurement networks worldwide (e.g. SMOSMANIA in France, Albergel et al, 2011; USCRN in the US, Bell et al., 2013). These datasets are described in Fairbairn et al. (2019).

4 Methods

4.1 IFS land model and the land data assimilation system (LDAS)

4.1.1 The ecLand IFS land model

This work is based on the IFS cycle 49r1 and its land model ecLand (Boussetta *et al.*, 2021). ecLand comprises 4 soil layers of 0.07 m, 0.21 m, 0.72 m and 1.89 m thicknesses. Soil moisture is analysed in the first three soil layers which represent the root-zone of 1m depth. EcLand employs the Faquhar photosynthesis model. The vegetation classes are derived from the European Space Agency-Climate Change Initiative Land Use/Land Cover (ESA CCI) dataset. ecLand relies on a satellite-derived monthly mean LAI climatology (Boussetta *et al.*, 2021; Boussetta and Balsamo, 2021). To distinguish between low (grassland, crop,

shrubland) and high (forest) vegetation types, the ecLand LAI is disaggregated into a low and high LAI for each grid cell.

4.1.2 The IFS Land data assimilations system (LDAS)

The IFS LDAS is composed of several components for the low-level meteorological parameters (2-m temperature and relative humidity), snow, and soil moisture, soil temperature, and snow temperature (de Rosnay et al., 2022). The screen level analysis and the snow analysis are conducted using a 2D-OI (2-Dimensional Optimal Interpolation). The soil temperature and snow temperature analyses are conducted using a 1D-OI. The soil moisture analysis is conducted using a simplified Extended Kalman Filter (SEKF) approach. The LDAS runs twice per day using 12-hour data assimilation windows. It can be used in a coupled (i.e. online) mode in the IFS where it is weakly coupled with the atmospheric analysis or in an offline mode forced by ERA5 reanalysis or operational meteorology. In its operational configuration, the SEKF approach updates the volumetric soil moisture of the first three soil layers of ecLand by assimilating the aforementioned pseudo-observations of 2-metre temperature and relative humidity, in combination with surface soil moisture retrievals from ASCAT. A Cumulative Distribution Function (CDF) matching is applied to the ASCAT soil moisture product prior to its assimilation. In the coupled data assimilation configuration, SMOS soil moisture products are also assimilated.

In this work, the assimilation of SIF or ASCAT are achieved using the SEKF. The analysis is calculated using the following Kalman filter equations:

$$\mathbf{x_a} = \mathbf{x_b} + \mathbf{K}\,(\mathbf{y} - H(\mathbf{x_b})) \quad (1)$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}\,(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}})^{-1} \quad (2)$$

where $\mathbf{x_a}$ is the analysis, $\mathbf{x_b}$ is the model background, $\mathbf{y}$ are the observations, H is the observation operator, $\mathbf{H}$ contains the Jacobians to link the model variables to the observed variables, $\mathbf{B}$ is the background error covariance matrix and $\mathbf{R}$ is the observation error covariance matrix, respectively. $\mathbf{K}$ is the Kalman gain and the innovation term ($\mathbf{y} - H(x_b)$) represents the departure between the observation and the model background in the observation space.

4.2 Assimilation of SIF in the offline LDAS

4.2.1 SIF ML observation operator

A ML-based observation operator was developed for assimilating SIF satellite observations at 0.1° spatial resolution and 8-day temporal resolution to update the LAI climatology in the offline LDAS. It relies on the assumption that at these spatial and temporal resolutions, the SIF signal is mainly driven by LAI for most vegetation types.

The ML model is based on the eXtreme Gradient Boosting (XGBoost) technique and was trained using the Copernicus Land Monitoring Service (CLMS) satellite LAI, week of the year, latitude and longitude as predictors. The ML model was trained over the 2019-2020 period, the hyperparameters were tuned over 2021 and the model was evaluated over 2022. Barren and desertic areas, orographic areas, snow and frozen soil areas and water bodies, for which the SIF signal is too uncertain or not meaningful, were removed from the training dataset.

4.2.2 Implementation in the offline LDAS

The assimilation of SIF was conducted into the offline LDAS forced by the ERA5 atmospheric reanalysis (Hersbach et al, 2020). The rationale for using the offline LDAS is that the SIF observation operator was not implemented in the coupled IFS model and that the use of the offline LDAS is a preparatory step to explore the potential impact of assimilation of SIF on coupled IFS forecasts before being implemented in the coupled data assimilation.

The SIF assimilation is conducted using the SEKF to update the ecLand LAI climatology once per day. The finite difference method was used to compute the Jacobians of SIF with respect to LAI. LAI increments are computed only when the Jacobians are positive and below 0.8 to filter out spurious correlations between SIF and LAI. The SIF observation error and the LAI background error was set to 0.6 mW m^{-2} nm^{-1} sr^{-1} and of 1 m^2 m^{-2} respectively.

4.2.3 SIF data assimilation experiments

The implementation of the SIF data assimilation was conducted in two steps:

First, SIF is assimilated in the offline LDAS to update the ecLand LAI climatology for the year 2022. Soil moisture is also analysed with the assimilation of 2 m temperature and humidity and ASCAT soil moisture. The analysis of soil moisture and LAI are conducted simultaneously within each assimilation window. But in the present work, the assimilation of SIF affects only LAI and not soil moisture. The LAI increments are evaluated against the CLMS satellite LAI product (see results in 5.1.1).

The updated LAI resulting from the SIF assimilation in the offline LDAS is used to feed coupled IFS experiments instead of the default LAI climatology. The experiments are conducted over summer 2022 from 01-06-2022 to 31-08-2022. A control IFS experiment based on the default LAI climatology was also run. The impacts on GPP, soil moisture and low-meteorological fields are assessed (see section 4.4 and results in 5.1.2).

4.3 Assimilation of ASCAT backscatter in the coupled data assimilation system

4.3.1 ASCAT backscatter ML-based observation operator

A four-year training database (2016-2019), which relates ASCAT backscatter at 40° to ERA-5 reanalysis variables was used. The ERA-5 model fields were interpolated at the time and location of the ASCAT observations. The spatial and temporal sampling of the training database is that of ASCAT (25 km and daily frequency over most locations). The IFS model fields used to predict ASCAT backscatter at 40° include soil moisture and soil temperature in the first 3 soil layers (up to 1m depth) and LAI. Frozen soil, water bodies, snow area and mountain area were excluded from the training database. The ASCAT ML model was trained over the 2016-2018 period and tested over 2019. As reported in the CORSO report D4-2, a feedforward neural network with 4 hidden layers, 60 neurons showed overall good performances to predict the ASCAT backscatter at 40° signal at global scale (Garrigues et al., 2024).

4.3.2 Implementation in the IFS

A new entry for ASCAT backscatter at 40° was created in the IFS observation database to enable its use in the online LDAS. The ML observation operator, developed with the PYTORCH package, was interfaced in the IFS Fortran code using the INFERO package developed at ECMWF.

The implementation in the IFS was achieved in two steps.

First, ASCAT backscatter at 40° was assimilated in the coupled LDAS to analyse soil moisture only. The Jacobians of backscatter with respect to soil moisture were computed in finite difference using a perturbation size of 0.01 m^3 m^{-3} (de Rosnay et al., 2013). The Jacobian values which were below and above zero and 30 dB m^3 m^{-3}, respectively, were filtered out to avoid spurious negative correlations between soil moisture and the backscatter signal. The ASCAT observation error standard deviation was computed as the standard deviation of the first guess departure (difference between the observed backscatter from ASCAT and the simulated backscatter from the ML observation operator) and was equal to 0.9 dB. A background error of 0.02 m^3 m^{-3} was used for soil moisture following the operational configuration of the SEKF in IFS cycle 49r1.

Then, LAI was added in the control vector to update it from the assimilation of ASCAT backscatter. However, since LAI is not a prognostic variable but a prescribed seasonally varying climatology the updated LAI is simply used to correct for the climatological value without any prognostic propagation. This aspect will be investigated in a future work once a prognostic LAI is in place in the IFS.

4.3.3 ASCAT data assimilation experiments

We first conducted IFS experiments where only the soil moisture of the first three soil layers are updated and not LAI. The objective is to assess the impact of assimilating ASCAT backscatter instead of the ASCAT soil moisture product on the soil moisture analysis and to evaluate the added value for the forecasts of GPP, soil moisture and 2m temperature and humidity. Separate experiments with the assimilation of ASCAT backscatter and ASCAT soil moisture retrieval were performed using two distinct configurations:

1) *Sensitivity configuration*: To isolate the impact of using ASCAT observations on the soil moisture increments, the assimilation of 2 m temperature and humidity and SMOS soil moisture were switched off.
2) *Operational configuration*: The experiments were run using ASCAT soil moisture or ASCAT backscatter along with the rest of observations which are used in the operational configuration of the soil moisture analysis.

In each configuration, we compared the soil moisture increments of the superficial soil layer and the root-zone (equivalent to 1 m of topsoil) obtained with the assimilation of ASCAT backscatter and ASCAT soil moisture. We evaluated the impact on GPP and NWP fields (2m temperature and humidity).

In a second step, we performed an IFS experiment where ASCAT backscatter was assimilated and both soil moisture and LAI were analysed. The experiment was done using the operational configuration where all the observations of soil moisture analysis were exploited. The evaluation was focused on the total LAI increments.

The IFS coupled experiments were conducted from the 1/06/2022 to 31/08/2022. For each configuration, we compared the experiment based on the assimilation of ASCAT backscatter with the one conducted with the assimilation of ASCAT soil moisture retrieval which is considered as the control (see results in 5.2.1).

Experiment names	Configurations	Control vector	ASCAT observations
Sens_B40	Sensitivity	Soil moisture	Backscatter 40°
Sens_SM	Sensitivity	Soil moisture	Soil moisture retrieval
Oper_B40	Operational	Soil moisture	Backscatter 40°
Oper_SM	Operational	Soil moisture	Soil moisture retrieval
Oper_B40_LAI	Operational	Soil moisture and LAI	Backscatter 40°

Table 1: List of IFS experiments with the assimilation of ASCAT data

4.4 Evaluation of GPP, soil moisture and low-level meteorological fields IFS forecasts:

The performances of each GPP forecast were quantified by the RMSE (root mean square error) between the predicted GPP and the FluxSat GPP used as a reference. The impact of the SIF or ASCAT assimilation on the GPP forecast performance was quantified by the GPP

RMSE differences between the IFS experiments with the assimilation of SIF or ASCAT and the control experiments based on the default operational configuration. FluxSat GPP was used as a reference.

Surface and root-zone soil moisture predictions were compared to in situ measurements from various measurement networks across the world (see section 3.5). The performances were quantified by the anomaly correlation of soil moisture computed in time at each in situ site and then averaged per network. The anomalies were computed as the difference between the soil moisture measurement time series and a centered 35-day moving average (more detailed can be found in Fairbairn et al., 2019).

Meteorological fields (2m temperature and humidity, geopotential height) were evaluated using the operational analysis as reference by computing the RMSE of the forecast and the operational analysis. The performances of each experiment were quantified by the RMSE normalized differences between the IFS experiment with the assimilation of SIF or ASCAT and the IFS control experiment, using the operational IFS analysis as a reference (Geer et al., 2016).

The evaluation results are presented in Section 5.1.2 for SIF and 5.2.2, 5.2.3 for ASCAT.

5 Results

5.1 Assimilation of SIF

5.1.1 Evaluation of LAI increments

Figure 1 highlights that the impacts of assimilating SIF compared to the use of the LAI climatology is limited to specific regions. The improvement of LAI with respect to the CLMS satellite product mainly concerns cropland as illustrated by the soybean production region in South America, the corn and wheat areas in the United-States and the wheat production areas in Australia (Figure 2). The assimilation of SIF generates meaningful regional spatial patterns of LAI increments in response to climate anomalies such as the reduction of LAI over Western Europe in July 2022 (Figure 3) which is consistent with the soil moisture deficit reported for the summer 2022 over Western Europe (C3S, 2023). However, LAI is degraded over tropical rainforests over the Amazon, central Africa and Indonesia.

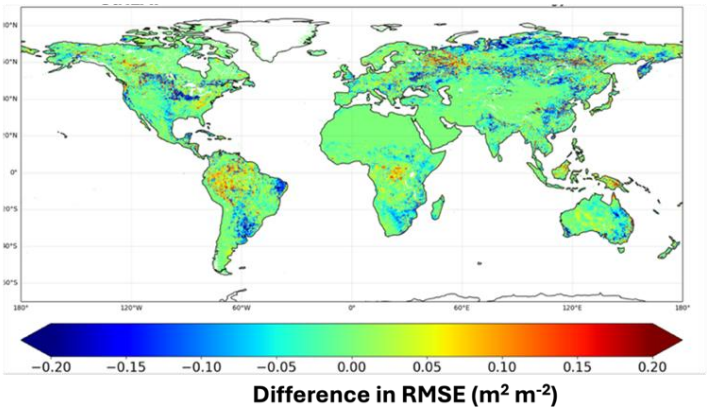


Figure 1: Impact of SIF data assimilation shown as RMSE differences with vs without SIF data assimilation, using CLMS LAI as a reference, for 2022, at global scale. Blue and red colours indicate improvement and degradation of LAI, respectively.

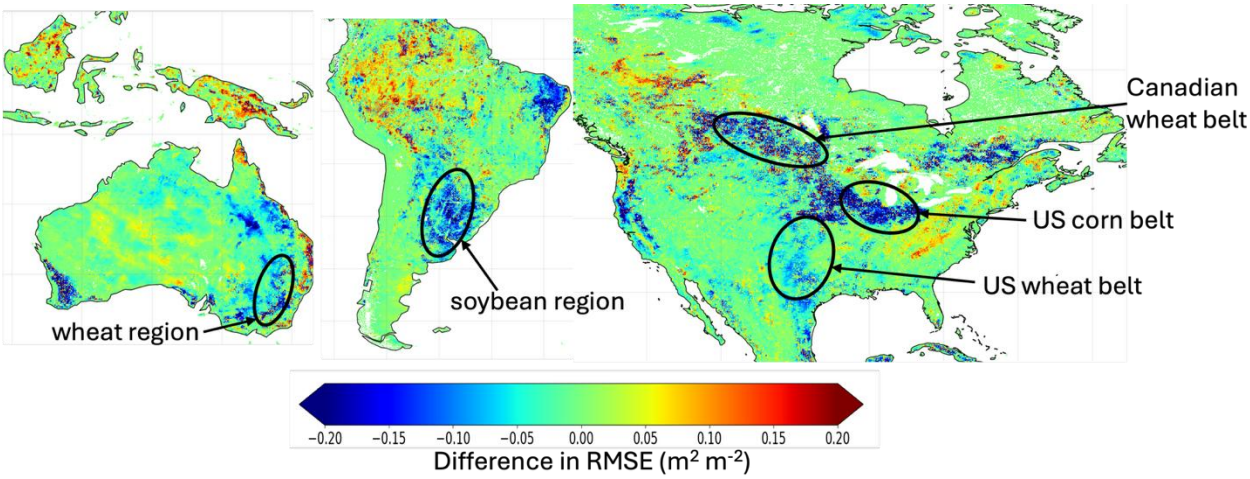


Figure 2: As Figure 1 but zoomed over North America, Australia and South America.

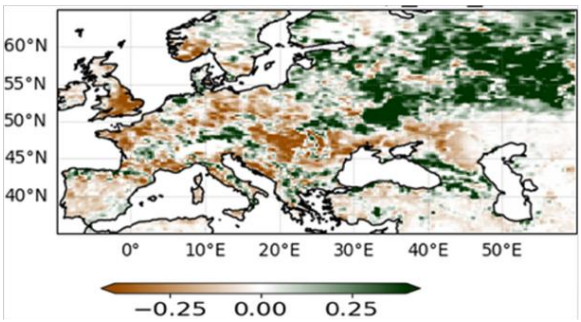


Figure 3: Regional monthly mean of LAI increments (unit: m^2 m^{-2}) over Europe in July 2022

5.1.2 Impact on coupled model forecasts

5.1.2.1 GPP

Figure 4 represents the global map of the GPP RMSE differences between the experiments done with versus without the updated LAI produced from the assimilation of SIF. The GPP RMSE of each experiment is computed using the FluxSat GPP dataset as reference. Figure 4 indicates mixed results with a degradation of GPP in the Amazon, in part of North America and Eurasia and local improvements in Africa, North America and central Europe. Over North America, GPP is improved over the wheat Canadian region and the US corn belt, but it is degraded over the US wheat belt.

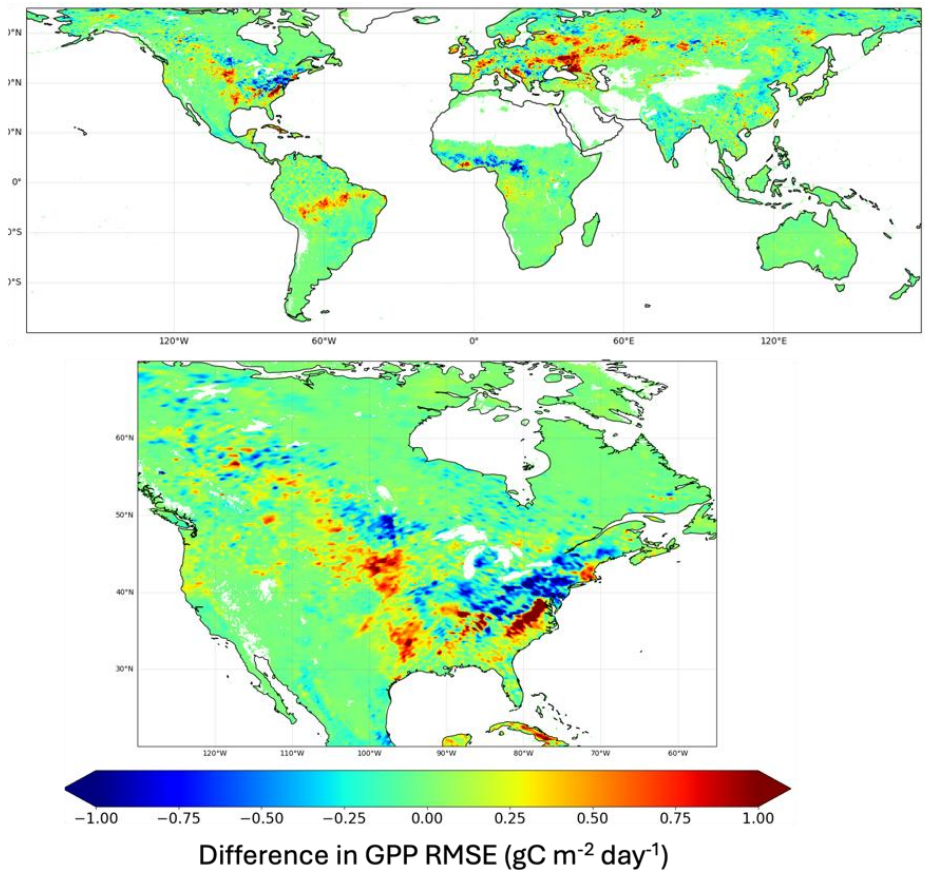


Figure 4: Impact of SIF DA on GPP forecast performance shown as GPP RMSE differences with vs without SIF DA, using FluxSat as reference, for summer 2022, at global scale and over North America.

5.1.2.2 Soil moisture

The assimilation of SIF generally leads to slight but non-significant improvements in the surface soil moisture forecast. The improvement is larger for the root-zone soil moisture which is expected since the assimilation of SIF improves the representation of vegetation processes.

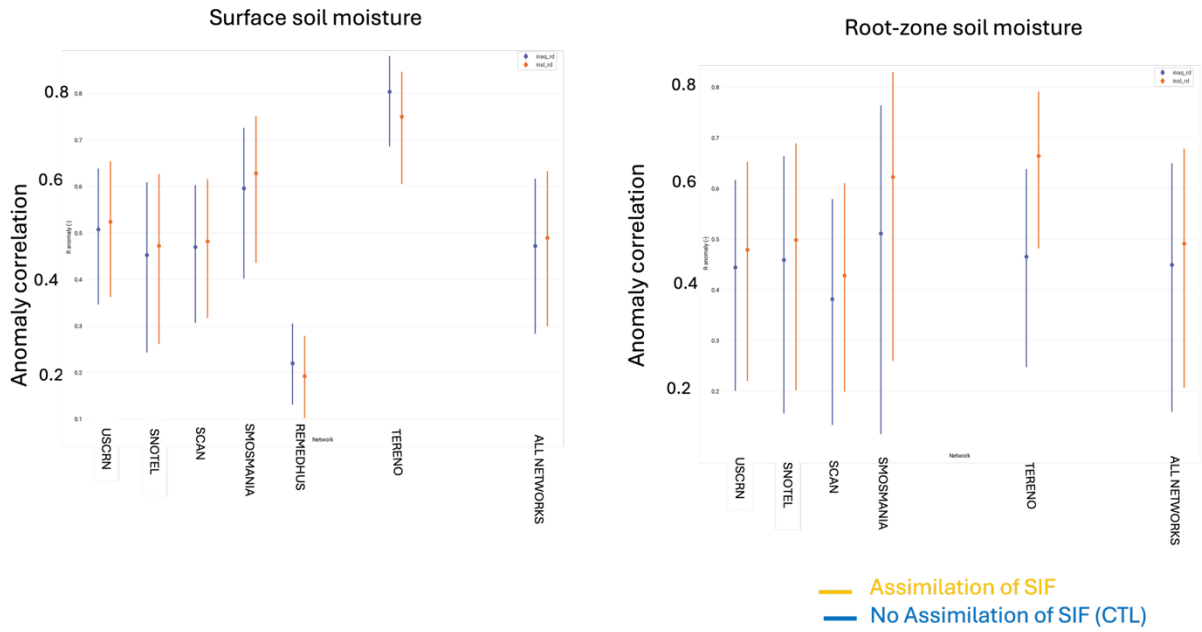


Figure 5: Evaluation of surface and root-zone soil moisture IFS prediction against in situ soil moisture measurements for various networks (names reported along the x-axis). The performance score is quantified by the anomaly correlation computed in time for each in situ network.

5.1.2.3 NWP fields

Figure 6 shows that the use of the updated LAI in the IFS forecast leads to degradation of 2 m temperature and humidity. While improvements are observed at short lead time (12h) over the Amazon for both 2m temperature and humidity, Central Europe for 2m temperature and Australia for 2m humidity, the signal does not persist over longer lead times and a steady degradation of 2m temperature and humidity appears over the Amazon from T+24h.

Figure 7 shows, however, a positive impact of the updated LAI on the general circulation over the tropic region as shown by the reduction in RMSE of geopotential height in the troposphere which persists across lead times.

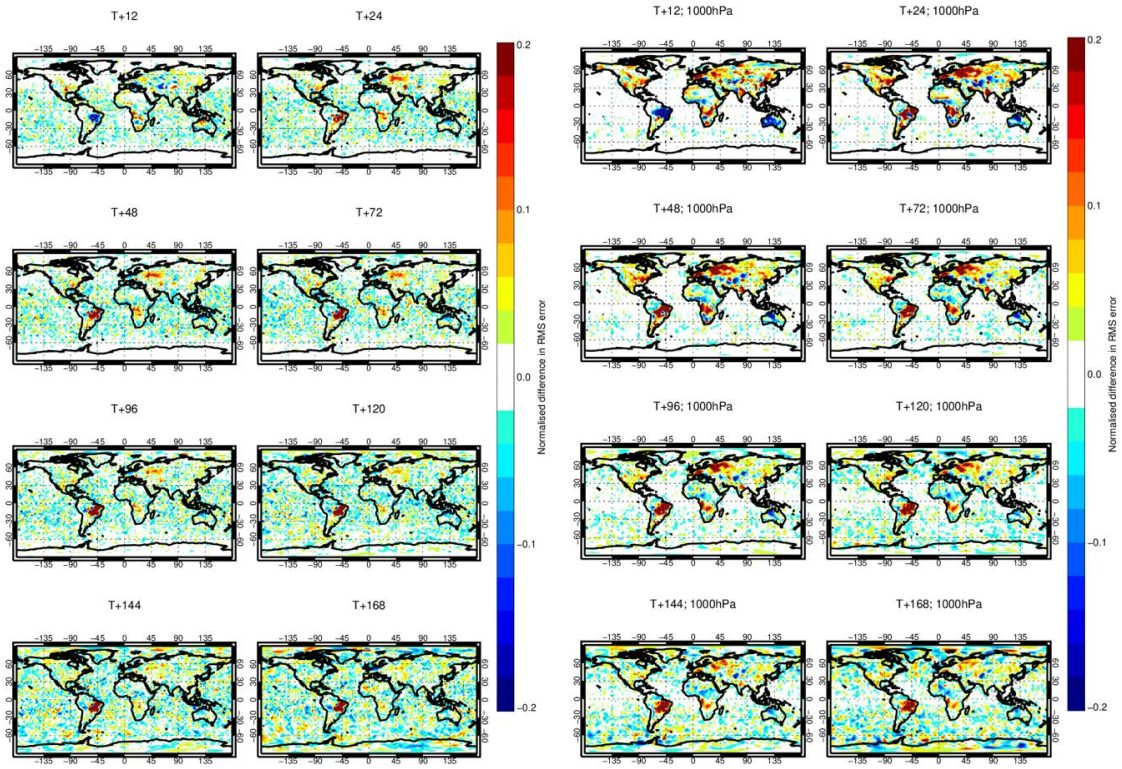


Figure 6: Impact of SIF data assimilation on 2 m temperature (left panel) and 1000 hPa humidity (right panel) forecasts from the IFS at different lead times (12h to 168h forecast) for the June-August 2022 period. The maps show the difference in RMSE between the experiment conducted with the updated LAI and the control based on the LAI climatology. The evaluation is done against the operational analysis.

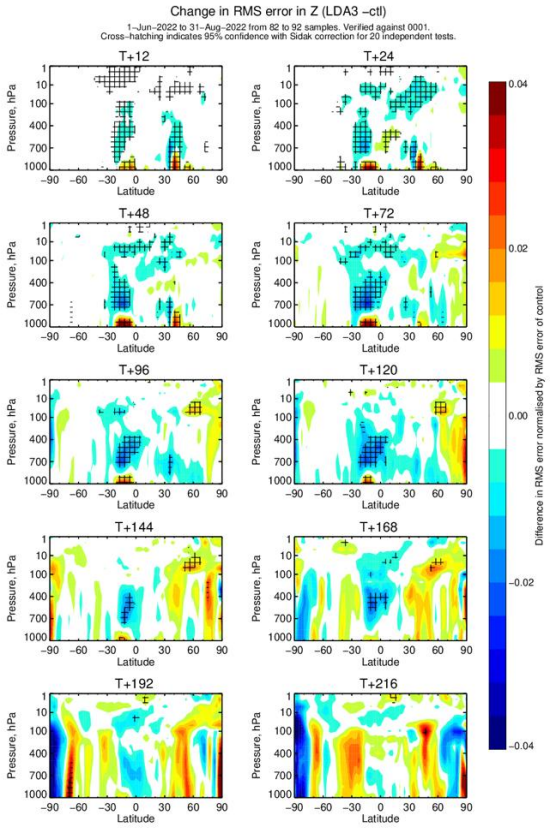


Figure 7: Impact of SIF data assimilation on geopotential height (Z) forecast from the IFS at different lead times (12h to 216h forecast) for the June-August 2022 period. The plots show the differences in RMSE between the experiment conducted with the updated LAI and the control based on the LAI climatology as a function of latitude and pressure level. The evaluation is done against the operational analysis by computing the RMSE of the forecast of geopotential height from each experiment and the analysis.

5.1.3 Discussion

The assimilation of SIF mainly improves the LAI climatology over cropland for which the LAI climatology is associated with larger uncertainties due to the large interannual variability generated by agricultural practices. The SIF ML-based observation operator trained on satellite LAI captures the strong statistical correlations between the SIF signal and LAI over cropland. This highlights that the SIF satellite retrieval is mainly influenced by the seasonal variability of the vegetation structure at this spatial and temporal resolution. The lower performances obtained for rainforests are related to the larger uncertainties in satellite LAI over dense vegetation and the inability of the ML observation operator to represent the SIF variability due to vegetation physiology which is the main driver of the SIF signal for this vegetation type. The improvement in LAI does not lead to systematic improvement in GPP which can be due to limitations in the current representation of vegetation in ecLand and the prevailing impact of other sources of biases in the coupled land-atmosphere model. Besides, satellite-based GPP estimates, used here as reference for the evaluation, are not direct measurements of GPP and are associated with various sources of uncertainties (Jung et al., 2020). Multi-GPP datasets should be exploited in future works to account for the uncertainties in the reference datasets. Finally, the lack of improvement of low-level meteorological fields can be due to the strong tuning of surface-atmosphere interactions in the IFS which limits the propagation of improved land surface variables into the atmospheric fields.

5.2 Evaluation of the assimilation of ASCAT backscatter

In sections 5.2.1 to 5.2.3 below, we present the results of experiments where only soil moisture is analysed. In section 5.2.4, the LAI increments produced from the experiment conducted with the analysis of both soil moisture and LAI are assessed.

5.2.1 Impact of assimilating ASCAT backscatter on soil moisture analysis

5.2.1.1 Comparison of soil moisture increments between ASCAT soil moisture and ASCAT backscatter data assimilation

Figures 8 to 10 display global maps of seasonal mean of soil moisture increments obtained in Summer 2022 from the assimilation of ASCAT using either backscatter (left) or soil moisture retrieval (right). In the absence of observations, null increment values were filtered out when the temporal average is computed.

For the sensitivity configuration, the increments of surface and root-zone soil moisture generated by the assimilation of ASCAT backscatter versus soil moisture retrieval exhibit different spatial distribution and order of magnitude (Figure 8 and 9). The assimilation of backscatter produces larger variability of increments with some contrasted structures. The difference in magnitude is particularly important for the root-zone soil moisture which reflects the higher sensitivity of the ML observation operator to the soil moisture of the deep soil layers. Figure 8 also highlights the larger spatial coverage of increments when assimilating backscatter than soil moisture observations. This is particularly clear in tropical rainforest regions where the soil moisture retrievals were filtered out during the quality control process.

For the operational configuration, the differences in the increment's magnitude and spatial distribution are smaller compared to the sensitivity configuration, because the increments are strongly driven by the assimilation of 2 m temperature and humidity pseudo-observations. But the assimilation of backscatter still produces larger soil moisture increments in the root-zone than the assimilation of retrieved soil moisture.

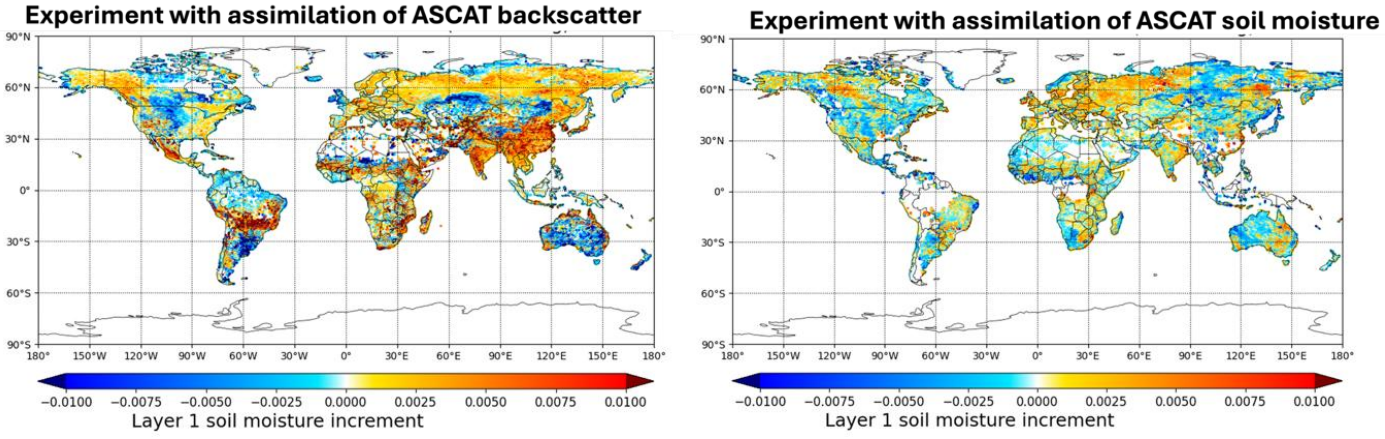


Figure 8: Seasonal average of surface soil moisture increments (m^3 m^{-3}) in the sensitivity configuration where only ASCAT observations are assimilated, using backscatter (left) and retrieved soil moisture (right), for the summer 2022 period.

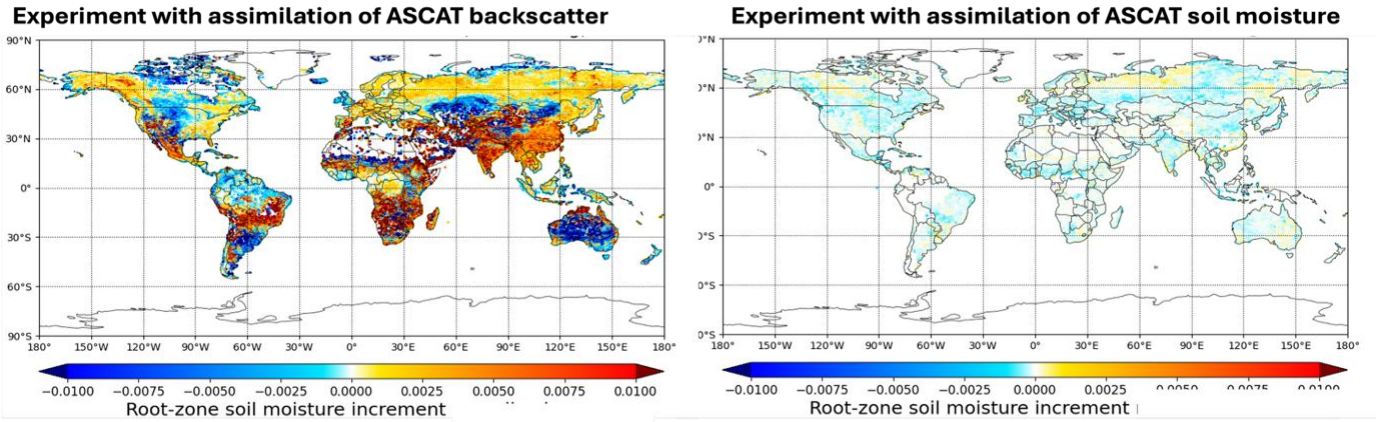
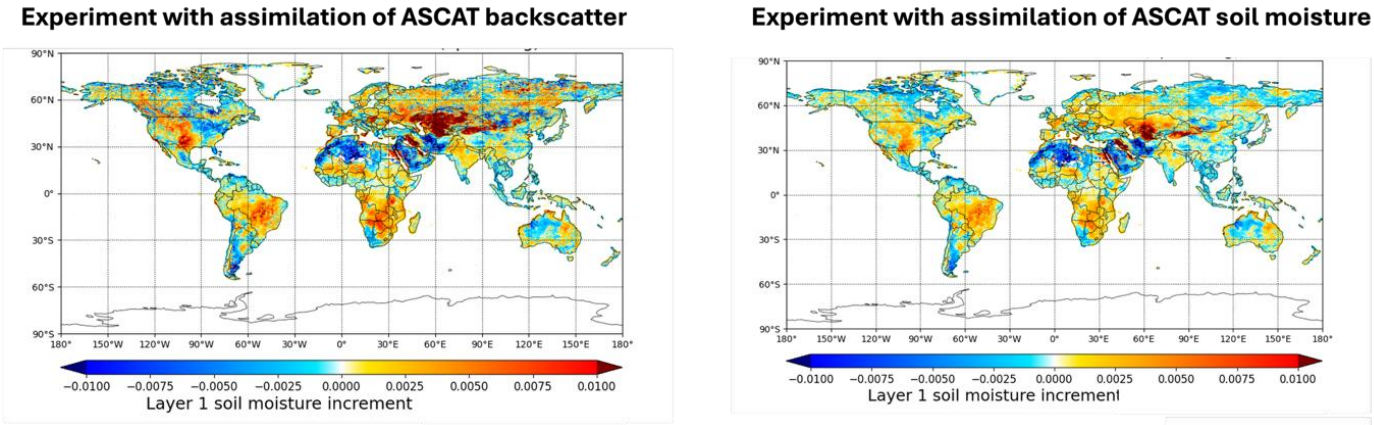


Figure 9: Same as Figure 8 but for the root zone (top metre of soil).

Figure 10: Seasonal average of surface soil moisture increments ($m^3\ m^{-3}$) in the operational



configuration using the full observing system including SMOS and low-level meteorological field observations, and ASCAT backscatter (left) and soil moisture (right), for the summer 2022 period

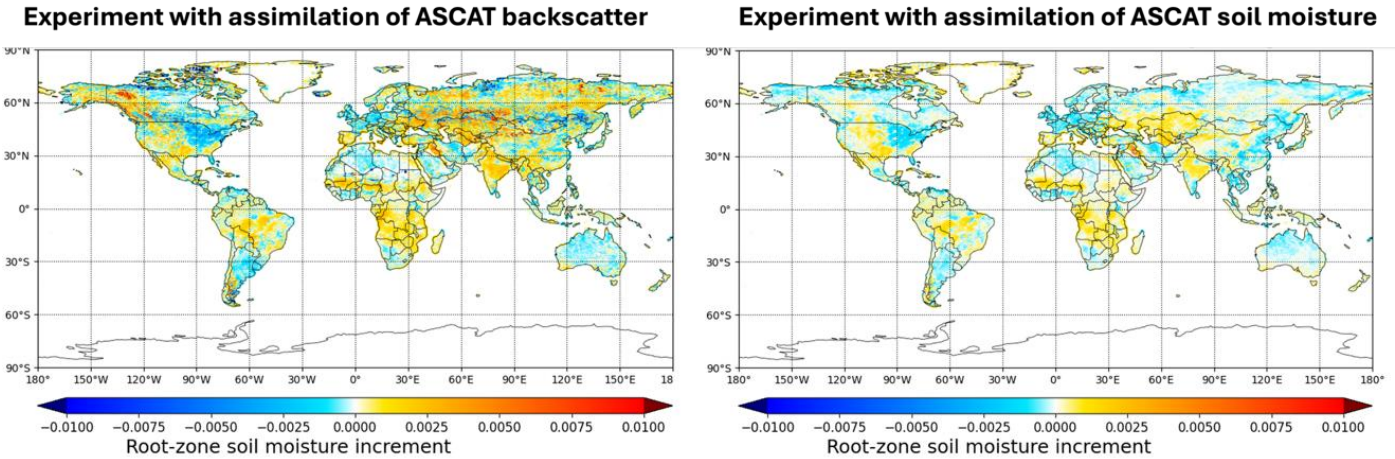


Figure 11: Same as Figure 10 for the root zone (top metre of soil).

5.2.1.2 Evaluation against in situ soil moisture measurements

The verification performed against various soil moisture in situ networks displayed in Figure 12 shows that, compared to the assimilation of retrieved soil moisture, the assimilation of backscatter leads to slight improvements in the simulated surface and root-zone soil moisture.

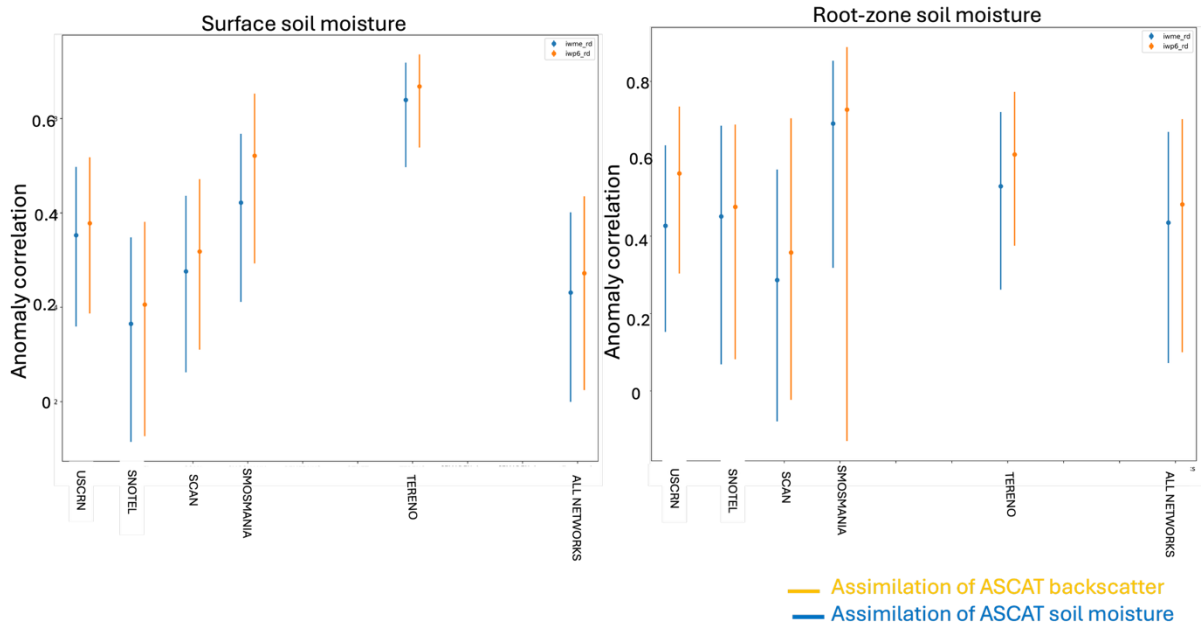


Figure 12: Evaluation of surface and root-zone soil moisture IFS prediction (m^3 m^{-3}) against in situ soil moisture measurements (m^3 m^{-3}) for various networks (names reported along the x-axis). The performance score is quantified by the anomaly correlation computed in time for each in situ network.

5.2.2 Impact of assimilating ASCAT backscatter on GPP

Figure 13 represents the global map of the GPP RMSE differences between the experiments done with the assimilation of ASCAT backscatter versus the assimilation of ASCAT soil moisture. The GPP RMSE of each experiment is computed using the FluxSat GPP dataset as reference. Figure 13 shows that the assimilation of ASCAT backscatter in lieu of ASCAT soil moisture leads to improvements in GPP forecast over Northern latitude particular in Europe and locations in South America and Australia while more mixed results are obtained over tropical rainforest and in North America.

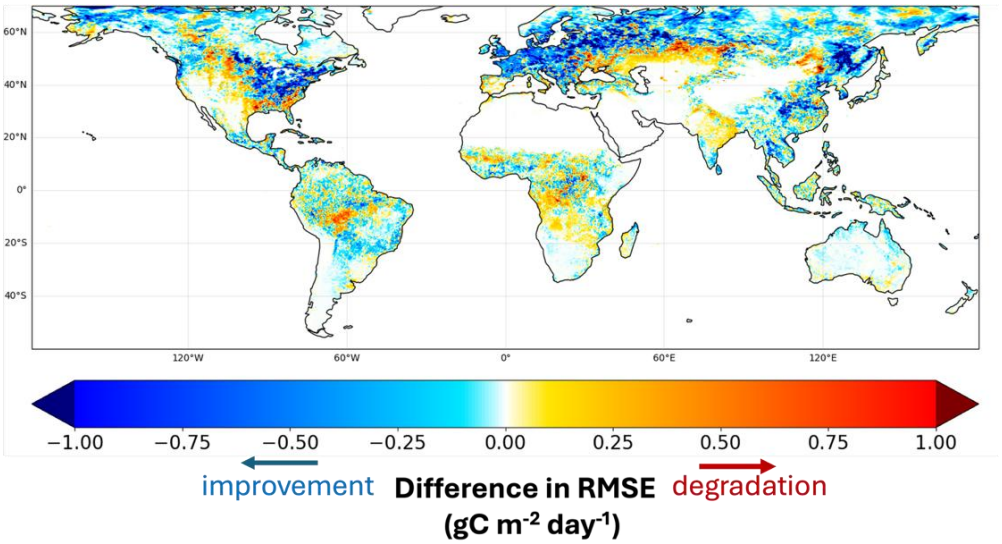


Figure 13: Impact of ASCAT backscatter on GPP forecast performances shown as GPP RMSE differences between the assimilation of ASCAT backscatter versus the assimilation of ASCAT soil moisture, using FluxSat as reference, for summer 2022 for the operational configuration.

5.2.3 Impact of assimilating ASCAT backscatter on NWP forecasts

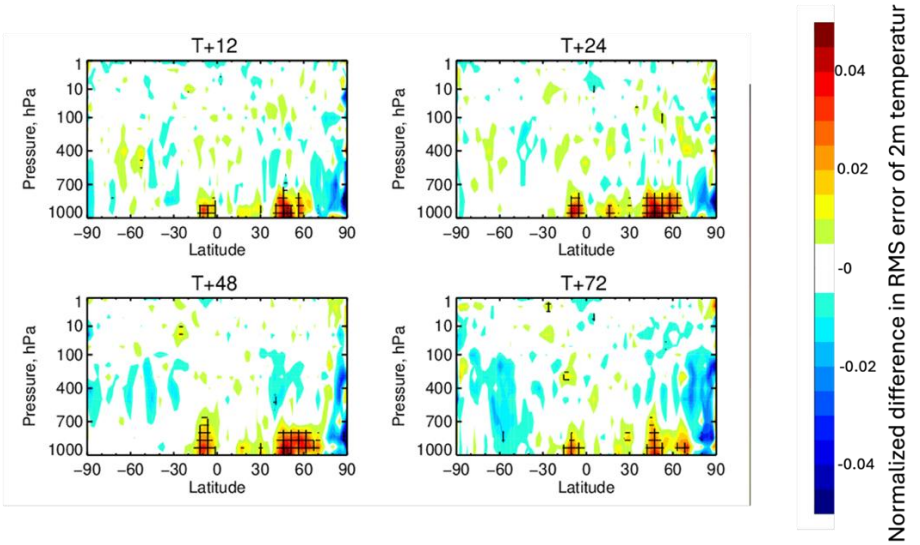


Figure 14: Impact of ASCAT backscatter on forecast performance of temperature vertical profile shown as RMSE normalized differences between the assimilation of ASCAT backscatter vs ASCAT soil moisture using the operational analysis as reference, for summer 2022, for the operational configuration. The RMSE difference is normalized by dividing by the RMSE related to the experiment done with ASCAT soil moisture. Cross hatching indicates statistically significant differences. The plots show the vertical distribution of RMSE across latitudes for distinct lead times.

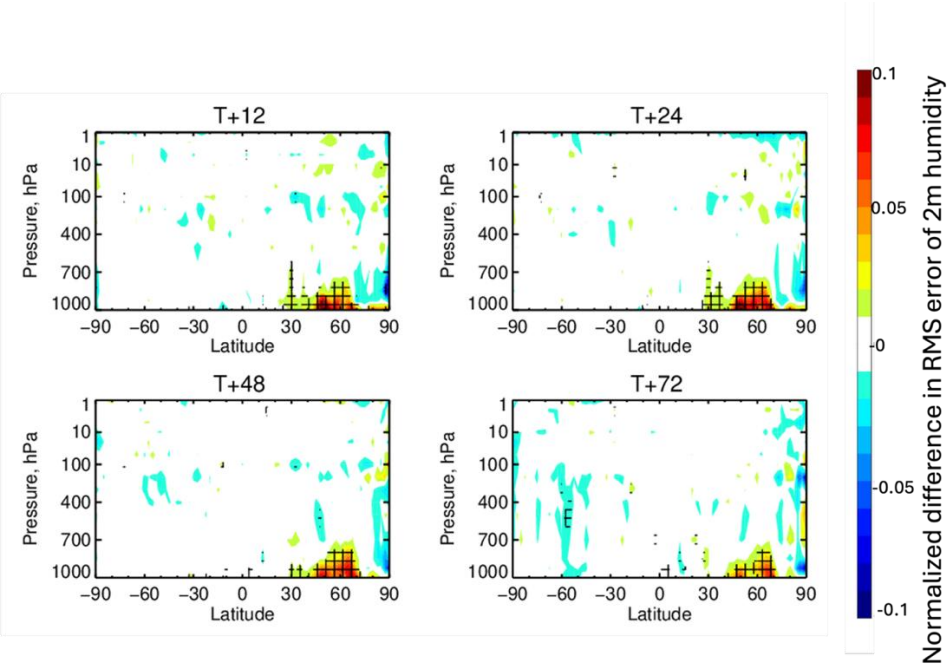


Figure 15: Impact of ASCAT backscatter on forecast performance of humidity vertical profile shown as RMSE differences between the assimilation of ASCAT backscatter vs ASCAT soil moisture using the operational analysis as reference, for summer 2022, for the operational configuration. The RMSE difference is normalized by dividing by the RMSE related to the experiment done with ASCAT soil moisture. Cross hatching indicates statistically significant differences. The plots show the vertical distribution of RMSE across latitudes for distinct lead times.

Figure 14 shows significant degradation in low-level temperature located in the tropics and Northern latitude in the operational configuration when ASCAT backscatter is used instead of soil moisture. Figure 15 indicates significant degradation of low-level humidity at Northern latitudes. The maps of RMSE (not shown here) indicate degradations across Europe and North America and some improvements in South America and Africa.

5.2.4 Impact on LAI increments

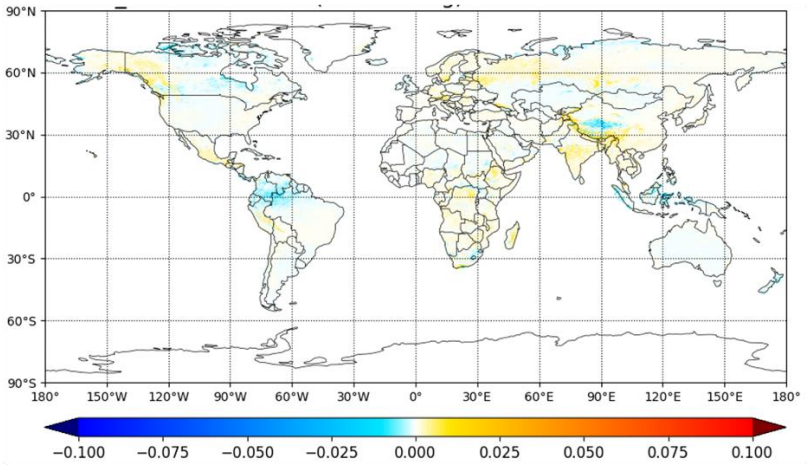


Figure 16: Global seasonal mean of LAI increments (unit: $m^2\ m^{-2}$) in summer 2022.

Figure 16 illustrates results when LAI is included along with soil moisture in the data assimilation control vector. It illustrates the capability of the ASCAT backscatter data assimilation system to produce LAI increments and update the IFS LAI climatology. The magnitude of the increments generated by the assimilation of backscatter is smaller than the one produced by SIF data assimilation but the comparison is difficult given SIF was assimilated in the offline LDAS while ASCAT is assimilated in the coupled LDAS where the weak coupling between the surface analysis and the atmospheric analysis plays a role.

5.2.5 Discussion

The differences in soil moisture increments generated by the assimilation of ASCAT backscatter versus soil moisture can be related to differences in information content between the backscatter signal and the soil moisture retrieval. The backscatter normalized at 40° is closer to the satellite measurement while the soil moisture retrieval is derived from a retrieval algorithm that relies on specific assumptions and vegetation maps. Besides, the application of a CDF matching to soil moisture retrieval and a ML observation operator used to simulate the model equivalent of ASCAT backscatter can play a role. Finally, differences in the quality control, particularly in the tropical regions, can generate differences in the number and location of the assimilated observations.

The assimilation of backscatter improves the simulation of some land surface variables such as soil moisture and GPP. The impact on GPP can result from improvements in soil moisture which was proved to be a key driver of the land carbon uptake variability (Humphrey et al., 2021). However, the degradation obtained for low-level meteorological fields shows that the improvement of land surface variables does not systematically translate into improvement in the atmosphere fields because of limitations in the vegetation representation in ecLand (prescribed LAI) and the strong tuning of land-surface interactions which are very well-known aspects of the IFS.

6 Conclusions

This work demonstrates the potential of ML observation operators to exploit land observations and strengthen the coupled water and carbon cycle land data assimilation system in the IFS. The assimilation of SIF provides a more realistic representation of the vegetation temporal dynamics over croplands which is very relevant for the monitoring of emissions over agriculture regions in the future CO2MVS. The assimilation of ASCAT backscatter represents a great potential for the joint analysis of soil moisture and leaf area index in the coupled model. The first results obtained by updating only soil moisture show positive impacts on soil moisture and GPP but some degradations for the meteorological fields. Enhanced consistency between surface improvements and atmospheric impact is expected with the prognostic LAI which is under development in the IFS.

This study also highlights that the improvement in vegetation processes representation provided by the assimilation of SIF or backscatter does not systematically translate into improved forecast scores for 2m temperature and humidity variables. This reflects the strong tuning of land-surface interactions which can limit the propagation of improved land surface variables into the atmospheric fields. These results need to be refined, and possible improvements need to be reassessed with the ongoing revision of the parametrization of land surface-atmosphere interactions and the development of a new prognostic LAI in the IFS.

The main recommendations and steps forwards are:

- Retrain the ML observation operator approach with the most recent IFS cycle to take advantage of the upcoming improvements of the vegetation process modelling in the IFS.

- Use the PYTORCH differentiation functionality to compute the ML observation operator Jacobian instead of using the finite difference method which substantially increases the computing cost in coupled simulations.
- Extend SIF and ASCAT backscatter data assimilation to the upcoming ecLand prognostic LAI in both the offline and online LDAS.
- Assess the joint analysis of soil moisture and LAI in the IFS coupled model with a focus on strong surface-atmosphere coupling conditions such as drought.
- Exploit SIF and ASCAT backscatter in purely data-driven forecast systems which may be capable to extract more efficiently the information content from those satellite observations compared to a traditional Earth System Model which can be limited by the lack of accurate enough representation of land surface processes.

7 References

Albergel, C., Zakharova, E., Calvet, J.-C., Zribi, M., Pard ´ e, M., Wigneron, J.-P., Novello, N., Kerr, Y., Mialon, A., and Fritz, N.: A first assessment of the SMOS data in southwestern France using in situ and airborne soil moisture estimates: the CAROLS airborne campaign, Remote Sens. Environ., 115, 2718–2728, doi:10.1016/j.rse.2011.06.012, 2011

Bell, J. E., M. A. Palecki, C. B. Baker, W. G. Collins, J. H. Lawrimore, R. D. Leeper, M. E. Hall, J. Kochendorfer, T. P. Meyers, T. Wilson, and H. J. Diamond. 2013: U.S. Climate Reference Network soil moisture and temperature observations. J. Hydrometeorol., 14, 977-988, https://doi.org/10.1175/JHM-D-12-0146.1;

Boussetta, S., G. Balsamo, G. Arduini, E. Dutra, J. McNorton, M. Choulga, A. Agustí-Panareda, A. Beljaars, N. Wedi, J. Munõz-Sabater, et al, 2021. ECLand: The ECMWF Land Surface Modelling System, *Atmosphere*; 12(6):723. https://doi.org/10.3390/atmos12060723.

CLMS LAI. Leaf Area Index 2014-present (raster 300 m), global, 10-daily – version 1. *European Union's Copernicus Land Monitoring Service information https://land.copernicus.eu/en/products/vegetation/leaf-area-index-300m-v1.0 (Accessed on 23.05.2025).*

de Rosnay P., M. Drusch, D. Vasiljevic, G. Balsamo, C. Albergel and L. Isaksen, 2013. A simplified Extended Kalman Filter for the global operational soil moisture analysis at ECMWF, Q. J. R. Meteorol. Soc., 139:1199-1213, 2013 doi: 10.1002/qj.2023

de Rosnay, P., Browne, P., de Boisséson, E., Fairbairn, D., Hirahara, Y., Ochi, K., et al., 2022. Coupled data assimilation at ECMWF: current status, challenges and future developments. Quarterly Journal of the Royal Meteorological Society, 148(747), 2672–2702. Available from: https://doi.org/10.1002/qj.4330

Fairbairn, D., P. de Rosnay, and P. A. Browne 2019. The new stand-alone surface analysis at ecmwf: Implications for land–atmosphere da coupling. Journal of Hydrometeorology, 20(10):2023–2042

Garrigues, S., de Rosnay, P., Pinnington, E; Weston, P., Agusti-Panareda, A., Boussetta, S., Calvet-JC., Fairbain, D., Bacour, C., Engelen, R., English., S. Machine learning-based observation operators to assimilate microwave and SIF satellite observations into the ECMWF integrated forecast system, IGARSS 2024

Garrigues, S., P. de Rosnay, P. Weston, C. Rüdiger, D. Fairbairn, E. Pinnington, S. Boussetta, A. Agusti-Panareda, R. Engelen, P. Vanderbecken, J.-C. Calvet: Solar-induced fluorescence satellite data in the land assimilation system, ECMWF Newsletter article, Summer 2025a https://www.ecmwf.int/en/newsletter/184/news/solar-induced-fluorescence-satellite-data-land-assimilation-system

Garrigues S., P. de Rosnay , P. Weston, C. Rüdiger, C. Bacour, D. Fairbairn, E. Pinnington, A. Agusti-Panareda, S. Boussetta, J.-C. Calvet, R. Engelen: "Assimilation of Satellite Solar Induced Fluorescence at the global scale in the ECMWF Integrated Forecast System " in review QJRMS, 2025b

Geer, A. J. (2016). Significance of changes in medium-range forecast scores. Tellus A: Dynamic Meteorology and Oceanography, 68(1). https://doi.org/10.3402/tellusa.v68.30229

Guanter, L., Bacour, C., Schneider, A., Aben, I., van Kempen, T. A., Maignan, F., Retscher, C., Köhler, P., Frankenberg, C., Joiner, J., and Zhang, Y., 2021. The TROPOSIF global sun-induced fluorescence dataset from the Sentinel-5P TROPOMI mission, Earth Syst. Sci. Data, 13, 5423–5440, https://doi.org/10.5194/essd-13-5423-2021, 2021.

Hersbach H, Bell B, Berrisford P, et al, 2020. The ERA5 global reanalysis. *Q J R Meteorol Soc*. 2020; 146: 1999–2049. https://doi.org/10.1002/qj.3803

Humphrey, V., Berg, A., Ciais, P. *et al.* Soil moisture–atmosphere feedBack dominates land carbon uptake variability. *Nature* **592**, 65–69 (2021). https://doi.org/10.1038/s41586-021-03325-5

Joiner, J., and Y. Yoshida. 2021. Global MODIS and FLUXNET-derived Daily Gross Primary Production, V2. ORNL DAAC, Oak Ridge, Tennessee, USA.https://doi.org/10.3334/ORNLDAAC/1835

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., Goll, D. S., Haverd, V., Köhler, P., Ichii, K., Jain, A. K., Liu, J., Lombardozzi, D., Nabel, J. E. M. S., Nelson, J. A., O'Sullivan, M., Pallandt, M., Papale, D., Peters, W., Pongratz, J., Rödenbeck, C., Sitch, S., Tramontana, G., Walker, A., Weber, U., and Reichstein, M.: Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach, Biogeosciences, 17, 1343–1365, https://doi.org/10.5194/bg-17-1343-2020, 2020.

Document History

Version	Author(s)	Date	Changes
1.0	Sebastien Garrigues	04/12/2025	
2.02	Sebastien Garrigues	17/12/2025	Implemented reviewer comments

Internal Review History

Internal Reviewers	Date	Comments
Susanne Preunkert	12.12.2025	
Samuel Hammer	12.12.2025	
Thomas Kaminski	16.12.2025	