ESA Contract Report

ESA Contract (4000144712/24/I-DT-bgh)

Contract Report to the European Space Agency

Advancing global live fuel moisture mapping through multi-sensor data assimilation

Authors: Siham El Garroussi, Patricia de Rosnay, Sebastien Garrigues, David Fairbairn, Joe McNorton, Francesca Di Giuseppe
Contract officer: Claudia Vitolo

June 2025

Series: ECMWF ESA Contract Report Series

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/en/publications/

Contact: library@ecmwf.int

Abstract

This deliverable outlines the methodology and results of Task 2 of Fuelity project, which focused on integrating Earth Observation data into the ECMWF land data assimilation system (LDAS) to constrain live fuel moisture content (LFMC) estimates. The aim is to support a high-resolution, observation-informed global LFMC reanalysis. Enhancements to ECMWF's ecLand system include targeted modifications to the Simplified Extended Kalman Filter to assimilate L-band Vegetation Optical Depth, Solar-Induced Fluorescence, and ASCAT backscatter data, thereby improving estimates of Leaf Area Index and soil moisture, which underpin LFMC. A machine learning-based post-assimilation bias correction further refines the LAI inputs. Evaluation shows improved seasonal vegetation dynamics and realistic LFMC spatial and temporal patterns. The system successfully captured canopy loss during the 2021 Evia wildfire, demonstrating its potential for operational fire monitoring and risk assessment.

Plain language summary

Satellite data can enhance fire forecasting by providing information about the dryness of vegetation; this is known as live fuel moisture content (LFMC). When vegetation is dry, it catches fire more easily. Accurately estimating LFMC is therefore essential for predicting wildfire risk.

As part of the Fuelity project, this report explains how satellite observations were used to improve LFMC estimates in the European Centre for Medium-Range Weather Forecasts (ECMWF)'s land surface model, called ecLand. LFMC mainly depends on two factors:

- Leaf Area Index: a measure of the amount of leaf cover in an area
- Soil moisture: the amount of water stored in the ground

To enhance both of these, the ecLand model was updated to include new types of satellite data:

- L-band Vegetation Optical Depth from the SMOS satellite and Solar-Induced Fluorescence from Sentinel-5P help estimate LAI
- ASCAT (a satellite radar instrument) provides measurements of soil moisture

These updates allowed the model to track seasonal changes in vegetation more accurately, particularly in forested areas. A machine learning method was also added to correct any remaining errors in the model's output by combining multiple data sources.

The improved system successfully detected vegetation loss during the 2021 wildfire in Evia, Greece. This demonstrates its potential to support fire risk monitoring worldwide and enhance early warning systems.

Executive summary

This deliverable presents the methodology and outcomes of Task 2, which aims to implement the technical components required to integrate Earth Observation (EO) data into the ECMWF forecasting system, using the European Centre for Medium-Range Weather Forecasts (ECMWF)'s land data assimilation system (LDAS) to constrain live fuel moisture content (LFMC) estimates. The overarching objective is to support the development of a high-resolution, observation-informed global LFMC reanalysis dataset.

The ECMWF land data assimilation framework is designed to be extensible to fuel load variables once the SPARKY fire module is coupled to the Integrated Forecasting System (IFS) and can also ingest additional EO data. The implementation of LFMC data assimilation is based on the formulation of LFMC within the SPARKY model, where improvements in LFMC rely on better estimates of both Leaf Area Index (LAI) and soil moisture (SM). Targeted modifications were made to ecLand's offline Simplified Extended Kalman Filter (SEKF) infrastructure. New routines and model variables were introduced in ecLand to explicitly represent fuel moisture. For instance, the SEKF system was extended to assimilate L-band Vegetation Optical Depth (L-VOD) from Soil Moisture and Ocean Salinity (SMOS) and Solar-Induced Fluorescence (SIF) from Sentinel-5P to improve LAI, as well as Advanced Scatterometer (ASCAT) backscatter data to enhance SM estimates. These developments allow the model to update fuel moisture states more accurately while accounting for associated uncertainties.

Evaluation of the system demonstrated that assimilating L-VOD and SIF to analyse LAI improved the representation of seasonal vegetation dynamics, particularly in temperate and boreal regions. However, assimilation performance varied regionally due to observational gaps and structural limitations. A post-assimilation bias correction was introduced to overcome these challenges.

The bias correction employs a machine learning (ML)-based approach that integrates leaf onset timing, recent Copernicus Global Land Service (CGLS) LAI observations, and increments from both VOD and SIF assimilation experiments. This process generates bias-corrected LAI inputs for LFMC estimation.

The diagnostic LFMC, computed from assimilated soil moisture and corrected LAI, captured realistic spatial patterns and seasonal variation. The August 2021 wildfire in northern Evia (Greece) demonstrated that the system effectively captured rapid canopy loss following fire events, closely tracking independent moderate resolution imaging spectroradiometer (MODIS) fire detections and CGLS LAI observations. These results confirm the system's capability to represent both gradual phenological changes and sudden disturbances, providing a valuable foundation for future operational fire monitoring and risk prediction.

Contents

1 Introduction

Live fuel moisture content (LFMC)—the ratio of water content to dry biomass in living vegetation—is a fundamental biophysical variable controlling vegetation flammability and wildfire behaviour [31]. It influences ignition probability, fire intensity, and rate of spread, making LFMC a central determinant in fire risk assessment frameworks and operational early warning systems. Critically, as climate change drives more frequent and severe droughts, ecosystems are increasingly vulnerable to transitioning into fire-prone states, with LFMC acting as a key threshold variable modulating this shift [23].

Despite its importance, monitoring LFMC globally and in near real-time remains a significant challenge. In situ measurements require destructive sampling and oven-drying of plant material, which is logistically intensive and geographically sparse. Furthermore, LFMC is inherently complex; it is governed not only by soil moisture availability and plant water uptake but also by species-specific physiological traits, canopy structure, and atmospheric demand. This complexity makes it difficult to generalise globally across ecosystems without dynamic, observation-informed estimates.

Several methods have been proposed in the literature to estimate LFMC from space. One widely used approach exploits optical reflectance, particularly in the shortwave infrared (SWIR) and near-infrared (NIR) spectral regions, where water strongly modulates leaf and canopy reflectance. Indices such as the Normalized Difference Water Index (NDWI) or the Global Vegetation Moisture Index (GVMI) are typically calibrated against field LFMC observations using empirical or machine learning models [27, 33]. Optical methods offer relatively high spatial resolution (e.g., 500 metres from the Moderate Resolution Imaging Spectroradiometer (MODIS)), but they suffer from cloud contamination and signal saturation in dense canopies. Moreover, they are largely passive and may lag physiological changes in vegetation under stress.

An alternative class of methods relies on microwave remote sensing. In particular, vegetation optical depth (VOD), retrieved from passive microwave sensors at L-, C-, and X-band frequencies, provides a proxy for vegetation water content, integrating both biomass and internal moisture signals. Because microwave signals can penetrate clouds and are sensitive to vegetation water status even under dense canopies, they offer complementary information to optical sensors. Forkel et al. [18] produced a global LFMC dataset based on VOD, demonstrating its potential for large-scale, all-weather monitoring. However, such products are typically coarse in resolution (e.g., 0.25°) and require careful interpretation, as VOD is influenced by both water content and vegetation structure.

A third category integrates remote sensing with land surface modelling. For example, Joe et al. [25] developed a semi-empirical global LFMC estimation system in which LFMC is modelled as an asymptotic function of leaf area index (LAI)—one half of the total leaf area per unit surface area—and soil moisture, using biome-specific parameterisations. Their system is used to provide daily global LFMC fields at approximately 9 km resolution for 2010–2019. This represents a major step towards near-real-time fuel condition monitoring within numerical weather prediction systems.

Recent advances in Earth observation provide new opportunities to improve LFMC estimation through multi-sensor data assimilation. For vegetation structure and function, VOD and solar-induced chlorophyll fluorescence (SIF) [19] offer independent and complementary constraints on LAI. As shown in the Shapley analysis presented in D1 [17], LAI was found to be a key driver for both SIF and VOD-L, with some seasonal variations sensitive to soil moisture. Weston et al. [22] demonstrated the use of VOD at multiple frequencies (L-, X-, and C-band) to estimate LAI via cumulative distribution function (CDF) matching. This technique enables LAI to be inferred even under cloud cover and provides consistent temporal dynamics. Complementarily, de Rosnay et al. [10] explored the use of SIF to constrain

LAI evolution, leveraging SIF's sensitivity to photosynthetic activity. Unlike traditional reflectance indices, SIF can decline rapidly in response to water stress, potentially anticipating structural changes and enabling earlier detection of fuel drying.

For soil moisture (SM), Scipal et al. [24] used the Advanced SCATterometer (ASCAT) radar and pseudo-observations to improve root-zone moisture estimates within land surface models. ASCAT provides frequent, weather-independent observations of surface wetness, which, when assimilated, improve the realism of plant water supply estimates—a key driver of LFMC. Subsequent studies confirmed and extended these benefits. Draper et al. [15] demonstrated that assimilating ASCAT-derived surface soil moisture into a hydrological model significantly reduced root-zone dry bias and improved hydrological consistency over multi-year periods, outperforming traditional vegetation-based metrics.

This deliverable presents the methodology and outcomes of Task 2, which aims to implement the technical components required to integrate Earth Observation (EO) data into the ECMWF forecasting system, using LDAS, for the purpose of constraining fuel moisture content estimates. The overarching objective is to support the development of a high-resolution, observation-informed global LFMC reanalysis dataset.

Although the project centres on fuel reanalysis generation, which could, to some extent, be informed directly by satellite observations (for instance, using LAI CGLS for LAI estimates), integrating EO data with a physical model offers significant advantages. This data assimilation-based approach enables:

- the assessment of the added value of satellite-derived products in updating fuel variables,
- the generation of daily (instead of 10-daily) LAI products,
- enhanced real-time fire risk monitoring capabilities,
- accounting for balanced observations and model uncertainties,
- consistently improving vegetation and soil moisture estimates.

Building on the observation operators for VOD at L-band (VOD-L) and SIF established in Task 1, we use decision tree-based ML algorithms, specifically Extreme Gradient Boosting (XGBoost), as an observation operator to characterise the nonlinear relationships between satellite-derived variables and LFMC. These methods and their validation are detailed in Deliverable D1 ("Towards enhanced fire fuel estimation with satellite-derived predictive models," December 2024) [17], where feature importance analysis identified SIF and VOD-L as the most informative predictors of LFMC across diverse biomes.

Informed by these findings, we propose a novel data assimilation framework that dynamically integrates VOD-L and SIF to LAI, alongside ASCAT-derived surface SM to constrain modelled soil moisture. By jointly assimilating these independent and complementary EO datasets, the framework aims to improve both the accuracy and temporal responsiveness of LFMC estimation at the global scale. This integration enables the system to ingest near-real-time information on vegetation structure, physiological stress, and surface moisture dynamics. We hypothesise that this multi-sensor fusion will enhance our ability to capture the spatio-temporal variability of LFMC, particularly during transitional periods such as rapid vegetation drying, post-disturbance recovery, or heightened fire risk.

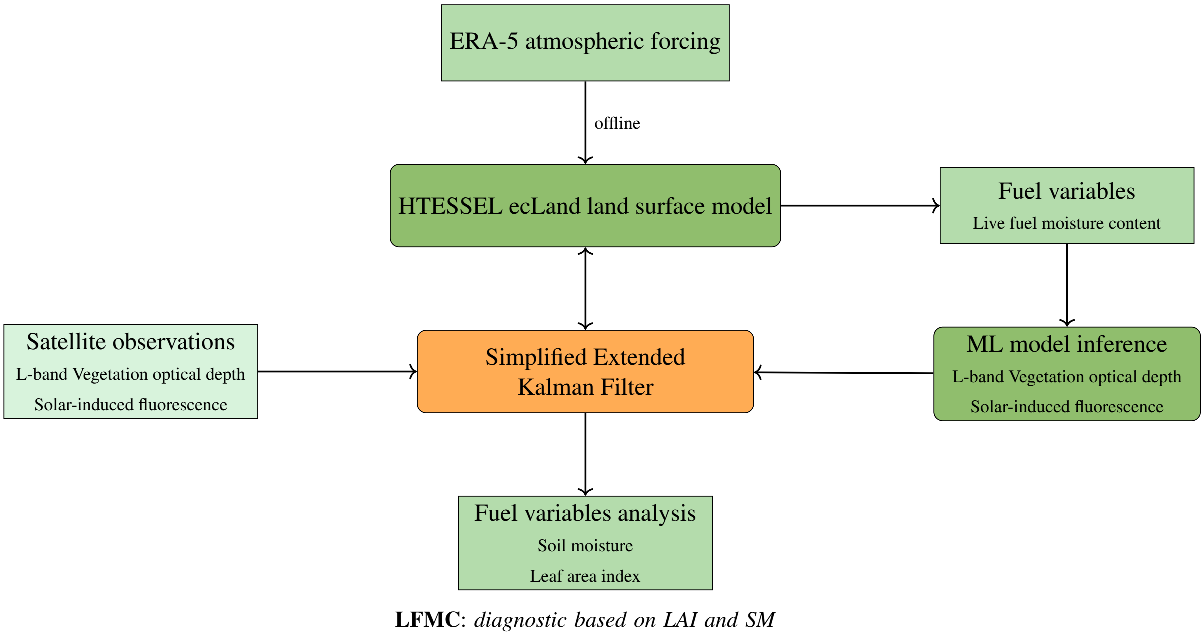

Figure 1: Schematic of the data assimilation system for LAI analysis: Satellite observations of VOD-L and SIF are assimilated using SEKF to update LAI, from which LFMC is diagnosed. A machine learning-based observation operator is used to map model variables to observation space.

2 Data and methodology

2.1 Overview of proposed framework

The goal of this study is to improve estimates of LFMC by correcting its key drivers: SM and LAI. To achieve this, we assimilate ASCAT soil moisture observations to update the SM state, and satellite-derived SIF and VOD-L observations to constrain LAI (Figure 1). The relationship between the vegetation observations and LAI is captured using machine learning–based observation operators developed in Deliverable D1 [17]. These operators simulate the model equivalent of the observation (i.e., what the satellite should have observed given the background state). For SIF assimilation, for example, the background is the LAI climatology, and the observation operator computes the expected SIF value from that LAI background. The innovation is the difference between the actual satellite observation and the simulated observation from the operator. Corrections are applied using the SEKF [12], which combines information from the background (e.g., LAI climatology or IFS soil moisture) and the observations, weighted by their respective uncertainties. The result is an updated (analysis) estimate that reflects the optimal combination of model and observation information. This correction is then applied to improve the estimates of the parameters, LAI and SM, thereby updating the LFMC.

An offline implementation of LDAS is used [28]. This setup is uncoupled, meaning it ingests ERA5 reanalysis data [20], specifically atmospheric forcing, without feedback to the atmosphere. While the one-way coupling limits interaction between the land and atmosphere, it allows for efficient, low-cost testing of developments over multi-year periods prior to integration into the coupled IFS LDAS [3].

Data assimilation was run for 2021, a year not seen by the machine learning observation operator. A coarse-resolution grid (TCo319, ∼36 km) was used to verify that the system setup works as expected.

Daily land surface analyses were produced at 00 UTC.

2.2 Data

2.2.1 Description of input datasets

Live fuel moisture content [25]: The initial dataset used to develop and calibrate the LFMC model was derived from Globe-LFMC [32], which compiles destructive sampling measurements collected between 2010 and 2018 across 1,383 sites in 11 countries. In total, 161,717 individual destructive sampling records were collected. These in situ measurements encompass a range of vegetation types and involve physically weighing live plant material before and after oven drying to calculate gravimetric water content. Given the substantial variability in LFMC across space and time, site-level observations were aggregated into monthly means to reduce sampling noise, resulting in 25,410 aggregated samples. These samples were used to train and constrain a semi-empirical model to estimate LFMC globally at a daily resolution of approximately 9 km from 2010 to 2019. The model assumes that LFMC responds asymptotically to SM and LAI, both of which serve as proxies for water availability and vegetation status. Parameters were optimised using a trust-region reflective algorithm, fitting the following equation:

$$\mathrm{LFMC} = \mathrm{LFMC}_{\max} - A \cdot e^{-(\alpha \cdot \mathrm{LAI} + \beta \cdot \mathrm{SM} + \gamma \cdot \mathrm{LAI} \cdot \mathrm{SM})} \tag{1}$$

Here, $\mathrm{LFMC}_{\max}$ represents the maximum attainable moisture content for a given vegetation type, and A denotes the range between the minimum and maximum LFMC. The coefficients α, β, and γ describe vegetation-specific sensitivity to LAI, SM, and their interaction, respectively—interpreted as proxies for dormancy, drought resistance, and active growth response.

Seven parameter sets were derived for major vegetation functional types (crops, short grass, evergreen needleleaf, deciduous broadleaf, mixed crops/grassland, deciduous shrubs, and broadleaf savannah), using subsamples of the Globe-LFMC data. Other types were assigned parameters from the closest analogue. LAI was sourced from the CONFESS dataset [7] at a monthly resolution of approximately 9 km, while SM was extracted from ERA5-Land [26], with values aggregated by rooting-depth distributions from ecLand [8].

Vegetation optical depth at L-band [1]: It provides information on vegetation water content and biomass by measuring the attenuation of microwave radiation as it passes through vegetation canopies. This is based on L-band (1.4 GHz; ∼21 cm wavelength) passive microwave observations from the SMOS satellite, which was launched by the European Space Agency in 2009. The L-band frequency is particularly well-suited for penetrating dense canopies and capturing signals from both vegetation and upper soil layers, making VOD-L especially valuable for monitoring woody biomass and structural changes in vegetation. The dataset has a spatial resolution of approximately 25 km and is available at a daily temporal resolution, allowing for consistent and frequent monitoring of vegetation dynamics on a global scale. Unlike optical sensors, L-band microwaves are largely unaffected by cloud cover or solar illumination, enabling consistent global coverage.

Solar-induced fluorescence [19] We use the TROPOSIF dataset, a global gridded solar-induced chlorophyll fluorescence (SIF) product derived from the TROPOMI instrument onboard the Sentinel-5 Precursor satellite. The TROPOSIF product is generated by ESA's pre-operational PAL framework and provides 8-day composites at 0.1° spatial resolution, which differs from TROPOMI's native pixel resolution (∼3.5 × 5.5 km). TROPOSIF is based on TROPOMI observations but aggregated temporally and spatially to improve signal stability and reduce noise. The dataset is publicly available at: https://data-portal.s5p-pal.com/products/troposif.html.

SIF is emitted by chlorophyll a during the de-excitation of absorbed light energy and is primarily observed in the red (∼687 nm) and far-red (∼740 nm) spectral regions. It serves as a direct proxy for photosynthetic activity, offering physiological insights into vegetation status that cannot be captured by traditional reflectance-based indices. The satellite-derived SIF signal—particularly at the coarse spatial and temporal resolution of TROPOSIF—is largely driven by vegetation structural properties such as LAI.

Soil moisture derived from the advanced scatterometer [12]: It represents the relative soil moisture index in the top approximately 2 cm of the soil surface and is derived from C-band (5.255 GHz) radar backscatter measurements. As an active microwave sensor, ASCAT transmits its own radar pulses and measures the signal reflected back from the Earth's surface. The strength of this backscatter is sensitive to the dielectric properties of the soil, which vary significantly with moisture content. The more water present in the soil, the higher the dielectric constant, resulting in a stronger return signal. ASCAT provides global soil moisture data with a spatial resolution of about 25 km and daily to sub-daily temporal coverage, depending on latitude. The product is distributed as a relative soil moisture index (in %), scaled between the historically driest and wettest conditions observed at each location, rather than representing absolute volumetric soil moisture.

2.2.2 Quality control

Prior to assimilation, VOD-L and SIF observations were filtered to exclude cases where data values fell outside their physical bounds, such as values below zero. Frozen surfaces were also screened out using a surface temperature threshold of 273 K, along with flags indicating the presence of snow or ice. Data were removed over unfavourable surface conditions, including snow cover, steep or complex terrain, and water bodies. Within the assimilation system, additional quality control procedures were implemented to reject observations that deviated excessively from the model background.

2.2.3 Validation data

The evaluation conducted in this study focuses on the year 2021. For vegetation, we use **the CGLS LAI 300 m product** [30], which provides 10-day composites at a 300 m resolution. The LAI product used in this study was developed within the framework of the Copernicus High-Resolution Vegetation Biophysical Variables for the Copernicus Global Land Service (CORSO) project, which aims to enhance the quality, timeliness, and consistency of biophysical vegetation products for near-real-time monitoring. In the CORSO project, the consolidated RT6 version—refined after six dekads—was produced and resampled to an 8-day frequency to align with the temporal resolution of SIF observations. This version was also used in the present study.

Since July 2020, imagery from Sentinel-3 OLCI has replaced PROBA-V. The CONFESS project's LAI climatology [7] relied on GEOV2 [5], which combines SPOT-VGT (1999–2014) and PROBA-V (2014–2019). Thus, the 2021 observations explored here differ from the CONFESS climatology due to both interannual variability and the change in the observing system (Sentinel-3 vs PROBA-V).

For soil moisture, we used **in situ observations from the International Soil Moisture Network** (ISMN; [14]) presented in Figure 2, which compiles quality-controlled point measurements from multiple regional and national monitoring networks. These ground-based observations serve as a reference standard for assessing the accuracy of satellite- and model-derived surface soil moisture estimates.

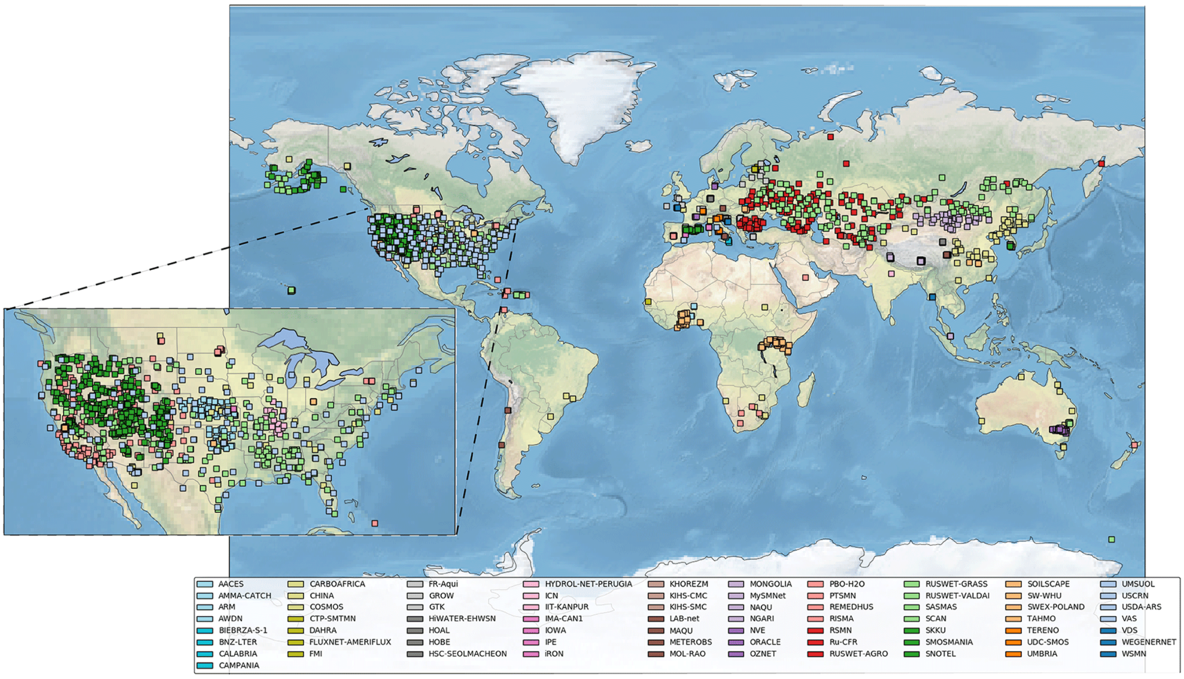

Figure 2: Locations of ISMN networks (status in July 2021), from [14].

2.3 Methodology

2.3.1 ECMWF land surface model - ecLand

The ECMWF land-surface modelling system, ecLand, is based on the HTESSEL (Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land) model [3, 8] which represents vertical movements of soil moisture using the equations of Richards (1931). The soil column is discretised into four layers with thicknesses of 0.07, 0.21, 0.72 and 1.89m from top to bottom. The vegetation representation in ecLand relies on a tile approach that accounts for dominant low (grassland, crop, shrubland) and high (forest) vegetation. In the current version of ecLand used in the IFS, vegetation parameters such as LAI are specified as seasonally varying climatological monthly mean maps in the ECMWF Numerical Weather Prediction (NWP) system. This climatology uses the latest CGLS LAI dataset (1993-2019) from the CONFESS project [7] and has been shown to have a significant impact on quality.

One of the main weaknesses of the current approach is that inter-annual variability in vegetation is not taken into account. Year-to-year differences in vegetation can be significant, driven most notably by land use change, as well as by fire events, meteorological extremes such as droughts or above-average rainfall, and fluctuations in 2-metre air temperature. These variations have become even more pronounced in recent years due to the effects of climate change.

Soil moisture estimates could benefit from incorporating dynamic vegetation indicators like LAI, which can help update soil moisture by providing more accurate, seasonally relevant vegetation states. These differences tend to be especially large during the transition seasons—spring and autumn—when vegetation characteristics change most rapidly as part of the annual cycle. This is particularly true in mid-latitude regions, where seasonal conditions vary the most.

2.3.2 Simplified extended Kalman filter

The Kalman Filter and its extended version (EKF) are sequential DA methods based on least-squares estimation. In this framework, the *background* is a short-range forecast generated by propagating the previous *analysis* forward in time using the forecast model $\mathscr{M}$. As new observations become available, they are sequentially incorporated to correct the background and produce an improved estimate known as the *analysis*. This makes the Kalman framework particularly well suited for real-time assimilation.

The SEKF implemented at ECMWF and initially developed for soil moisture analysis, operates as a point-wise data assimilation scheme. Following the notation of [21], the analysed soil moisture state vector $\mathbf{x}^a$ is computed at time t_i for each grid point as:

$$\mathbf{x}^a(t_i) = \mathbf{x}^b(t_i) + \mathbf{K}_i[\mathbf{y}^o(t_i) - \mathscr{H}_i(\mathbf{x}^b)] \tag{2}$$

with superscripts a, b, o standing for analysis, background and observations, respectively. $\mathbf{x}$ is the model state vector, $\mathbf{y}$ is the observation vector and $\mathscr{H}$ the nonlinear observation operator.

The *increment* in data assimilation is defined as the difference between the analysis and the background:

$$\delta\mathbf{x} = \mathbf{x}^a - \mathbf{x}^b \tag{3}$$

This increment represents the adjustment applied to the background, based on the innovation (i.e., the difference between observations and their model equivalents), weighted by the respective uncertainties of the model and observations.

The Kalman gain matrix $\mathbf{K}_i$ is computed at time t_i as:

$$\mathbf{K}_i = [\mathbf{B}^{-1} + \mathbf{H}_i^T\mathbf{R}^{-1}\mathbf{H}_i]^{-1}\mathbf{H}_i^T\mathbf{R}^{-1} \tag{4}$$

where $\mathbf{H}_i$ is the linearised observation operator, $\mathbf{B}$ is the approximate background error covariance matrix associated with $\mathbf{x}$ and $\mathbf{R}$ is the observation errors covariance matrix.

The linearisation of the observation operator is approximated using finite differences, by applying a small perturbation δx_n to the n^{th} component of the model state vector and evaluating the resulting change in the observation space. One perturbed simulation is required for each element of the control state vector. For each perturbed simulation, the initial background state vector is perturbed by a vector $\delta\mathbf{x}_n^b$ that contains δx_n for the perturbed n^{th} element and zero for all the other elements. Using index m to represent the m^{th} element of the observations vector, the Jacobian elements $\mathbf{H}_{mn,i}$ of the observation operator at time t_i can be expressed as:

$$\mathbf{H}_{mn,i} = \frac{\mathscr{H}_{m,i}(\mathbf{x}^b + \delta\mathbf{x}_n^b) - \mathscr{H}_{m,i}(\mathbf{x}^b)}{\delta x_n} \tag{5}$$

The model state vector evolution from time t_i to time t_{i+1} is then defined as:

$$\mathbf{x}^b(t_{i+1}) = \mathscr{M}_i[\mathbf{x}^a(t_i)] \tag{6}$$

with $\mathscr{M}$ the nonlinear forecast model.

2.3.3 Experiment description

In the system implemented for this study, we use an offline DA setup, meaning that the analysis corrections do not feed back into the atmospheric forcing [11]. The state vector includes soil moisture from the top three soil layers ($SM_{l1:3}$) and LAI. For the computation of the LFMC analysis, increments of LAI and the three top layers were applied; no increment was applied to the deepest soil layer (SM_{l4}).

The observation vector, background state vector, and observation operator are structured as follows:

$$\mathbf{y^o} = \begin{bmatrix} 2mT \\ 2mRH \\ ASCAT \\ VOD_L \\ SIF \end{bmatrix} \qquad \mathbf{x}^b = \begin{bmatrix} SM_{l1} \\ SM_{l2} \\ SM_{l3} \\ LAI \end{bmatrix} \qquad \mathscr{H}(\mathbf{x}^b) = \begin{bmatrix} 2mT \\ 2mRH \\ SM_{top} \\ VOD_{L,b} \\ SIF_b \end{bmatrix} \tag{7}$$

We use the notation VOD_L instead of "VOD-L" in mathematical formulations to avoid confusion with a minus sign.

In Equation 7, variables highlighted in orange represent the existing observations and model states already assimilated in the system. Specifically, 2-metre temperature (2mT) and 2-metre relative humidity (2mRH) pseudo-observations, as well as ASCAT soil moisture observations, were assimilated to improve soil moisture estimates in the top three model layers. Variables highlighted in purple correspond to the newly assimilated datasets introduced in this study, specifically VOD-L and SIF, to improve the representation of LAI.

As described in [16], the background error covariance matrix (**B**) and the observation error covariance matrix (**R**) were assumed to be static and diagonal, with diagonal terms defined by error variances (σ_b^2 for background errors and σ_o^2 for observation errors). The values in Table 1 represent the final configuration retained for the assimilation system. They were selected based on sensitivity experiments in which background and observation standard deviations were systematically varied to assess their impact on assimilation performance. The tested values are provided in Table A.1 in the annex, and the corresponding analysis increments, zoomed over Europe, are shown in Figure A.1.

A larger observation error (σ_o) results in smaller analysis increments, giving more weight to the background, while smaller observation errors increase the influence of observations. Similarly, larger background errors (σ_b) increase the weight of observations in the analysis.

Table 1: Standard deviation values (σ_b for background errors and σ_o for observation errors) retained for this study. These values define the diagonal terms of the background error covariance matrix (**B**) and the observation error covariance matrix (**R**).

Variable	Standard Deviation	Unit
Background variables (σ_b)		
Soil moisture (top three layers)	0.01	$m^3\ m^{-3}$
Leaf area index (LAI)	1	$m^2\ m^{-2}$
Observation variables (σ_o)		
2-metre temperature	1	K
2-metre relative humidity	4	%
ASCAT soil moisture	0.05	$m^3\ m^{-3}$
VOD-L	0.16	unitless
SIF	0.12	$mW\ m^{-2}\ sr^{-1}\ nm^{-1}$

Background errors for soil moisture, 2-metre temperature, 2-metre relative humidity, and ASCAT soil moisture follow the statistics reported in [24, 12]. The LAI background error standard deviation was set to a static value of 1 $m^2\,m^{-2}$ based on sensitivity analysis (see Annex section A) and is consistent with literature reporting typical static LAI errors in the 0.4–1.2 $m^2\,m^{-2}$ range [29, 2, 4, 10].

LAI was assimilated against a climatological background representing the mean seasonal cycle, which does not capture interannual or synoptic variability. Its associated uncertainty was intentionally inflated toward the upper range of literature values to reflect spatial heterogeneity, model structural limitations, and observation–model mismatches. Assigning a larger background error allows the analysis to place greater weight on observational information.

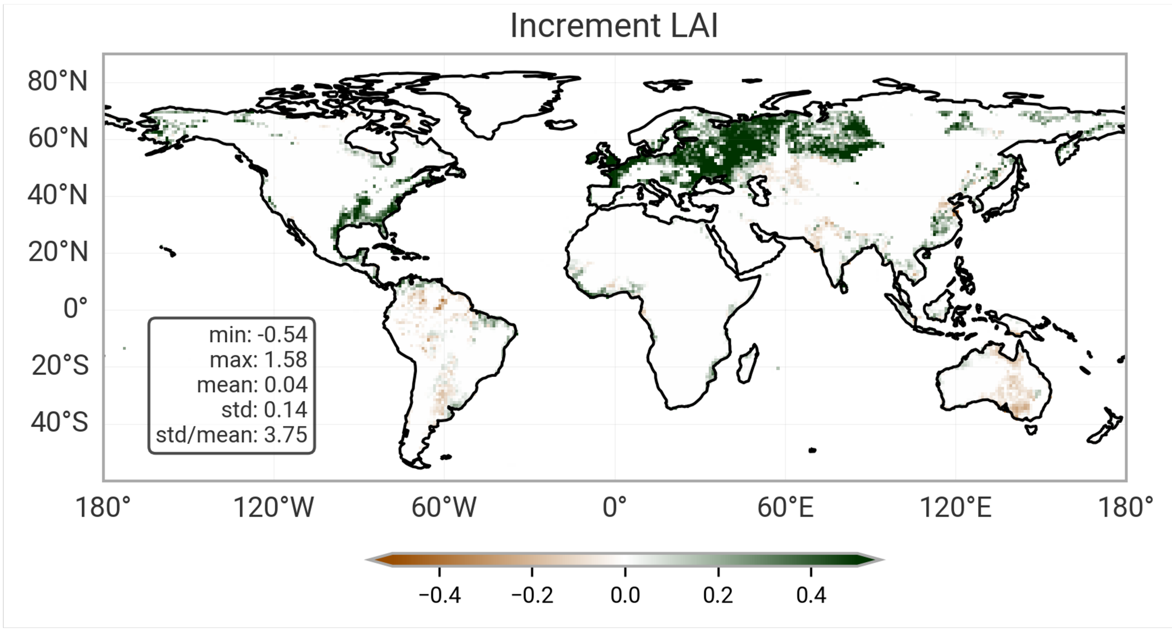

Figure 3: Spatial distribution of LAI analysis increments for August 2021, derived from the data assimilation system using climatological background LAI. Positive values (green) indicate regions where the assimilation increased LAI relative to the background, while negative values (brown) show where LAI was reduced. Increments are derived from the assimilation of satellite-based SIF and VOD-L observations. White areas correspond to locations with no overlapping SIF and VOD-L observations during this period. The inset box provides global statistics: minimum, maximum, mean, standard deviation, and the ratio of standard deviation to mean of the increments.

Observation error standard deviations for SIF and VOD-L were derived from O–B statistics over the year 2021. Desroziers diagnostics [13] applied to single-observation-type assimilation experiments indicated similar error magnitudes despite imperfect convergence.

The Jacobian matrix **H** was computed via finite differences as follows:

$$\mathbf{H} = \begin{bmatrix} \frac{T_{2m\_pert1} - T_{2m}}{\delta SM_{l1}} & \frac{T_{2m\_pert2} - T_{2m}}{\delta SM_{l2}} & \frac{T_{2m\_pert3} - T_{2m}}{\delta SM_{l3}} & \frac{T_{2m\_pert4} - T_{2m}}{\delta LAI} \\ \frac{RH_{2m\_pert1} - RH_{2m}}{\delta SM_{l1}} & \frac{RH_{2m\_pert2} - RH_{2m}}{\delta SM_{l2}} & \frac{RH_{2m\_pert3} - RH_{2m}}{\delta SM_{l3}} & \frac{RH_{2m\_pert4} - RH_{2m}}{\delta LAI} \\ \frac{SM_{l1\_pert1} - SM_{l1,ASCAT}}{\delta SM_{l1}} & \frac{SM_{l1\_pert2} - SM_{l1,ASCAT}}{\delta SM_{l2}} & \frac{SM_{l1\_pert3} - SM_{l1,ASCAT}}{\delta SM_{l3}} & \frac{SM_{l1\_pert4} - SM_{l1,ASCAT}}{\delta LAI} \\ \frac{VOD_{L\_pert1} - VOD_{L}}{\delta SM_{l1}} & \frac{VOD_{L\_pert2} - VOD_{L}}{\delta SM_{l2}} & \frac{VOD_{L\_pert3} - VOD_{L}}{\delta SM_{l3}} & \frac{VOD_{L\_pert4} - VOD_{L}}{\delta LAI} \\ \frac{SIF_{\_pert1} - SIF}{\delta SM_{l1}} & \frac{SIF_{\_pert2} - SIF}{\delta SM_{l2}} & \frac{SIF_{\_pert3} - SIF}{\delta SM_{l3}} & \frac{SIF_{\_pert4} - SIF}{\delta LAI} \end{bmatrix}$$

Following [16], the soil moisture perturbations ($\delta SM_{l1:3}$) were set to 0.01 m^3m^{-3}. Similarly, the LAI perturbation (δLAI) was set to 0.01 $m^2 m^{-2}$ based on sensitivity analysis.

In this study, the assimilation of VOD-L and SIF was intentionally restricted to updating only the LAI component of the state vector, with their sensitivities (Jacobians) with respect to soil moisture set to zero. Furthermore, increments were applied only when both VOD-L and SIF observations were simultaneously available. This approach ensures that vegetation updates are based on coherent multi-sensor information, thereby minimising the risk of assimilating inconsistent or insufficiently constrained data.

3 Results

3.1 Multi-sensor data assimilation

In the experiment shown in Figure 3, SIF and VOD-L were jointly assimilated to constrain LAI, with increments applied only when both observations were simultaneously available to ensure consistent, co-located vegetation updates. In parallel, ASCAT soil moisture, 2-metre air temperature, and relative humidity pseudo-observations were assimilated to inform the soil moisture state, following the error specifications presented in Table 1.

Green areas indicate increases in LAI driven by the observational signal, while brown areas reflect reductions. Regions shaded white correspond to areas where no simultaneous SIF and VOD-L data were available during the month, and therefore no update was applied. The inset box provides summary statistics: a global mean increment close to zero (0.04 m^2 m^{-2}), with values ranging from -0.54 to 1.58 m^2 m^{-2}, and a relatively high ratio of standard deviation to the mean (3.75), highlighting the spatially heterogeneous nature of the corrections. The corrections introduced by data assimilation are not uniform. In northern temperate regions, such as western Russia, parts of central and eastern Europe, and the eastern United States, LAI was consistently adjusted upwards, suggesting that vegetation activity during August 2021 exceeded what was captured in the long-term climatological baseline. In contrast, significant reductions in LAI were noted over parts of South America and Australia, regions known to have experienced dry anomalies during the same period [6]. However, large swathes of the globe remain observationally unconstrained, particularly in arid and high-latitude regions where either SIF or VOD-L—or both—are missing or filtered (see section 2.2.2).

The effectiveness of the LAI assimilation system was evaluated by comparing analysis increments against the CGLS 300 m LAI product. Figure 4 shows the spatial distribution of changes in RMSE (ΔRMSE) between the climatological background and assimilation analysis, computed over two months: May and August 2021. Across both months, large areas exhibit reduced RMSE (shaded red), indicating that the assimilation consistently improved the agreement with the independent reference dataset. These improvements are particularly notable over temperate regions such as central and eastern Europe, parts of eastern North America, and southeastern South America. In contrast, some regions—particularly in the tropics and parts of Australia—show small, localised increases in RMSE (shaded red), which may reflect unresolved surface heterogeneity, limitations in the observation operator in those regions, or the need for further tuning of the assimilation system.

Assimilating ASCAT-derived surface soil moisture along with 2 m temperature and relative humidity had a limited impact on modelled soil moisture. The spatial evaluation (Figure 5, top panel) reveals that, for most stations, changes in correlation between the model and observations were not statistically significant. Improvements and degradations were largely balanced, with only a few stations showing

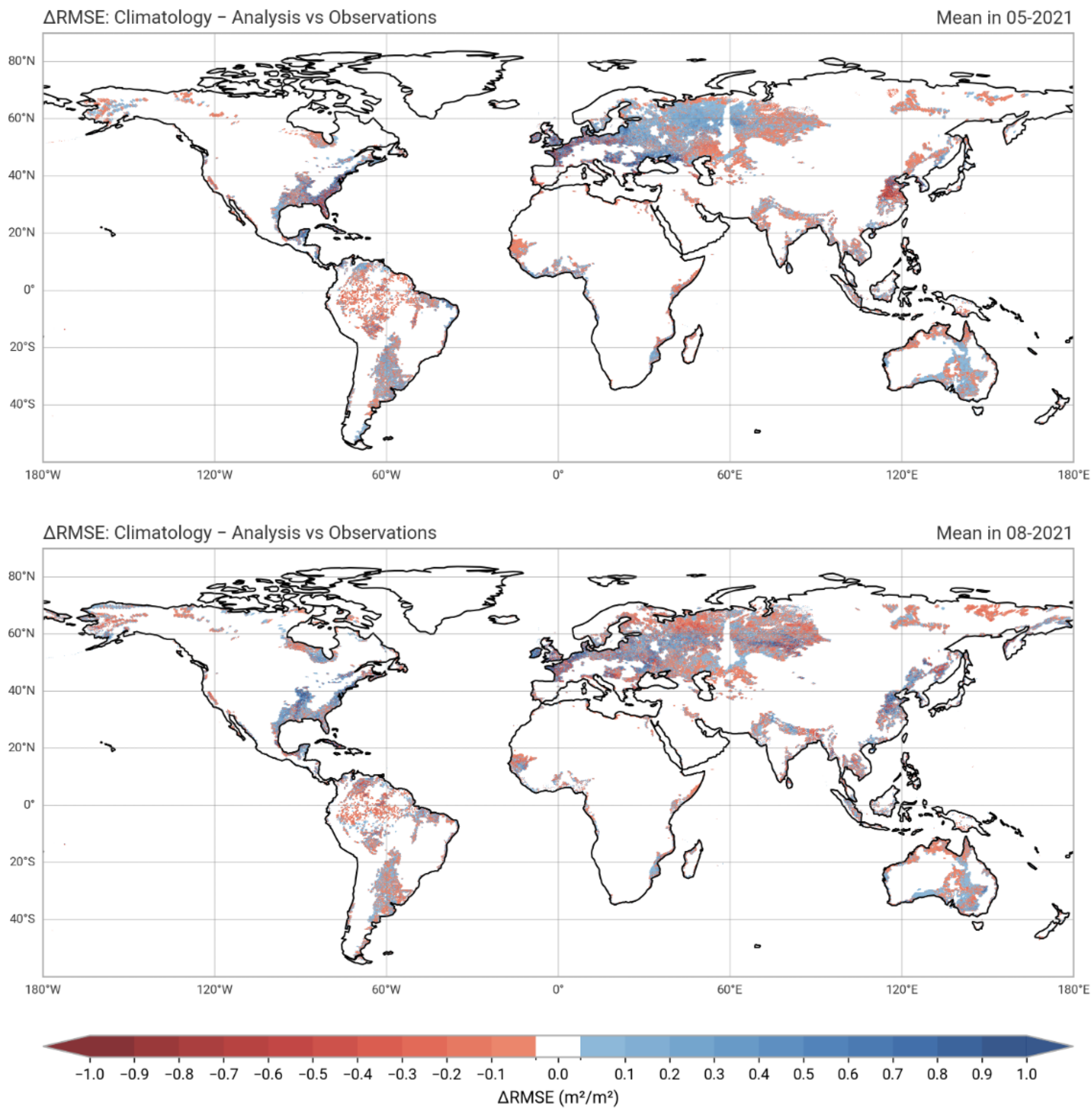

Figure 4: Spatial distribution of the change in root-mean-square error (ΔRMSE) between model estimates and satellite observations of LAI for May (top) and August (bottom) 2021. ΔRMSE is calculated as RMSE(climatology, observations) minus RMSE(analysis, observations), where observations refer to the Copernicus Global Land Service 300 m LAI product. Positive values (blue) indicate improved agreement with observations due to assimilation, while negative values (red) reflect degraded performance. The analysis is limited to grid cells where assimilation increments were applied and valid CGLS data were available.

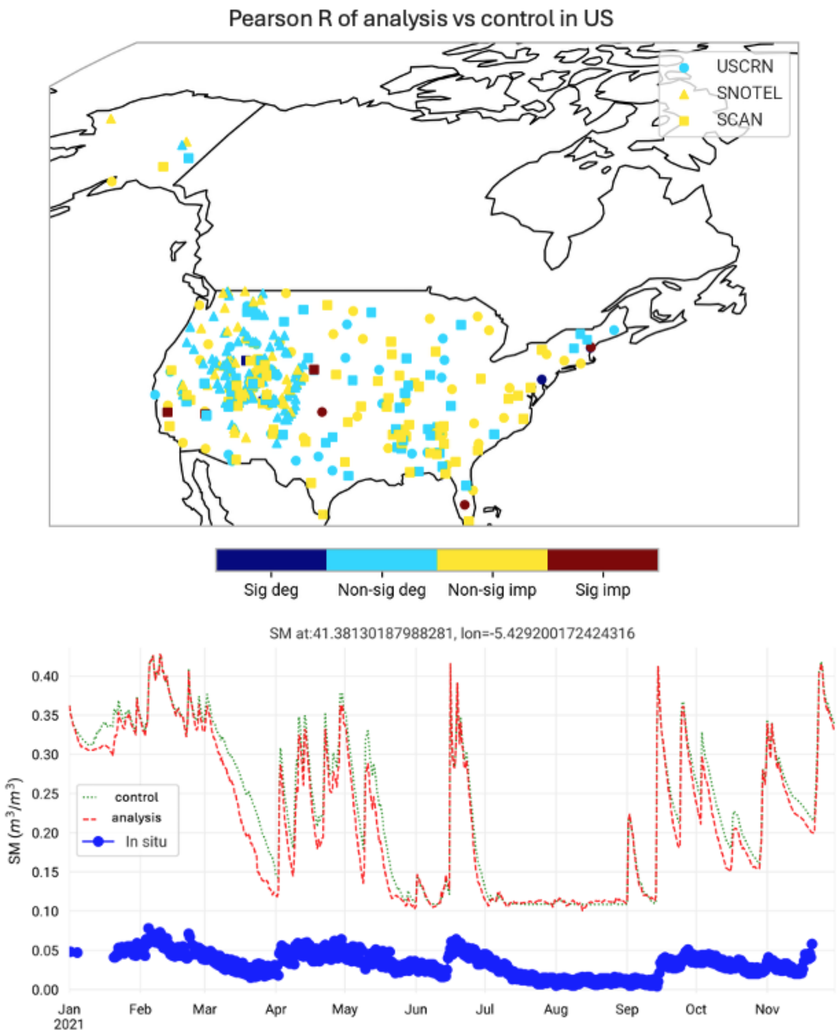

Figure 5: Evaluation of surface soil moisture against ISMN in situ measurements across the continental United States in 2021. The top panel shows the spatial distribution of changes in Pearson correlation (R) between the model and in situ observations when ASCAT surface soil moisture, 2 m air temperature, and 2 m relative humidity were assimilated (analysis) compared to an open-loop control experiment without data assimilation. Stations from the USCRN, SNOTEL, and SCAN networks are included. Colours indicate whether the change in correlation is significant and in which direction: significant degradation (dark blue), non-significant degradation (light blue), non-significant improvement (yellow), and significant improvement (brown). The bottom panel presents time series of daily soil moisture at an example SCAN site in Iowa (Midwestern region of the United States) comparing the analysis (green), control (red), and in situ observations (blue points).

significant changes—either positive or negative. This suggests that the assimilation configuration, while stable, did not systematically enhance skill relative to the open-loop simulation.

The time series at a representative site in Iowa (Figure 5, bottom panel) highlights a typical scale mismatch issue when comparing local in situ soil moisture measurements to grid-scale model estimates. This example illustrates that part of the assimilation performance assessment is inherently influenced by representativeness errors, as the model integrates soil moisture over much larger spatial domains than the point-scale sensors capture. The persistent bias observed at this site is consistent with this known issue and demonstrates the challenge of directly validating coarse-resolution soil moisture simulations with point-based observations.

Despite the potential value of simultaneously assimilating both VOD-L and SIF, the current setup disregards potentially informative signals due to inconsistencies in data availability within a single assimilation cycle. The lack of synergy between the datasets likely contributes to the observed degradations in performance, as indicated by the positive RMSE differences in Figure 4, where red regions indicate areas where the LAI analysis performed worse than the climatology when compared to the observations. To address these shortcomings, a post-processing strategy was implemented to correct biases in the LAI increments.

3.2 LAI post-processing

The LAI post-processing step was introduced to correct systematic biases in the LAI analysis increments generated by the assimilation system. In its current configuration, the system applies increments only in areas where both VOD-L and SIF observations are simultaneously available within the same assimilation cycle. This conservative choice was made to ensure that LAI updates are supported by co-located, multi-sensor evidence. However, this constraint limits the spatial and temporal extent of the updates, leading to spatial inconsistencies, particularly in regions with sparse satellite coverage.

To address this issue, a post-processing correction was implemented to produce more consistent and unbiased LAI increments across space and time.

Although one alternative would be to pre-fill satellite data gaps prior to assimilation—through temporal interpolation, collocation analysis, or model-based gap-filling—this approach risks smoothing or distorting true dynamics, particularly under extreme conditions or in data-sparse regions. Given these trade-offs, the post-assimilation correction offers a conservative solution. It preserves the raw assimilation structure while addressing systematic deviations, ultimately enhancing the realism and usability of the reanalysis product.

3.2.1 Correction of LAI increments

The correction combines predictors from the standalone assimilation experiments—specifically, the raw increments from the VOD-L-only and SIF-only experiments—alongside phenological and observational information. This includes leaf onset timing [9], which governs the seasonal trajectory of LAI, as well as the most recent valid LAI observation from the CGLS 300 m product.

Leaf onset is particularly valuable as it captures the timing of vegetation green-up, a key phenological transition that strongly influences LAI evolution throughout the season. An ML model was trained to predict corrected, observation-consistent LAI increments using these predictors, thereby improving the overall coherence of the LAI analysis.

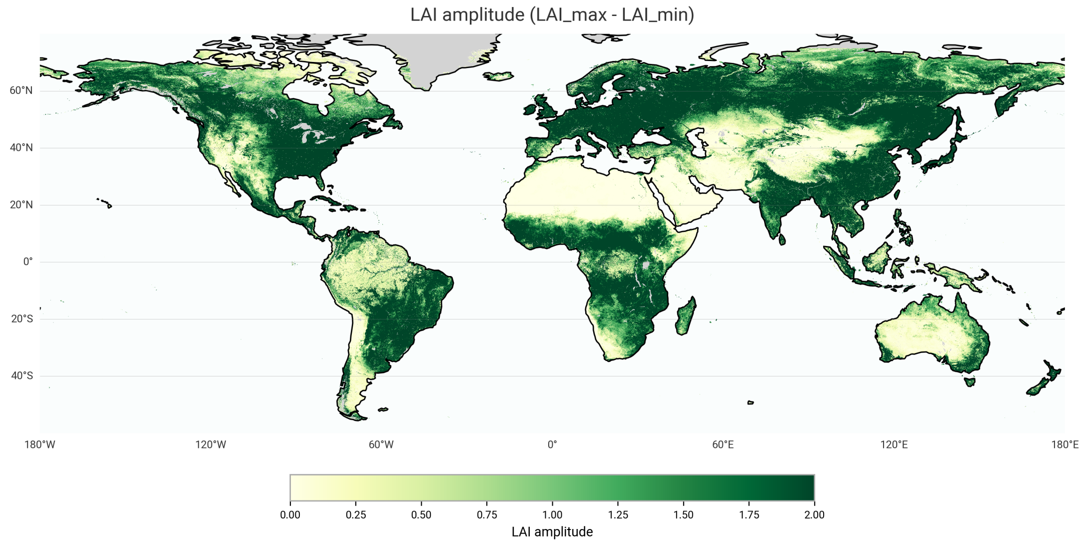

Figure 6: LAI seasonal amplitude, derived from CGLS observations, computed as the difference between the maximum and minimum LAI values over the 2021 growing season. Higher amplitude values (dark green) indicate ecosystems with strong seasonal variability, typically found in temperate and boreal forests, croplands, and semi-arid regions. Areas with low amplitude (yellow) correspond to evergreen tropical forests or sparsely vegetated regions with minimal seasonal LAI change.

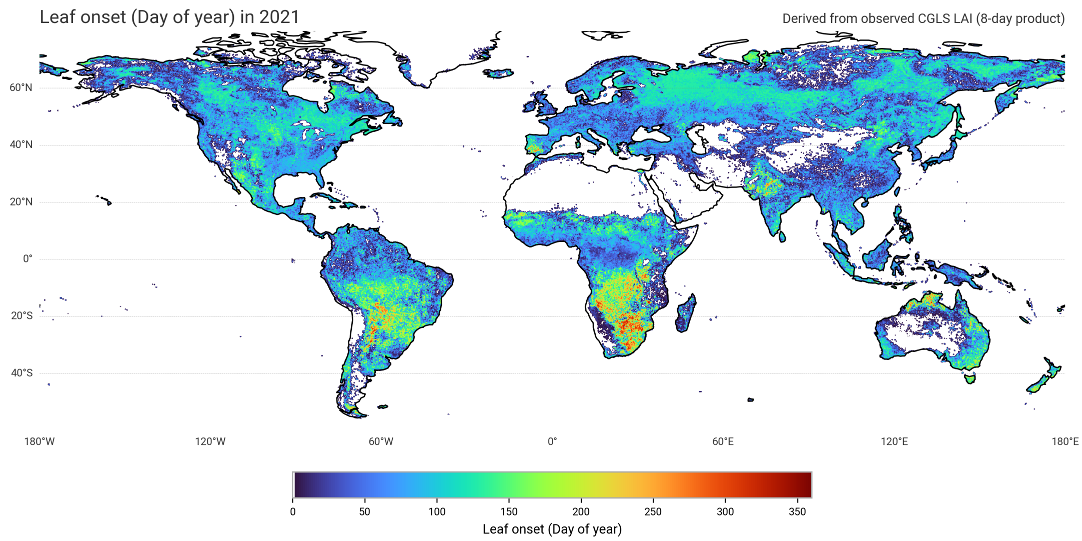

Figure 7: Leaf onset timing in 2021, expressed as the day of year (DOY) when the most significant seasonal increase in LAI occurred. This product is derived from the smoothed daily CGLS LAI dataset using the curvature-based method of Brut et al. (2009) [9]. Areas with no meaningful seasonal signal—defined as locations where the LAI amplitude was below 0.2 m^2 m^{-2}—were masked.

The seasonal amplitude of LAI (Figure 6) provides an essential diagnostic of vegetation phenology and productivity. High-amplitude regions, such as parts of central North America, eastern Europe, and northeastern China, reflect ecosystems with strong seasonal leaf development and senescence, where accurate timing of LAI onset is critical for capturing vegetation dynamics. Conversely, low-amplitude areas—such as the tropics and high-latitude tundra—exhibit minimal intra-annual variation, suggesting that LAI remains relatively stable year-round or that seasonal changes are less pronounced in the satellite record.

Building on this, the derived leaf onset map (Figure 7) identifies the timing of spring green-up using the method of Brut et al. (2009) [9], which detects the inflection point of the LAI time series by analysing the second derivative of a fitted seasonal curve. Only locations with meaningful seasonal variability (amplitude > 0.2 m^2 m^{-2}) were retained, as changes below this threshold were considered indistinguishable from noise or climatologically stable. In such cases, the background LAI is unlikely to require significant correction, and assimilation-induced changes would have limited impact. The resulting onset map captures broad latitudinal gradients, with earlier onset in subtropical and temperate zones and later onset in boreal regions. This phenological signal serves as a valuable input in our bias correction model, helping to align assimilation increments with the expected seasonal trajectory of vegetation growth, as shown in Figure 8.

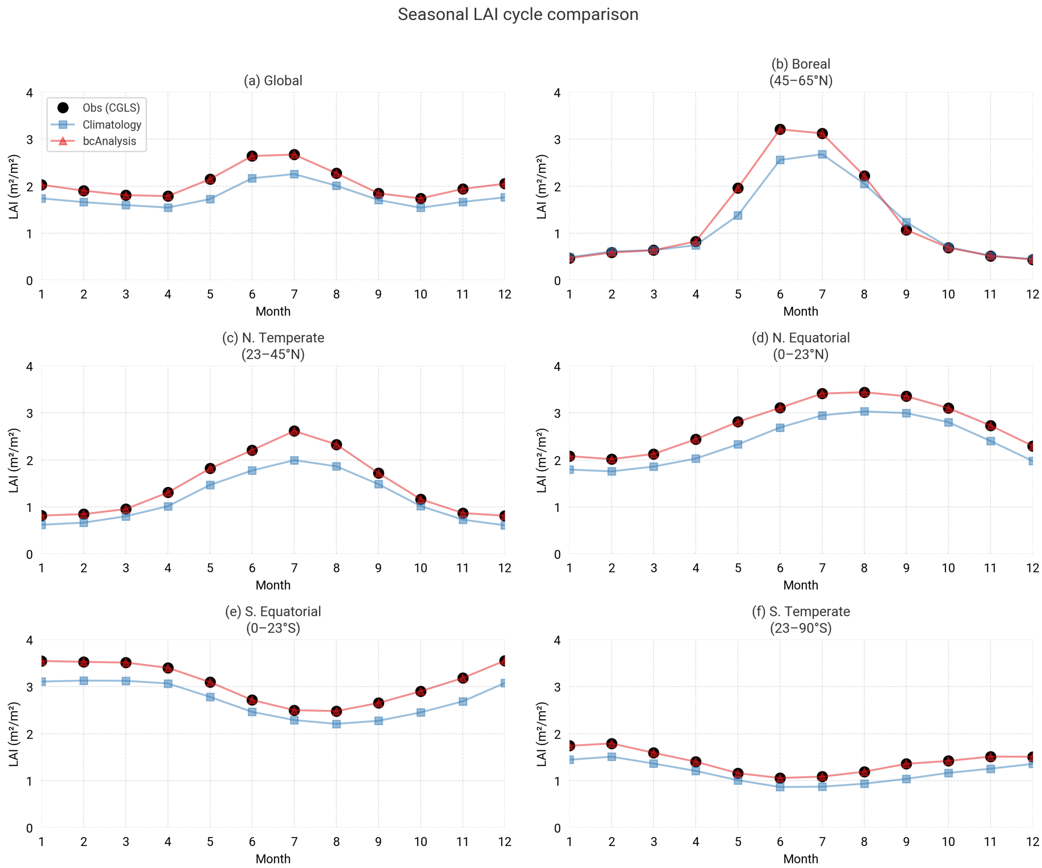

Figure 8: Seasonal cycles of LAI across different latitudinal zones in 2021 are compared. Observations from the CGLS LAI 300 m product (black dots) are evaluated against the climatological LAI background (blue line) and the bias-corrected analysis (bcAnalysis) produced through post-processing (red line; see Section 3.2 for details). The panels present the global mean (a) and zonal averages for the Boreal (b), Northern Temperate (c), Equatorial (d), Southern Equatorial (e), and Southern Temperate (f) regions. The post-processing step enhances the seasonal agreement between the analysis and the observed LAI cycle.

3.2.2 Impact on LFMC estimates

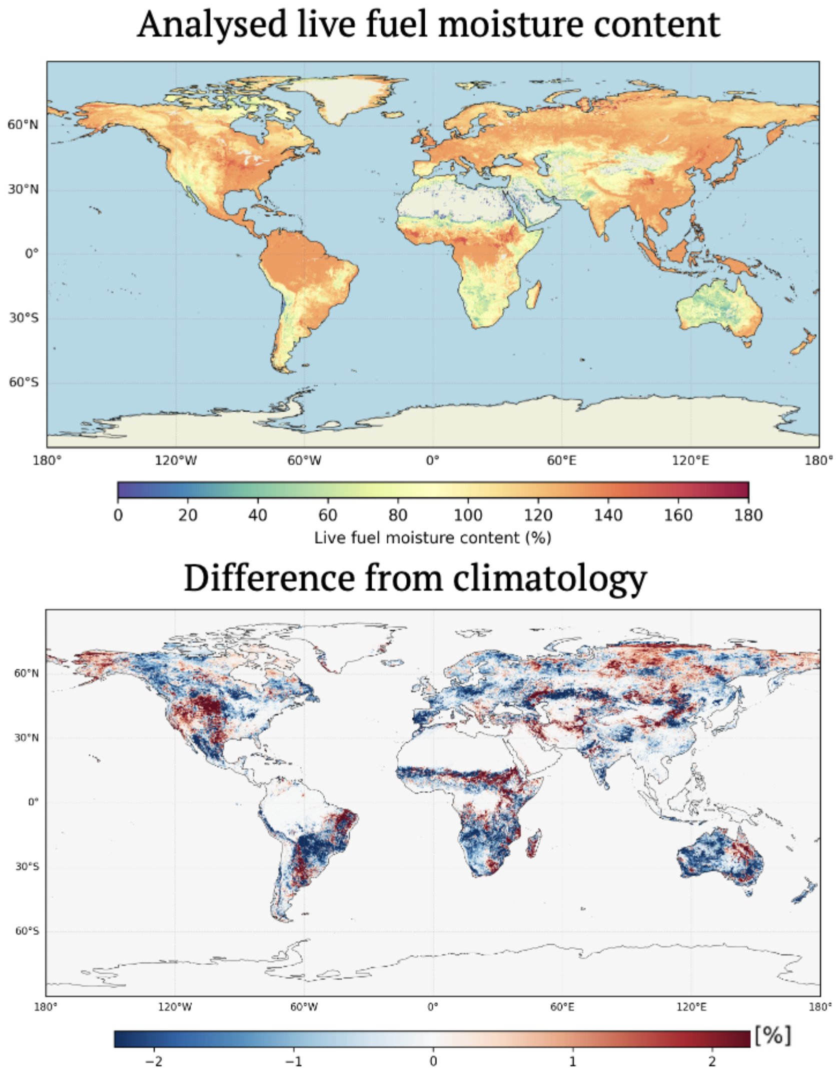

Figure 9: Top: Analysed live fuel moisture content (LFMC) [%] for August 2021, derived from a diagnostic model (Equation 1) driven by assimilated surface soil moisture and bias-corrected LAI. Bottom: Relative difference (%) between the analysed LFMC and a LAI-based climatology LFMC, highlighting areas with anomalously dry (blue) or moist (red) vegetation conditions.

The LFMC was estimated using equation Eq. 1, which integrates two key variables from the land data assimilation system: surface soil moisture from ASCAT assimilation and unbiased LAI derived through post-processing of SIF- and VOD-L–based assimilation experiments. The top panel of Figure 9 presents the global LFMC distribution for August 2021. Expected spatial patterns emerge, with high LFMC values observed in tropical rainforests and moist temperate regions, where vegetation remains well-hydrated year-round. In contrast, arid and semi-arid zones, such as southern Australia, southern Africa, and parts of the western United States, display much lower LFMC values, consistent with their seasonal dry periods and high fire potential at this time of year.

The bottom panel illustrates the impact of data assimilation on LFMC, expressed as the difference between the diagnostic LFMC estimate based on analysed LAI and climatological LAI. These increments represent how satellite observations (ASCAT, SIF, VOD-L, and CGLS LAI) adjust the vegetation moisture state in the model. Negative increments dominate in key fire-prone regions, including the Brazilian cerrado, southern Africa, eastern Australia, and parts of the western United States. The assimilation reduced LFMC values by up to 2 percentage points, indicating that the control simulation likely overestimated vegetation moisture during the peak fire season. Smaller positive increments appear where vegetation greening or moisture recovery was observed, such as in parts of eastern Europe and Central Asia.

Although most adjustments fall within ±2 percentage points, they are significant in fire-prone regions, where even subtle changes can shift the system across ignition thresholds. The sensitivity of fire ignition and spread to LFMC depends strongly on vegetation type. In fine, herbaceous fuels such as grasses and savannas, LFMC is closely coupled to short-term drying and responds rapidly to atmospheric and soil moisture conditions. In contrast, in woody fuels—such as Mediterranean shrublands, conifer forests, or eucalyptus stands—small drops in LFMC can sharply increase flammability by affecting leaf combustibility, volatility, and energy release. In these systems, a decrease of just 1–2 points can mark the transition from low to extreme fire risk, particularly when combined with high temperatures and wind. The spatial pattern of negative LFMC increments aligns, to some extent, with regions that experienced major wildfire activity in 2021.

3.3 Study case: Greece fires - August 2021

In August 2021, Greece experienced one of the most destructive wildfire seasons in recent history, with the island of Evia (Euboea) being particularly hard hit. Northern Evia, a region dominated by dense pine forest and shrubland, saw extensive fire activity between 3 and 11 August, resulting in a total burnt area of over 500 km^2, as captured by the MODIS Burnt Area product (Figure 10, a). The spatial concentration of fire activity in this region was exceptionally high, making it one of the largest single fire events in Europe during that summer.

The evolution of vegetation state over the affected area is illustrated in Figure (10, b). The assimilated LAI analysis (black curve), which integrates satellite-derived VOD-L and SIF observations, closely tracks the CGLS 300 m observational time series (red crosses) until the fire event. Prior to August, the vegetation followed a typical Mediterranean seasonal trajectory, peaking in late spring to early summer, in line with climatological LAI. However, during and immediately after the active fire window (highlighted in orange), the LAI dropped abruptly, reflecting the destruction of canopy structure due to combustion and vegetation mortality. This drop is captured both in the assimilated analysis and in the direct satellite observations, highlighting the system's sensitivity to abrupt disturbance events. Climatological values for low and high vegetation types (green and yellow dots) are also displayed. Post-fire LAI falls well below the typical values for low vegetation, confirming that the magnitude of vegetation loss was extreme. Figure (10, c) focuses on the temporal evolution of LAI during the critical two-week period surrounding the fire. Here, the abrupt LAI decline in early August is clearly visible, closely aligned with MODIS active fire detections. The rapid loss of LAI is an indicator of severe canopy disruption, consistent with ground-based assessments of the area, which reported widespread deforestation and long-term ecological damage.

Together, these results highlight the ability of data-assimilated LAI systems to capture both the slow evolution of seasonal vegetation cycles and the rapid onset of extreme disturbances such as wildfire. By integrating real-time satellite observations, such systems provide improved vegetation state estimates and a timely, quantitative window into the ecological impacts of climate-driven events. This capacity is

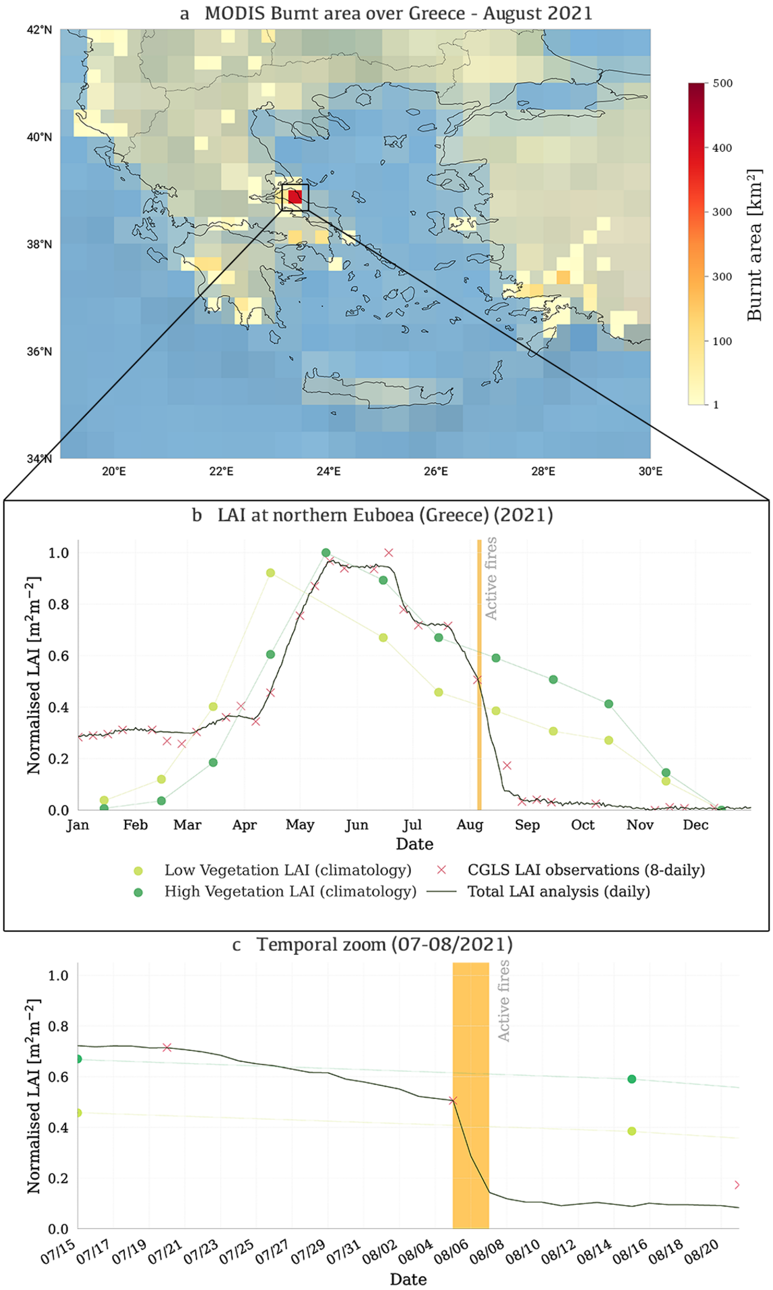

Figure 10: Wildfire impact on vegetation in northern Evia, Greece, during August 2021. (a) MODIS-derived burnt area over Greece for August 2021, highlighting a major wildfire event in northern Evia (Euboea), where over 500 km² of land was affected. (b) Seasonal evolution of the normalised LAI over the red square in panel (a), which corresponds to northern Euboea, comparing the daily assimilated LAI analysis (black line), CGLS LAI observations (red crosses), and low and high climatological LAI (light and dark green dots, respectively). The timing of active fire detections (orange shaded region) is derived from MODIS thermal anomalies. (c) A zoomed view of the LAI time series from July to August 2021, illustrating the sharp decline in LAI that coincides with the fire period.

critical for post-fire assessment, land surface reanalysis, and monitoring the pace of ecosystem recovery in the aftermath of large disturbances.

The alignment between assimilated LAI, satellite observations, and MODIS fire detections highlights the potential of integrated Earth observation systems to support both retrospective reanalysis and near-real-time monitoring of landscape disturbances.

4 Conclusions and next steps

Both LFMC and fuel load are key controllers of fire activity, particularly relevant for predicting and characterising extreme fires. LFMC influences vegetation flammability, with low moisture levels significantly increasing the probability of ignition, fire intensity, and spread. In parallel, high fuel loads—resulting from biomass accumulation over time—provide the combustible material needed to sustain large, severe fires. These two factors are increasingly shaped by hydroclimatic whiplash, where alternating wet and dry extremes promote rapid fuel build-up followed by intense drying. Wet periods lead to vigorous vegetation growth, while subsequent drought phases desiccate live fuels, lowering LFMC and creating highly flammable conditions. This dynamic is especially evident in regions like California and the Pantanal, where fire activity has become increasingly decoupled from traditional fire weather metrics. Incorporating LFMC, fuel load, and hydroclimatic variability into fire danger assessments is therefore essential to improve predictive accuracy and guide preparedness for extreme fire behaviour.

In the SPARKY ECMWF fire model, both LFMC and fuel load currently depend on a climatological LAI, which limits the model's ability to capture rapid changes in fuel conditions. This includes modifications driven by preceding fires, unusually wet or dry seasons, and land use changes such as crop harvesting at varying times or deforestation. The use of a static LAI does not reflect the dynamic nature of vegetation and fuel availability. However, the increasing availability of satellite-derived observations offers the potential to incorporate interannual variability and improve the model's responsiveness to real-time changes in land surface conditions.

Thus far, FUELITY has explored the use of VOD-L and SIF observations to inform LFMC by developing a comprehensive data assimilation system based on an extended ensemble Kalman filter. This system employs machine learning to model the relationship between satellite observations and vegetation variables. It shows potential for improving the representation of vegetation temporal dynamics in the IFS and has been successfully tested for LFMC estimation; it will be extended to include fuel load.

Initial results are promising:

1. The use of multiple sensors is advantageous, as they provide complementary information across different biomes.

2. The assimilation increments move in the expected direction when validated against independent observations.

3. The frequent updates allow the system to capture rapid shifts in fuel conditions, such as the sharp onset of fire activity observed during the Greece event.

The next stage involves incorporating fuel load and running the system over a longer time period to assess its impact on fire dynamics.

While results are promising, better performance could be achieved by using a prognostic LAI, as the current approach resets the benefits at each time step. A prognostic LAI would provide the system with memory, allowing it to retain and build on vegetation changes over time. This is a direction of development that ECMWF is embracing.

Annex: Sensitivity experiments

Table A.1: The standard deviation values listed here represent a selected subset of the sensitivity experiments conducted in this study. These values were varied to evaluate their impact on the performance of the data assimilation system. The corresponding analysis increments, focused over Europe, are presented in Figure A.1, and the final configuration retained for the data assimilation system is provided in Table 1.

ID	**LAI** ($\mathrm{m^2\,m^{-2}}$)	**SIF** ($\mathrm{mW\,m^{-2}\,sr^{-1}\,nm^{-1}}$)	**VOD-L** (unitless)
iqjq	1.2	0.10	0.10
iq5r	0.4	0.10	0.10
iqjx	0.8	0.10	0.32
iqjz	0.8	0.10	0.65
iq5z	0.8	0.20	0.10
iqjr	0.8	0.45	0.10

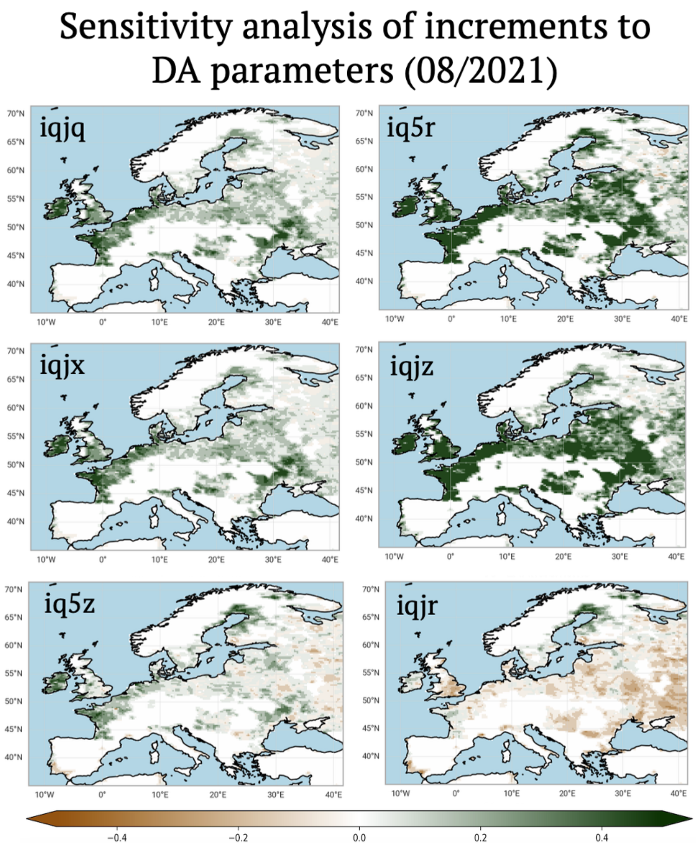

Figure A.1: Spatial patterns of LAI analysis increments for August 2021 from the sensitivity experiments listed in Table A.1. Each panel represents one experiment where the standard deviation of background LAI, satellite-based SIF, or VOD-L observations was varied to assess its influence. The increments illustrate how different uncertainty configurations impact the LAI analysis across Europe. White areas indicate regions where no simultaneous SIF and VOD-L observations were available during August 2021, and thus no data assimilation was performed there.

This annex provides details of the sensitivity experiments conducted to evaluate the influence of background and observation error standard deviations on assimilation performance. Table A.1 lists the tested configurations, and Figure A.1 illustrates the resulting LAI analysis increments across Europe. The final standard deviation values selected for the study (Table 1) were those that yielded the most robust and realistic system performance.

References

[1] A. Al Bitar, A. Mialon, Y. H. Kerr, F. Cabot, P. Richaume, E. Jacquette, A. Quesney, A. Mahmoodi, S. Tarot, M. Parrens, A. Al-Yaari, T. Pellarin, N. Rodriguez-Fernandez, and J.-P. Wigneron. The global SMOS Level 3 daily soil moisture and brightness temperature maps. *Earth System Science Data*, 9:293–315, 2017. doi:10.5194/essd-9-293-2017.

[2] C. Albergel, S. Munier, A. Bocher, B. Bonan, Y. Zheng, C. Draper, D. J. Leroux, and J.-C. Calvet. LDAS-Monde Sequential Assimilation of Satellite Derived Observations Applied to the Contiguous US: An ERA-5 Driven Reanalysis of the Land Surface Variables. *Remote Sensing*, 10(10), 2018. doi:10.3390/rs10101627.

[3] G. Balsamo, P. Viterbo, A. Beljaars, B. van den Hurk, M. Hirsch, A. Betts, and K. Scipal. A revised hydrology for the ECMWF model: verification from field site to terrestrial water storage and impact in the Integrated Forecast System. *Journal of Hydrometeorology*, 10:623–643, 2009.

[4] A.-L. Barbu, J.-C. Calvet, J. Demarty, V. Masson, and J.-F. Mahfouf. Assimilation of Soil Wetness Index and Leaf Area Index into the ISBA-A-gs land surface model: grassland case study. *Biogeosciences*, 8:1971–1986, 2011. doi:10.5194/bg-8-1971-2011.

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

[6] Jessica Blunden, Timothy S. Boyer, and Howard J. Diamond. State of the climate in 2021. *Bulletin of the American Meteorological Society*, 103(9):S1–S475, 2022. URL: https://journals.ametsoc.org/view/journals/bams/103/9/BAMS-D-22-0092.1.xml, doi:10.1175/BAMS-D-22-0092.1.

[7] S. Boussetta and G. Balsamo. Vegetation dataset of Land Use/Land Cover and Leaf Area Index. https://confess-h2020.eu/wp-content/uploads/2021/08/confess-d1-1-v1-0-.pdf, 2021. Deliverable D1.1, CONFESS Project, Horizon 2020.

[8] S. Boussetta, G. Balsamo, G. Arduini, E. Dutra, J. McNorton, M. Choulga, A. Agustí-Panareda, A. Beljaars, N. Wedi, J. Muñoz-Sabater, P. de Rosnay, I. Sandu, I. Hadade, G. Carver, C. Mazzetti, C. Prudhomme, D. Yamazaki, and E. Zsoter. ECLand: The ECMWF Land Surface Modelling System. *Atmosphere*, 12(6):723, 2021. doi:10.3390/atmos12060723.

[9] A. Brut, C. Rüdiger, S. Lafont, J.-L. Roujean, J.-C. Calvet, L. Jarlan, A.-L. Gibelin, C. Albergel, P. Le Moigne, J.-F. Soussana, K. Klumpp, D. Guyon, J.-P. Wigneron, and E. Ceschia. Modelling LAI at a regional scale with ISBA-A-gs: comparison with satellite-derived LAI over southwestern France. *Biogeosciences*, 6(8):1389–1404, 2009.

[10] P. de Rosnay, C. Bacour, B. Bonan, J.-C. Calvet, T. Corchia, S. Garrigues, T. Kaminski, W. Knorr, F. Maignan, P. Peylin, M. Scholze, V. Tartaglione, P. Vanderbecken, M. Voßbeck, and J. Vural. Report on SIF data assimilation method and preliminary testing in the IFS, 2024. URL: https://corso-project.eu/sites/default/files/2025-02/CORSO-D4-3-V1.5.pdf.

[11] P. de Rosnay, P. Browne, E. de Boisséson, D. Fairbairn, Y. Hirahara, K. Ochi, D. Schepers, P. Weston, H. Zuo, M. Alonso-Balmaseda, G. Balsamo, M. Bonavita, N. Bormann, A. Brown, M. Chrust, M. Dahoui, G. De Chiara, S. English, A. Geer, S. Healy, H. Hersbach, P. Laloyaux, L. Magnusson,

S. Massart, A. McNally, F. Pappenberger, and F. Rabier. Coupled data assimilation at ECMWF: current status, challenges and future developments. *Quarterly Journal of the Royal Meteorological Society*, 148(747):2672–2702, 2022. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.4330, doi:10.1002/qj.4330.

[12] P. de Rosnay, M. Drusch, D. Vasiljevic, G. Balsamo, C. Albergel, and L. Isaksen. A simplified Extended Kalman Filter for the global operational soil moisture analysis at ECMWF. *Quarterly Journal of the Royal Meteorological Society*, 139(674):1199–1213, 2013. doi:10.1002/qj.2023.

[13] G. Desroziers, L. Berre, B. Chapnik, and P. Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3385–3396, 2005. doi:10.1256/qj.05.108.

[14] W. A. Dorigo, A. Gruber, R. A. M. De Jeu, W. Wagner, T. Stacke, A. Loew, C. Albergel, L. Brocca, D. Chung, R. Parinussa, R. Reichle, C. Rüdiger, R. Van Der Schalie, M. Vreugdenhil, and S. Zwieback. The international soil moisture network: serving earth system science for over a decade. *Hydrology and Earth System Sciences*, 25:5749–5804, 2021. doi:10.5194/hess-25-5749-2021.

[15] C. S. Draper, R. H. Reichle, G. J. M. De Lannoy, S. P. Mahanama Q. Liu, and R. D. Koster. Assimilation of ASCAT near-surface soil moisture into the SIM hydrological model over France. *Hydrology and Earth System Sciences*, 15(12):3829–3841, 2011. doi:10.5194/hess-15-3829-2011.

[16] M. Drusch, K. Scipal, P. de Rosnay, G. Balsamo, E. Andersson, P. Bougeault, and P. Viterbo. Towards a Kalman Filter based soil moisture analysis system for the operational ECMWF Integrated Forecast System. *Geophysical Research Letters*, 36(10), 2009. doi:10.1029/2009GL037716.

[17] S. El Garroussi and J. McNorton. Towards enhanced fire fuel estimation with satellite-derived predictive models. Technical report, ECMWF, ESA Contract Report, 01/2025. doi:10.21957/6d8a1ebcc1.

[18] M. Forkel, W. A. Dorigo, Y. Liu, Y. Zhang, J. Schwaab, and G. Camps-Valls. Estimating leaf moisture content at global scale from passive microwave observations of vegetation optical depth. *Hydrology and Earth System Sciences*, 27(1):39–68, 2023.

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

[20] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J.-R. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. J. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, and J.-N. Thépaut. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020. doi:10.1002/qj.3803.

[21] K. Ide, M. Ghil, and A. Lorenc. Unified Notation for Data Assimilation: Operational, Sequential and Variational. *Journal of the Meteorological Society of Japan*, 75(N 1B):181–189, 05 1997.

[22] J.-C. Calvet, B. Bonan, O. Rojas-Munoz, A. Agusti-Panareda, P. de Rosnay, P. Weston, P. Peylin, C. Bacour, V. Bastrikov, F. Maignan, T. Kaminski, W. Knorr, M. Vossbeck, and M. Scholze. D3.4 Demonstrator systems for using remote sensing data (LAI, VOD, SIF) in online global prior fluxes for the CO2MVS prototype. Technical report, CoCO2: Prototype system for a Copernicus CO_2 service, 2023. URL: https://coco2-project.eu/sites/default/files/2023-11/CoCO2-D3-4-V2-1.pdf.

[23] W. M. Jolly, M. A. Cochrane, P. H. Freeborn, Z. A. Holden, T. J. Brown, G. J. Williamson, and D. M. J. S. Bowman. Climate-induced variations in global wildfire danger from 1979 to 2013. *Nature Communications*, 6:7537, 2015.

[24] K. Scipal, M. Drusch, and W. Wagner. Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system. *Advances in water resources*, 2008. doi:10.1016/j.advwatres.2008.04.013.

[25] J. McNorton and F. Di Giuseppe. A global fuel characteristic model and dataset for wildfire prediction. *Biogeosciences*, 21(2):279–300, 2024.

[26] J. Muñoz-Sabater, E. Dutra, A. Agustí-Panareda, C. Albergel, G. Arduini, G. Balsamo, S. Boussetta, M. Choulga, S. Harrigan, H. Hersbach, B. Martens, D. G. Miralles, M. Piles, N. J. Rodríguez-Fernández, E. Zsoter, C. Buontempo, and J.-N. Thépaut. ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13:4349–4383, 2021. doi:10.5194/essd-13-4349-2021.

[27] X. Quan, M. Yebra, N. Tapper, V. Resco de Dios, J. P. Guerschman, Z. Zhu, and Y. He. Global fuel moisture content mapping from MODIS. *International Journal of Applied Earth Observation and Geoinformation*, 105:102597, 2021.

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

[29] J. M. Sabater, C. Rüdiger, J.-C. Calvet, N. Fritz, L. Jarlan, and Y. Kerr. Joint assimilation of surface soil moisture and LAI observations into a land surface model. *Agricultural and Forest Meteorology*, 148:1362–1373, 2008. doi:10.1016/j.agrformet.2008.04.003.

[30] A. Verger, M. Weiss, F. Baret, F. Camacho, and H. Makhmara. *Leaf Area Index 2014-present (raster 300 m), global, 10-daily – version 1*. Copernicus Global Land Service, 2019. URL: https://land.copernicus.eu/en/products/vegetation/leaf-area-index-300m-v1.0, doi:10.2909/219fdc9f-616b-444b-a495-198f527b4722.

[31] M. Yebra, P. E. Dennison, E. Chuvieco, D. Riaño, P. Zylstra, Jr. Hunt, E. R., F. M. Danson, Y. Qi, and S. Jurdao. A global review of remote sensing of live fuel moisture content for fire danger assessment: Moving towards operational products. *Remote Sensing of Environment*, 136:455–468, 2013.

[32] M. Yebra, G. Scortechini, A. Badi, M. Beget, M. Boer, R. Bradstock, E. Chuvieco, F. Danson, P. Dennison, V. Resco de Dios, C. Di Bella, G. Forsyth, P. Frost, M. Garcia, A. Hamdi, B. He,

M. Jolly, T. Kraaij, M. Martín, F. Mouillot, G. Newnham, R. Nolan, G. Pellizzaro, Y. Qi, X. Quan, D. Riaño, D. Roberts, M. Sow, and S. Ustin. Globe-LFMC, a global plant water status database for vegetation ecophysiology and wildfire applications. *Scientific Data*, 6:155, 2019. doi:10.1038/s41597-019-0164-9.

[33] Z. Zhu, M. Yebra, X. Quan, Q. Sun, X. Yang, J. P. Guerschman, and Y. He. Live fuel moisture content estimation from MODIS: A deep learning approach. *Remote Sensing of Environment*, 252:112150, 2021.