

LDAS-Monde Sequential Assimilation of Satellite Derived Observations Applied to the CONtiguous US: An ERA-5 Driven Reanalysis of the Land Surface Variables

Clement Albergel¹, Simon Munier¹, Aymeric Bocher¹, Bertrand Bonan¹, Yongjun Zheng¹, Clara Draper², Delphine J. Leroux¹, Jean-Christophe Calvet¹

¹CNRM UMR 3589, Météo-France/CNRS, Toulouse, France

²CIRES / NOAA Earth System Research Laboratory, Boulder, CO, USA

Corresponding author : Clement Albergel, clement.albergel@meteo.fr

Abstract - LDAS-Monde, an offline land data assimilation system with global capacity, is applied over the CONtiguous US (CONUS) domain to enhance monitoring accuracy for water and energy states and fluxes. LDAS-Monde ingests satellite-derived Surface Soil Moisture (SSM) and Leaf Area Index (LAI) estimates to constrain the Interactions between Soil, Biosphere, and Atmosphere (ISBA) Land Surface Model (LSM) coupled with the CNRM (Centre National de Recherches Météorologiques) version of the Total Runoff Integrating Pathways (CTRIP) continental hydrological system (ISBA-CTRIP). LDAS-Monde is forced by the ERA-5 atmospheric reanalysis from the European Center For Medium Range Weather Forecast (ECMWF) from 2010 to 2016 leading to a 7-yr, quarter degree spatial resolution offline reanalysis of Land Surface Variables (LSVs) over CONUS.

The impact of assimilating LAI and SSM into LDAS-Monde is assessed over North America, by comparison to satellite-driven model estimates of land evapotranspiration from the Global Land Evaporation Amsterdam Model (GLEAM) project, and upscaled ground-based observations of gross primary productivity from the FLUXCOM project. Also, taking advantage of the relatively dense data networks over CONUS, we also evaluate the impact of the assimilation against in-situ measurements of soil moisture from the USCRN network (US Climate Reference Network) are used in the evaluation, together with river discharges from the United States Geophysical Survey (USGS) and the Global Runoff Data Centre (GRDC). Those data sets highlight the added value of assimilating satellite derived observations compared to an open-loop simulation (i.e. no assimilation). It is shown that LDAS-Monde has the ability not only to monitor land surface variables but also to forecast them, by providing improved initial conditions which impacts persist through time. LDAS-Monde reanalysis has a potential to be used to monitor extreme events like agricultural drought, also. Finally, limitations related to LDAS-Monde and current satellite-derived

observations are exposed as well as several insights on how to use alternative datasets to analyze soil moisture and vegetation state.

Keywords: Land Surface Data Assimilation, remote sensing, ERA5

1 Introduction

One of the major scientific challenges in relation to the adaptation to climate change is observing and simulating the response of land biophysical variables to extreme events, making Land Surface Models (LSMs) constrained by high-quality gridded atmospheric variables and coupled with river-routing models key tools (Schellekens et al., 2017; Dirmeyer et al., 2006). The modelling of terrestrial variables can be improved through the dynamical integration of observations. Remote sensing observations are particularly useful in this context due to their global coverage, higher and higher spatial resolution. The current fleet of Earth observation missions holds an unprecedented potential to quantify land surface variables (LSVs) (Lettenmaier et al., 2015) and many satellite-derived products relevant to the hydrological and vegetation cycles are already available at high spatial resolution. Satellite remote sensing observations exhibit however spatial and temporal gaps and not all key LSVs can be observed. LSMs are able to provide LSV estimates at all times and locations using physically based equations but as remotely sensed observations they are affected by uncertainties (e.g., parametrization representation, atmospheric forcing, initialisation). Through a weighted combination of both, LSVs can be better estimated than by either source of information alone (Reichle et al., 2007): data assimilation techniques enable to spatially and temporally integrate observed informations into LSMs in a consistent way to unobserved locations, time steps and variables.

In the past recent years, several Land Data Assimilation System (LDAS) have emerged at different spatial scales: regional like like the Coupled Land Vegetation LDAS (CLVLDAS, Sawada and Koike, 2014, Sawada et al., 2015), the Famine Early Warning Systems Network (FEWSNET) LDAS (FLDAS, McNally et al., 2017), continental like the North American LDAS (NLDAS, Mitchell et al., 2004; Xia et al., 2012), the National Climate Assessment LDAS (NCA-LDAS Kumar et al., 2018) as well as at global scale like the Global Land Data assimilation (GLDAS, Rodell et al., 2004) and more recently LDAS-Monde (Albergel et al., 2017). LDAS-Monde has been developed to constrain the CO₂-responsive version of the ISBA (Interactions between Soil, Biosphere, and Atmosphere) LSM (Noilhan and Mahfouf, 1996; Calvet et al., 1998, 2004; Gibelin et al., 2006) using satellite derived observations within the open-source SURFEX modelling platform (SURFace Externalisee, Masson et al., 2013) of Meteo-France. LDAS-Monde has been implemented in a monitoring chain of terrestrial water and carbon fluxes. Unlike most of the above mentioned LDAS, LDAS-Monde is able to jointly and sequentially assimilate vegetation products such as Leaf Area

Index (LAI) together with Surface Soil Moisture (SSM) observations (Barbu et al., 2011; 2014; Fairbairn et al., 2017; Albergel et al., 2017). Albergel et al. (2017) tested LDAS-Monde over Europe and the Mediterranean basin for the 2000-2012 period. A long term, global scale, multi-sensor satellite-derived surface soil moisture dataset (ESA CCI SSM, Liu et al., 2011a, 2012; Dorigo et al., 2015, 2017) along with satellite derived LAI (GEOV1, <http://land.copernicus.eu/global/> last access, June 2018), were jointly assimilated. LDAS-Monde was forced by WFDEI (WATCH-Forcing-Data-ERA-Interim) observations based atmospheric forcing dataset (Weedon et al., 2011, 2014) at half degree spatial resolution. Analysis impact was successfully carried out using (i) agricultural statistics over France, (ii) river discharge observations, (iii) satellite-derived estimates of land evapotranspiration from the Global Land Evaporation Amsterdam Model (GLEAM) project and (iv) spatially gridded observations based estimates of up-scaled gross primary production and evapotranspiration from the FLUXNET network (Albergel et al., 2017).

In this study, LDAS-Monde is applied and tested in a data-rich area: the CONtiguous US (CONUS, defined here as longitudes from 130.0°W to 60.0°W, latitudes from 20.0° to 55.0° N, as shown in Figure 1). LDAS-Monde is forced by the latest ERA-5 atmospheric reanalysis from the European Center For Medium Range Weather Forecast (ECMWF) from 2010 to 2016 leading to a 7-yr, quarter degree spatial resolution offline reanalysis of the Land Surface Variables (LSVs). Albergel et al. (2018) assessed ERA-5 ability to force the ISBA LSM by comparison to satellite-derived products and in-situ observations covering a substantial part of the land surface storage and fluxes. They found that using ERA-5 in place its predecessor, ERA-Interim, led to significant improvements in the representation of the LSVs linked to the terrestrial water cycle (surface soil moisture, river discharges, snow depth and turbulent atmospheric fluxes), but did not improve the LSVs linked to the vegetation cycle (evapotranspiration, carbon uptake and LAI). In that respect, the assimilation of LAI through ERA-5 driven reanalysis from LDAS-Monde is expected to bring a clear improvements (Albergel et al., 2017). In this study, the impact of LDAS-Monde analysis with respect to an open-loop (i.e. model run without assimilation) is assessed using satellite-driven model estimates of land evapotranspiration from the Global Land Evaporation Amsterdam Model (GLEAM) project and upscaled ground-based observations of gross primary productivity from the FLUXCOM project, together with river discharges from the United States Geophysical Survey (USGS) and the Global Runoff Data Centre (GRDC). Over CONUS, in-situ measurements of soil moisture from the USCRN network (US Climate Reference Network) are used in the evaluation, also. Section 2 describes the different components of LDAS-Monde as well as the evaluation data-sets and strategy. Section 3 provides a set of statistical diagnostics to assess and evaluate the impact of the assimilation. Finally, Sect. 4 provides perspectives and future research directions.

2 Data and Methods

2.1 LDAS-Monde system components

LDAS-Monde allows sequential assimilation of satellite derived land observations at a global scale. The assimilation is performed into the open-access SURFEX modelling platform of Météo-France (SURFace Externalisée, Masson et al., 2013). It produces offline reanalyses of LSVs using (i) an LSM along with data assimilation techniques, (ii) observations and (iii) atmospheric forcing. Those components of LDAS-Monde are briefly described below.

2.1.1 The SURFEX modelling platform

LDAS-Monde uses the CO₂-responsive version of ISBA embedded in the SURFEX platform. The most recent version of SURFEX (version 8.1) is used in this study with the “NIT” plant biomass monitoring option for ISBA. In this configuration, ISBA simulates leaf-scale physiological processes and plant growth (Calvet et al., 1998, 2004; Gibelin et al., 2006), while transfers of water and heat through the soil rely on a multilayer diffusion scheme (Boone et al., 2000 and Decharme et al., 2013). The ISBA parameters are defined for 12 generic land surface patches. They include nine plant functional types (needle leaf trees, evergreen broadleaf trees, deciduous broadleaf trees, C3 crops, C4 crops, C4 irrigated crops, herbaceous, tropical herbaceous, and wetlands) as well as bare soil, rocks, and permanent snow and ice surfaces.

This version of ISBA is coupled to the CTRIP river routing model through OASIS-MCT (Volodire et al., 2017) in order to simulate streamflows of the main rivers (Decharme et al., 2010, 2012, Vergnes and Decharme, 2012, Vergnes et al., 2014). Besides, a single-source energy budget of a soil/vegetation composite is computed. SURFEX also involves data assimilation techniques to analyse LSVs from the ISBA LSM.

This study makes use of the Simplified version of an Extended Kalman Filter (SEKF) as already used and described in Barbu et al. (2011), Fairbairn et al. (2017), Albergel et al. (2017) and Leroux et al. (2018). The SEKF uses finite differences from perturbed simulations to estimate the linear tangent model linking the model state control variables to the observed variables. Satellite derived Surface Soil Moisture (SSM) and Leaf Area Index (LAI) are simultaneously assimilated to update eight model state control variables (i.e. control variables): LAI and soil moisture from seven layers of soil, from 1 cm to 100 cm. Assimilating SSM and LAI within LDAS-Monde results in updates of the LSM variables in different ways. First, the SEKF computes the Kalman gain to update the model variables corresponding to the observations. It can be variables of the model that are directly observed or related to the observations through an observation operator (that can be a simple transfer function). Secondly, control variables are updated through their functional link to the observed variables. For example, the assimilation of LAI impacts LAI itself but also soil moisture

from the 7 layers present in the state vector and the assimilation of SSM impacts LAI, also. Finally, other variables are indirectly modified by the SEKF process through biophysical processes and feedbacks in the model by updates of the control variables.

2.1.2 ESA CCI Surface Soil Moisture and CGLS Leaf Area Index

In this study the European Space Agency and Climate Change Initiative (ESA CCI) SSM-combined version of the product (v4.1) is assimilated into LDAS-Monde (<http://www.esa-soilmoisture-cci.org>, last access June 2018). The CCI merges SSM observations from seven different microwave radiometers (SMMR, SSM/I, TMI, ASMR-E, WindSat, AMSR2, SMOS) and four different scatterometers (ERS-1 and 2 AMI and MetOp-A and B ASCAT) into a single combined data set covering the time period from November 1978 to December 2016. Data are expressed in volumetric (m^3m^{-3}) units and quality flags are provided (i.e. snow coverage or temperature below 0° and dense vegetation). For a more comprehensive overview of the product see Dorigo et al. (2015, 2017). Topographic relief is known to negatively affect satellite remote sensing retrievals of SSM (Mätzler and Standley, 2000), hence the time series for pixels whose average altitude exceeds 1500 m above sea level were not accounted for. Data on pixels with urban land cover fractions larger than 15% were discarded too, to limit the effects of artificial surfaces. These thresholds were set according to Draper et al. (2011), Barbu et al. (2014) and Albergel et al. (2017) who processed satellite-based SSM retrievals for data assimilation experiments with the ISBA LSM. Data are available almost every day with a spatial resolution of $0.25^\circ \times 0.25^\circ$. Similarly to previous studies, the ESA CCI SSM product has been transformed into the model-equivalent SSM to address possible misspecification of physiographic parameters, such as the wilting point and the field capacity. The linear rescaling approach described in Scipal et al. (2008) (using the first two moments of the cumulative distribution function, CDF) was used. It consists of a linear rescaling enabling a correction of the differences in the mean and variance of the distribution. It has been applied at a seasonal scale (i.e. for each specific month) following Albergel et al. (2017).

The GEOV1 LAI is assimilated, also. It is produced by the European Copernicus Global Land Service project (<http://land.copernicus.eu/global/>) last access June 2018). Boussetta et al. (2015) proposed an evaluation of this product in the context of Numerical Weather Prediction (NWP). LAI observations are retrieved from the SPOT-VGT (from 1999 to 2014) and then from PROBA-V (from 2014 to present) satellite data according to the methodology proposed by Baret et al. (2013). The 1 km spatial resolution observations are interpolated by an arithmetic average to the $0.25^\circ \times 0.25^\circ$ model grid points, if at least 50 % of the observation grid points are observed (i.e half the maximum amount). LAI observations have a temporal frequency of 10 days at best (e.g., in

presence of clouds no observations are available). Both assimilated datasets are illustrated by Figure 1, averaged over 2010-2016.

2.1.3 ERA-5 atmospheric reanalysis

ERA-5 (Hersbach and Dee, 2016) is the fifth generation of European reanalyses produced by the ECMWF and a key element of the EU-funded Copernicus Climate Change Service (C3S). ERA-5 important changes relative to ERA-Interim former ECMWF's atmospheric reanalysis include (i) a higher spatial and temporal resolution as well as (ii) a more recent version of ECMWF Earth system model physics and data assimilation system (corresponding to ECMWF's cycle CY41R2, <https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model/ifs-documentation>, last access June 2018). It makes it able to use modern parameterizations of Earth processes compared to older versions used in ERA-Interim. For instance, in addition to being applied to satellite observations, a variational bias scheme is also applied to aircraft and surface ozone and pressure data. ERA-5 also benefits from reprocessed data sets that were not ready yet during the production of ERA-Interim. Two other important features of ERA-5 are the more frequent model output and improved model spatial resolution, going from 6-hourly output in ERA-Interim to hourly output analysis in ERA-5, and from 79 km (horizontal dimension) and 60 levels (vertical dimension), to 31 km and 137 levels in ERA-5. Finally, ERA-5 also provides an estimate of uncertainty through the use of a 10-member Ensemble of Data Assimilations (EDA) at a coarser resolution (63 km horizontal resolution) and 3-hourly frequency. ERA-5 is foreseen to replace ERA-Interim reanalysis. All ERA-5 atmospheric variables were interpolated at $0.25 \times 0.25^\circ$ spatial resolution. A bilinear interpolation from the native reanalysis grid to the regular grid was made.

2.2 Evaluation datasets and methods

LDAS-Monde analysis impact was assessed with respect to the open-loop model run (i.e. no assimilation). The system was spun-up by running year 2010 twenty times. Table I presents the set up of the different experiments used in this study, the openloop and the analysis as well as two additional model runs: (i) *Ini_model*, a 12-month model run starting on 1 January 2016 (initialised by the model simulation run from 2010 to 2015) and (ii) *Ini_analysis*, a 12-month model run initialised by initial conditions from the analysis on 1 January 2016. The two above-mentioned assimilated datasets (ESA-CCI SSM and LAI GEOF1) were used as a way to check to what extent the assimilation system was able to produce analyses closer to these two datasets that were assimilated than the open-loop. Then two independent spatially distributed datasets, namely evapotranspiration from the GLEAM project (Miralles et al., 2011, Martens et al., 2017) and Gross Primary Production (GPP) from the FLUXCOM project (Tramontana et al., 2016, Jung et al., 2017)

were used in the evaluation process. Ground based measurements of soil moisture from the USCRN network (US Climate Reference Network, Bell et al., 2013) were used, also, along with river discharge observations from the United States Geophysical Survey (USGS) and the Global Runoff Data Centre (GRDC).

LDAS-Monde ability to represent SSM, LAI, evapotranspiration and GPP, was assessed using the correlation coefficient (R) and Root Mean Square Difference (RMSD). These metrics were applied at a seasonal scale (i.e. for each month) over 2010-2016. For ground based measurements of SSM, R was calculated for both absolute and anomaly time series in order to remove the strong impact from the SSM seasonal cycle on this specific metric (see e.g., Albergel et al., 2017, 2018). Ground measurements at a depth of 5 cm were compared to soil moisture of the third layer of soil (between 4 and 10 cm depth) from both the model and the analysis for months from April to September over the 2010-2016 time period to avoid frozen conditions. Only stations with significant R values for the two experiments (with p-value < 0.05) were kept for the evaluation.

In order to provide an easier measurement of the added value of the analysis, statistics were also normalized with respect to the model. The so called Normalized Information Contribution index (NIC as in Kumar et al., 2009; Albergel et al., 2018) was applied to the correlation coefficient (Eq. 1, for both volumetric and anomaly time-series) and to RMSD (Eq. 2) to quantify the improvement or degradation from the analysis with respect to the model.

$$NIC_R = \frac{R_{(Analysis)} - R_{(Model)}}{1 - R_{(Model)}} \times 100 \quad \text{Eq.1}$$

$$NIC_{RMSD} = \frac{RMSD_{(Analysis)} - RMSD_{(Model)}}{RMSD_{(Model)}} \times 100 \quad \text{Eq.2}$$

NIC scores were then classified according to three categories: (i) negative impact from the analysis with respect to the model with values smaller than -3 %, (ii) positive impact from the analysis with respect to the model with values greater than +3 % and (iii) neutral impact from the analysis with respect to the model with values between -3 % and 3 %.

Over the 2010-2016 time period, river discharge from the analysis and model runs were compared to daily streamflow data from USGS and GRDC. Data were selected for sub-basins with rather large drainage areas ($10,000 \text{ km}^2$ or greater) due to the low resolution of CTRIP ($0.5 \times 0.5^\circ$) and with a long observation time series (48 months or more). As commonly found in the literature observed and simulated river discharge (Q) data are expressed in m^3s^{-1} . However given that the observed drainage areas may differ from the simulated ones, specific discharge in mm.d^{-1} (the ratio of Q to the drainage area) was used in this study, similarly to Albergel et al., 2017, 2018. Stations with drainage areas differing by more than 20 % from the simulated ones were discarded. Impact on Q was evaluated using the Kling–Gupta Efficiency (KGE; Gupta et al., 2009) score:

$$KGE = 1 - \sqrt{RE_{\mu}^2 + RE_{\sigma}^2 + (1 - R)^2} \quad \text{Eq. 3}$$

with RE_{μ} and RE_{σ} the relative error of simulated or analysed mean and standard deviation (Eq.4) and (Eq. 5) respectively, R the correlation coefficient between the observed discharges and either the modelled or analysed river discharges.

$$RE_{\mu} = \frac{Q_{\mu}}{Q_{(obs.)\mu}} - 1 \quad \text{Eq.4}$$

$$RE_{\sigma} = \frac{Q_{\sigma}}{Q_{(obs.)\sigma}} - 1 \quad \text{Eq.5}$$

KGE represents the Euclidean distance from the ideal point in the $[RE_{\mu}, RE_{\sigma}, R]$ score space. RE_{μ} , RE_{σ} and R constitute a set of mathematically independent metrics quantifying the fit of simulated/analysed discharge time series. At best, RE_{μ} and RE_{σ} are equal to 0 and R is equal to 1 (leading to a perfect KGE value of 1), indicating that simulated or analysed time series are identical to the measured one. NIC (Eq. 1) was applied to KGE (Eq.6) as well, for stations with KGE values greater than 0, only. Finally RE_{μ} and RE_{σ} metrics were normalised, following Eq. 7, Eq. 8 to appreciate the added value from the analysis with respect to the model, also.

$$NIC_{KGE} = \frac{KGE_{(Analysis)} - KGE_{(Model)}}{1 - KGE_{(Model)}} \times 100 \quad \text{Eq.6}$$

$$N_{RE_{\mu}} = 100 * \frac{RE_{\mu(Analysis)} - RE_{\mu(Model)}}{RE_{\mu(Model)}} \quad \text{Eq.7}$$

$$N_{RE_{\sigma}} = 100 * \frac{RE_{\sigma(Analysis)} - RE_{\sigma(Model)}}{RE_{\sigma(Model)}} \quad \text{Eq.8}$$

3 Results

3.1 Analysis impact on assimilated variables

Being the model equivalents of the assimilated observations, LAI and soil moisture from the second layer of soil are expected to be the two variables most affected by the assimilation. Figure 2 presents 10-day time series of LAI averaged over the whole domain for the 2010-2016 time period. From Figure 2 one can see that the open-loop simulation tends to overestimate the observed LAI in winter periods and that the senescence phase of vegetation is too late over the autumn when compared to the observations. To that respect, the assimilation is efficiently correcting the model ; however analysed LAI does not reach LAI maximal values of the observations. Figure 3 shows maps of LAI for the model (Fig. 3a), the observations (Fig. 3b) and the analysis (Fig. 3c) averaged over 2010-2016. It is clearly visible that the model overestimates LAI in the eastern part of the domain. Also, some geographical patterns visible in the observations (e.g., the Mississippi and red river areas in Fig. 3b) are not represented in the model (Fig. 3a). After assimilation, the analysis presents reduced LAI values in the eastern part of the domain and the above-mentioned geographical patterns are visible, too (fig.3c). This shows the ability of the assimilation to integrate

geographical information into the model. Figure 3 also presents seasonal scores between the model and the observations and between the analysis and the observations for RMSD and R, over the 2010-2016 time period. The analysis leads to better fit between the model forecasts and the subsequent assimilated observations, for both metrics. On average for the whole period, RMSD values drop from $1.10 \text{ m}^2\text{m}^{-2}$ (model vs. observations) to $0.65 \text{ m}^2\text{m}^{-2}$ (analysis vs. observations) while R values increase from 0.69 (model vs. Observations) to 0.88 (Analysis vs. Observations). Figure 4 presents the same information content for soil moisture. As ESA CCI SSM was rescaled in order to match the modelled SSM probability distribution, differences are hardly visible from fig.4a, b and c. From fig.4d & e however, one can appreciate the added value of the analysis: RMSD values drop from $0.046 \text{ m}^3\text{m}^{-3}$ (model vs. Observations) to $0.044 \text{ m}^3\text{m}^{-3}$ (Analysis vs. observations) while R values increase from 0.85 (model vs. observations) to 0.87 (analysis vs. observations). Finally, Figure 5 shows maps of analysis increments for 4 (out of 8) control variables averaged over the whole 2010-2016 time period (LAI, second, fourth and sixth layers of soil from left to right, respectively). It can be noticed that the magnitude of increments is decreasing with depth. It can also be noticed that over almost the whole domain, the analysis tends to add water in the soil near the surface (positive increments) while it removes water in the deepest layers (negative increments).

3.2 Evaluation using independent datasets

3.2.1 *Evapotranspiration and GPP*

On average, R increases from 0.80 to 0.81 when comparing evapotranspiration from the model and from the analysis, respectively, to the independent estimates. Average RMSD decreases from $0.89 \text{ kg.m}^{-2}.\text{d}^{-1}$ to $0.85 \text{ kg.m}^{-2}.\text{d}^{-1}$. When compared to GPP estimates, averaged correlations rise from 0.74 to 0.78 and RMSD drops from $2.198 \text{ kg(C).m}^{-2}.\text{d}^{-1}$ to $1.908 \text{ kg(C).m}^{-2}.\text{d}^{-1}$ when considering the model or the analysis, respectively. Figure 6 presents spatial maps of N_{RMSD} (fig.6a and c) and NIC_R (fig.6b and d) resulting from the comparison with evapotranspiration (figure 6, top row) and GPP (figure 6, bottom row) of their modeled and analysed equivalent. For, N_{RMSD} (fig.6a and c) blue colors represent an improvement from the analysis regarding RMSDs (i.e. the latter better represents either evapotranspiration or GPP than the model) while for NIC_R (fig.6b and d) red colors represent an improvement from the analysis. Figure 6 shows that both evapotranspiration and GPP are improved almost everywhere, in terms of correlation and RMSD. Finally Figure 7 shows the annual cycle of monthly RMSD and R values. From fig.7a and c one can appreciate the positive impact of the assimilation, with reduced RMSD values all year long, especially for GPP. From fig.7b, it is visible that the analysis has a positive impact on the R values for evapotranspiration on a limited time period (April to May) and from fig.7d, a strong positive impact on R values for GPP is noticeable all year long.

3.2.2 *Soil moisture*

The statistical scores for surface soil moisture from the model and the analysis (third layer of soil between 4 and 10 cm depth) over 2010–2016 when compared to ground measurements from the USCRN network (at 5 cm depth) are presented in Table II. Median R values on volumetric soil moisture time-series (anomaly time series) along with their 95 % confidence interval of the median derived from a 10 000 samples bootstrapping are: 0.72 ± 0.02 (0.60 ± 0.02) and 0.74 ± 0.02 (0.60 ± 0.02) while median ubRMSD are 0.049 ± 0.004 and 0.048 ± 0.004 for the model and the analysis, respectively. Figures 8a & b illustrate correlation values on volumetric and anomaly time-series, respectively, between the model and the observations, for each stations. Figures 8c & d represent the added value of the analysis expressed through the NIC index (Eq.1) applied for correlations (NIC_R) values on volumetric and anomaly time-series: large blue circles represent a positive impact from the analysis at NIC_R greater than +3 (i.e. R values are better when the analysis is used than when the model is used), large red circles a degradation from the analysis at NIC_R smaller than -3 while diamond symbols represent a rather neutral impact at NIC_R between [-3;+3]. While 46% (81%) of the pool of stations present a rather neutral impact for R values on volumetric (anomaly) time series, stations more impacted by the analysis tend to be positively impacted at 46% (18%), to be compared with 8% (1%) of negative impacts. Although differences between the model run and the analysis are rather small, these results underline the added value of the analysis with respect to the model run.

3.2.3 Streamflow

A subset of 258 out of 531 gauging stations was selected for the evaluation according to the criteria described in the methodology section, with KGE scores within the [0, 1] interval. Figure 9 presents the performance of analysed streamflow with respect to the one from the model run for this pool of stations, with a focus on the eastern part of the domain. NIC_{KGE} values are presented following the same classification as NIC_R applied to soil moisture. Scores are presented in Table III. Looking at NIC_{KGE} , 62% of the pool of stations (258 stations) present a rather neutral impact (at NIC_{KGE} between [-3;3]) and 26% of the stations present a positive impact (at $\text{NIC}_{\text{KGE}} > +3$) while only 12 % of stations have a negative impact (at $\text{NIC}_{\text{KGE}} < -3$). NIC_R , $N_{\text{RE}\sigma}$ and $N_{\text{RE}\mu}$ follow the same classification (with even a smaller percentage of stations being negatively affected by the analysis; 1%); when the analysis is impacting streamflow representation, it tends to be a positive impact.

4 Potential applications, discussions and perspectives

4.1 Could LDAS-Monde be used to monitor agricultural droughts?

The previous section has highlighted the LDAS-Monde ability to enhance the monitoring accuracy for land surface variables. It should then be possible to use it to better represent extreme events like agricultural droughts. Figure 10 represents monthly LAI anomalies averaged over the USA corn belt

(simplified as a box from 110°W to 70°W and 30°N to 50°N) with respect to 2010-2016 means from the model, the analysis and the observations. As showed by Figure 10, for the second part of the year 2012, LAI observations exhibit a strong negative anomaly at this domain scale. While it is also visible in the model, the latter clearly overestimates the intensity of the observed anomaly. The analysed LAI anomaly is closer to the observed one than the model. This extreme drought event is known as the August 2012 US corn belt drought. The U.S. Department of Agriculture (USDA, www.nass.usda.gov, last access June 2018) estimated that corn yield (per acre of planted crop) was 26% below the expectation that they had at the beginning of the 2012 growing season. The 2012 corn yield deficit and the implied climatic impact was classified as an '*historic event*' (Hoerling et al., 2014). As visible on figure 10, spring 2012 presents a positive anomaly for vegetation. Ault et al. (2013) defined spring 2012 as the earliest false spring in North American record (i.e., a period of weather in late winter or early spring allowing to bring vegetation out of dormancy prematurely). It has contributed to an earlier dry out of the soil. Figure 11 presents maps of LAI anomaly for this specific month for the model, observations and analysis from left to right, respectively. Compared to the observations (fig.11b), the area affected by the anomaly in the model (fig.11a) is too large and too intense while the analysis (fig.11c) better matches the observed pattern both in space and intensity. This impact is valid when comparing to most of the severe droughts events that occurred over CONUS (data from the National Oceanic and Atmospheric Administration -NOAA- state of the climate website, last access April-2018 <https://www.ncdc.noaa.gov/sotc/drought/201803>, not shown). Hence, LDAS-Monde provides a better tool than the model alone to monitor extreme events like agricultural droughts.

4.2 Could LDAS-Monde provide accurate initial conditions for model forecast?

In the context of NWP, assimilation of satellite observations in atmospheric models has the capacity of mitigating model deficiencies, leading to better estimates of system states. This has been the main driver of the improvement of both weather forecast skill and lead time (Bauer et al., 2015). Data assimilation is able to produce similar benefits for LSVs forecasting. Seeking to foster link with applications, LDAS-Monde could be used not only to monitor the LSVs but also be integrated in a forecasting system (at different time scales) assuming that it can provide better initial conditions than a model run and that its impact lasts in time. Many applications could benefit from a better representation of the LSVs, from NWP (de Rosnay et al., 2013), to early warning systems of e.g. agricultural drought, yield forecasts. As a first step towards such early warning systems, Figure 12 shows a comparison between LAI from the two last simulations presented in Table I: *Ini_Model* and *Ini_Analysis*. Fig.12a(b) shows monthly RMSD (R) values for the year 2016 for LAI. Not only a strong impact is visible from the beginning of the two simulations but also a few months later (up to April). The four maps of fig.12c show RMSD differences between *Ini_analysis* and *Ini_model*

from January to April. All maps are dominated by negative values suggesting that Ini_analysis present a better match with the observed LAI and that these effects last in time. Such results are strongly linked to the time of the year when the simulation is initialised by the analysis, i.e. the greater the prior difference between the model and the analysis experiments will be, the stronger the impact. As for LAI, and according to Figure 2 a marked impact would be expected from July to March. It is also very promising that the impact of LAI initialisation last in time for several weeks, or even months.

4.3 Which alternative data to better constrain LDAS-Monde?

LDAS-Monde reanalyses presented above were repeated assimilating only SSM or LAI, results suggested that most of the skill came from the assimilation of LAI (not shown). While assimilating SSM does mainly affect the first layers of soil (layer 2, 1 cm to 4 cm and layer 3, 4 cm to 10 cm), assimilating LAI has an impact on deeper layers (up to 60 cm) and is more efficient to analyse the root zone soil moisture, too. This has also been suggested by Albergel et al. (2017), when analysing the ISBA LSM sensitivity to the assimilated observations through the SEKF Jacobians.

However, the LAI product used in this study is available every 10 days at best, making it less efficient to constrain the ISBA LSM, particularly in areas of the world affected by clouds for long periods of time (e.g. areas affected by the monsoon regime).

Microwave remote sensing over land has mainly been focusing on soil moisture retrieval (Entekhabi et al., 1994; Reichle et al., 2001) and vegetation was mostly considered during the retrieval of surface soil moisture as a by-pass product affecting the signal penetration to the surface (Kurum et al., 2011; 2012). The attenuation of the signal (i.e. when passing through the vegetation) depends on the Vegetation Optical Depth (VOD). VOD describes the attenuation of radiation due to scattering and absorption within the vegetation layer, which is caused by the water contained in the vegetation. It is function of the frequency of the microwave sensor, the water content of the plant (trunk, branches, leaves) as well as on the biomass (e.g. Meesters et al., 2005; Liu et al., 2011b; Konings et al., 2016; Tian et al., 2016). VOD can be retrieved from microwave data e.g. from the L-band Soil Moisture and Ocean Salinity (SMOS) mission (Fernandez-Moran et al., 2017) or the C-band Advanced Scatterometer (ASCAT) mission on-board the meteorological operational satellite A (MetOp-A) (Vreugdenhil et al., 2016a, 2016b). VOD can be related to LAI (e.g., Zribi et al., 2011; Kim et al., 2012; Sawada et al., 2016; Momen et al., 2017). Figures 13 presents a map of temporal correlation coefficient values between modelled LAI and microwave-derived VOD from radar backscatter measurements of ASCAT (Vreugdenhil et al., 2016a, 2016b) (fig.13a) as well as their

distribution (fig.13b) for 2010-2016. High correlations values are found in large parts of the domain, with a median value of 0.57. The northern part of the domain shows R values greater than 0.7 while smaller R values (an even negative R values) are found in the southern part of the domain. Over dry soils, sub-surface scattering from the microwave signal potentially affects the VOD estimates (Wolfgang Wagner, TU WIEN, personal communication, April 2018). The same VOD dataset has a higher median R value with the observed LAI, 0.88. Consequently it better correlates with the analysis (median R values of 0.61) than with the model, also. If a strong statistical relationship between C-band VOD and LAI can be obtained through the use of e.g. machine learning techniques (like Neural Network techniques, Rodriguez-Fernandez et al., 2017), it could enable obtaining a surrogate of LAI based on C-band VOD that would have the advantage of having higher temporal frequency than the current LAI product (low frequency microwave observations are not affected much by clouds and are not affected by solar elevation). Such a product could then be assimilated into LDAS-Monde to better constrain the system. Looking at such relationship for data assimilation purposes is currently under study at CNRM.

Also, retrieved soil moisture is assimilated in LDAS-Monde, from active radar backscatter (σ_0) observations. Retrieval methods usually make use of land surface parameters and auxiliary information (like vegetation, texture and temperature), possibly being inconsistent with specific model simulations (which also include these parameters but potentially from different sources). Also, if retrievals and model simulations rely on similar types of auxiliary information, their errors may be cross-correlated, potentially degrading the system performance (De Lannoy and Reichle, 2016). This leads to an increasing tendency towards the direct assimilation of σ_0 observations (and of passive radiometer brightness temperature, T_b , as well) (De Lannoy et al., 2013; Han et al., 2014; Lievens et al., 2015, 2016). CNRM is also investigating the direct assimilation of σ_0 . It requires the implementation of a forward model for σ_0 in the ISBA LSM. Lievens et al. (2016) used the Water Cloud Model (Attema and Ulaby, 1978) to relate surface soil moisture from the Global Land Evaporation Amsterdam Model (GLEAM, Miralles et al., 2011a,b) to σ_0 for data assimilation purposes. Within LDAS-Monde, both surface soil moisture and leaf area index could be related to the radar backscatter.

5 Conclusions

In this study, LDAS-Monde sequential assimilation of satellite derived surface soil moisture and leaf area index, forced by ERA-5 latest atmospheric re-analysis was applied to the CONtiguous US

domain. Albergel et al. (2018) have highlighted the added value of using the ERA-5 atmospheric reanalysis to force the ISBA Land Surface Model over the CONtiguous US for the 2010-2016 period. They found that the use of ERA-5 instead of ERA-Interim leads to significant improvements in the representation of the land surface variables linked to the terrestrial water cycle (e.g. surface soil moisture, river discharges, snow depth and turbulent fluxes) but to a rather neutral impact on land surface variables linked to the vegetation cycle (e.g. evapotranspiration, carbon uptake and leaf area index). Assimilating satellite derived observations linked to vegetation (LAI in this application) through LDAS-Monde forced by ERA-5 not only leads to a clear improvement in the representation of the vegetation cycle in ISBA but brings further improvement on the representation of the terrestrial water cycle. Results have highlighted the stronger impact of LAI observations assimilation with respect to soil moisture assimilation. Other vegetation-related observations such as Vegetation Optical Depth could be used, under specific circumstances, as a surrogate of LAI limiting the negative impact of the rather low temporal frequency of the LAI product. LDAS-Monde is a powerful tool to track the evolution of land surface variables and to monitor extreme events such as agricultural drought. Since LDAS-Monde analysis is more accurate than a simple model run, it can be used to initialise a forecast experiment of the land surface variables. Preliminary results suggest that its impact on forecast experiments, in particular with respect to vegetation, is positive and lasts in time. It opens the way towards applications from monitoring to forecasting land surface states. For that purpose, LDAS-Monde could be forced by other ECMWF atmospheric products like the high resolution forecast (HRES, current spatial resolution of ~9km), which also gives daily forecasts up to 10 days ahead and/or the ensemble forecast (ENS, current spatial resolution of ~18km), giving daily forecasts up to 15 days (46 days twice a week).

Acknowledgement- The authors would like to thank the Copernicus Global Land Service for providing the satellite-derived LAI products and the Vienna University of Technology (Vienna, Austria) for the Vegetation Optical Depth datasets. Emanuel Dutra (Instituto Dom Luiz, IDL, Faculty of Sciences, University of Lisbon, Portugal) is thanks for his help processing the ERA-5 data.

Code availability- LDAS-Monde is a part of the ISBA land surface model and is available as open source via the surface modelling platform called SURFEX. SURFEX can be downloaded freely at <http://www.umr-cnrm.fr/surfex/> using a CECILL-C Licence (a French equivalent to the L-GPL licence; http://www.cecill.info/licences/Licence_CeCILL-C_V1-en.txt). It is updated at a relatively low frequency (every 3 to 6 months). If more frequent updates are needed, or if what is required is not in Open-SURFEX (DrHOOK, FA/LFI formats, GAUSSIAN grid), you are invited to follow the

procedure to get a SVN account and to access real-time modifications of the code (see the instructions at the first link). The developments presented in this study stemmed on SURFEX version 8.0 and are now part of the version 8.1 (revision number 4621).

Data availability-The ERA-5 datasets are distributed by ECMWF (<http://apps.ecmwf.int/datasets/>, ECMWF, last access: June 2018). The ECOCLIMAP dataset is distributed by CNRM (<https://opensource.umr-cnrm.fr/projects/ ecoclimap>, CNRM, 2013). The SURFEX model code is distributed by CNRM (<http://www.umr-cnrm.fr/surfex/>, CNRM, 2016). The satellite-derived LAI GEOV1 observations are freely accessible from the Copernicus Global Land Service (<http://land.copernicus.eu/global/>; last access: June 2018). The ESA CCI surface soil moisture dataset is distributed by ESA (<http://www.esa-soilmoisture-cci.org/>, last access: June 2018, Dorigo et al., 2017). The satellite-driven model estimates of land evapotranspiration are freely accessible at <http://www.gleam.eu> (last access: June 2018; Martens et al., 2017). The upscaled estimates of gross primary production are freely accessible at <https://www.bgc-jenna.mpg.de/geodb/projects/Home.php> (last access: June 2018; Jung et al., 2017). In situ measurements of soil moisture are freely available at <https://www.ncdc.noaa.gov/crn> (last access: June 2018; Bell et al., 2013). In situ measurements of streamflow are freely available at <https://nwis.waterdata.usgs.gov/nwis> (last access: June 2018, USGS).

References

Albergel, C., Munier, S., Leroux, D. J., Dewaele, H., Fairbairn, D., Barbu, A. L., Gelati, E., Dorigo, W., Faroux, S., Meurey, C., Le Moigne, P., Decharme, B., Mahfouf, J.-F., and Calvet, J.-C.: Sequential assimilation of satellite-derived vegetation and soil moisture products using SURFEX_v8.0: LDAS-Monde assessment over the Euro-Mediterranean area, *Geosci. Model Dev.*, **10**, 3889–3912, <https://doi.org/10.5194/gmd-10-3889-2017>, 2017.

Albergel, C., Dutra, E., Munier, S., Calvet, J.-C., Munoz-Sabater, J., de Rosnay, P., and Balsamo, G.: ERA-5 and ERA-Interim driven ISBA land surface model simulations: Which one performs better? *Hydrol. Earth Syst. Sci.*, **22**, 3515–3532, 2018, <https://doi.org/10.5194/hess-22-3515-2018>

Attema, E., Ulaby, F., 1978. Vegetation modeled as a water cloud. *Radio Sci.* **13**, 357–364.

Ault, T. R., G. M. henebry, K. M. de Beurs, M. D. Schwartz, J. L. Betancourt, and D. Moore: The False Spring of 2012, Earliest in North American Record. *Eos*, Vol. 94, No. 20, 14 May 2013

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

Barbu, A.L., J.-C. Calvet, J.-F. Mahfouf, and S. Lafont: Integrating ASCAT surface soil moisture and GEOV1 leaf area index into the SURFEX modelling platform: a land data assimilation application over France. *Hydrol. Earth Syst. Sci.*, **18**, 173–192, doi:10.5194/hess-18-173-2014, 2014.

Bauer, P., Thorpe ,A. and Brunet, G.: The quiet revolution of numerical weather prediction. *Nature* **525**, 47–55, doi:10.1038/nature14956, 2015.

Boussetta, S., Balsamo, G., Dutra, E., Beljaars, A. and Albergel, C.: Assimilation of surface albedo and vegetation states from satellite observations and their impact on numerical weather prediction, *Remote Sens. Environ.*, **163**, 111–126, <https://doi.org/10.1016/j.rse.2015.03.009>, 2015.

Calvet, J.-C., Noilhan, J., Roujean, J.-L., Bessemoulin, P., Cabelguenne, M., Olioso, A., and Wigneron, J.-P.: An interactive vegetation SVAT model tested against data from six 780 contrasting sites, *Agric. For. Meteorol.*, **92**, 73–95, 1998.

de Lannoy, G. J. M., Reichle, R. H., and Pauwels, V. R. N.: Global calibration of the GEOS-5 L-band microwave radiative transfer model over nonfrozen land using SMOS observations, *J. Hydrometeorol.*, **14**, 765–785, 2013.

de Lannoy, G. J. M. and Reichle, R. H.: Global assimilation of multiangle and multipolarization SMOS brightness temperature observations into the GEOS-5 catchment Land Surface Model for soil moisture estimation, *J. Hydrometeorol.*, **17**, 669–691, 2016

de Rosnay, P., Drusch, M., Vasiljevic, D., Balsamo, G., C. Albergel and Isaksen, L.: A simplified extended Kalman filter for the global operational soil moisture analysis at ECMWF, *Q. J. Roy. Meteor. Soc.*, **139**, 1199–1213, 2013.

Dirmeyer, P. A., Gao, X., Zhao, M., Guo, Z., Oki, T., and Hanasaki N.: The Second Global Soil Wetness Project (GSWP-2): Multi-model analysis and implications for our perception of the land surface, *B. Am. Meteorol. Soc.*, **87**, 1381–1397, <https://doi.org/10.1175/BAMS-87-10-1381>, 2006.

Dorigo, W.A., A. Gruber, R.A.M. De Jeu, W. Wagner, T. Stacke, A. Loew, C. Albergel, L. Brocca, D. Chung, R.M. Parinussa and R. Kidd: Evaluation of the ESA CCI soil moisture product using ground-based observations, *Remote Sensing of Environment*, <http://dx.doi.org/10.1016/j.rse.2014.07.023>, 2015.

Dorigo, W., Wagner, W., Albergel, C. et al.: ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions, *Remote Sensing of Environment*, <http://dx.doi.org/10.1016/j.rse.2017.07.001>, 2017.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kallberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. Roy. Meteorol. Soc.*, **137**, 553–597, doi:10.1002/qj.828, 2011.

Fairbairn, D., Barbu, A. L., Napoly, A., Albergel, C., Mahfouf, J.-F., and Calvet, J.-C.: The effect of satellite-derived surface soil moisture and leaf area index land data assimilation on streamflow simulations over France, *Hydrol. Earth Syst. Sci.*, **21**, 2015–2033, <https://doi.org/10.5194/hess-21-2015-2017>, 2017.

Fernandez-Moran, R., Wigneron, J.-P., De Lannoy, G., Lopez-Baeza, E., Parrens, M., Mialon, A., Mahmoodi, A., Al-Yaari, A., Bircher, S., Al Bitar, A., Richaume, P., Kerr, Y., A new calibration of the effective scattering albedo and soil roughness parameters in the SMOS SM retrieval algorithm, *Int J Appl Earth Obs Geoinformation*, **62**, 27–38, 2017. <https://doi.org/10.1016/j.jag.2017.05.013>

Gibelin, A.-L., Calvet, J.-C., Roujean, J.-L., Jarlan, L., and Los, S. O.: Ability of the land surface model ISBA-A-gs to simulate leaf area index at global scale: Comparison with satellite products, *J. Geophys. Res.*, **111**, 1–16 D18102, doi:10.1029/2005JD006691, 2006.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, **377**, 1–2, <https://doi.org/10.1016/j.jhydrol.2009.08.003>, 2009.

Han, X., Hendricks Franssen, H.-J., Montzka, C., and Vereecken, H.: Soil moisture and soil properties estimation in the community land model with synthetic brightness temperature observations, *Water Resour. Res.*, **50**, 6081–6105, 2014.

Hersbach, H. and Dee, D.: “ERA-5 reanalysis is in production”, ECMWF newsletter, number 147, Spring 2016, p. 7, 2016.

Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P., Gans, F., Ichii, K., Jain, A. K., Kato, E., Papale, D., Poulter, B., Raduly, B., Rödenbeck, C., Tramontana, G., Viovy, N., Wang, Y.-P., Weber, U., Zaehle, S., and Zeng, N.: Compensatory water effects link yearly global land CO₂ sink changes to temperature, *Nature*, **541**, 516–520, <https://doi.org/10.1038/nature20780>, 2017.

Kim, Y., T. Jackson, R. Bindlish, H. Lee and S. Hong, "Radar Vegetation Index for Estimating the Vegetation Water Content of Rice and Soybean," in *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 4, pp. 564-568, July 2012. doi: 10.1109/LGRS.2011.2174772

Konings, A.G., Gentine, P., 2016. Global variations in ecosystem-scale isohydricity. *Glob. Change Biol.* **23** (2), 891–905. <http://dx.doi.org/10.1111/gcb.13389>.

Kumar, S.V., M. Jasinski, D. Mocko, M. Rodell, J. Borak, B. Li, H. Kato Beudoing, and C.D. Peters-Lidard: NCA-LDAS land analysis: Development and performance of a multisensor, multivariate land data assimilation system for the National Climate Assessment. *J. Hydrometeor.*, **0**, <https://doi.org/10.1175/JHM-D-17-0125.1>, 2018.

Kurum, M., Lang, R. H., O'Neill, P. E., Joseph, A. T., Jackson, T. J., & Cosh, M. (2011). A first-order radiative transfer model for microwave radiometry of forest canopies at L-band. *IEEE Transaction on Geoscience and Remote Sensing*, **49** (9), 3167–3179.

Kurum, M., O'Neill, P.E., Lang, R.H., Joseph, A.T., Cosh, M.H., Jackson, T.J., 2012. Effective tree scattering and opacity at L-band. *Remote Sens. Environ.* **118**, 1–9.

Lettenmaier, D. P., D. Alsdorf, J. Dozier, G. J. Huffman, M. Pan, and E. F. Wood: Inroads of remote sensing into hydrologic science during the WRR era, *Water Resour. Res.*, **51**, 7309–7342, doi:10.1002/2015WR017616, 2015.

Leroux, D.J.; Calvet, J.-C.; Munier, S.; Albergel, C. Using Satellite-Derived Vegetation Products to Evaluate LDAS-Monde over the Euro-Mediterranean Area. *Remote Sens.* **2018**, *10*, 1199.

Liu, Y. Y., R. M. Parinussa, W. A. Dorigo, R. A. M. De Jeu, W. Wagner, A. I. J. M. van Dijk, M. F. McCabe, & J. P. Evans. Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. *Hydrol. Earth Syst. Sci.*, **15**, 425–436, doi:10.5194/hess-15-425-2011, 2011.

Liu, Y.Y., Jeu, R.D., McCabe, M.F., Evans, J.P., van Dijk, A.I.J.M., 2011b. Global long-term passive microwave satellite-based retrievals of vegetation optical depth. *Geophys. Res. Lett.* **38** (18). <http://dx.doi.org/10.1029/2011GL048684/abstract>.

Liu, Y. Y., W. A. Dorigo, R. M. Parinussa, R. A. M. De Jeu, W. Wagner, M. F. McCabe, J. P. Evans & A. I. J. M. Van Dijk. Trend-preserving blending of passive and active microwave soil moisture retrievals, *Remote Sens. Environ.*, **123**, 280–297, doi:10.1016/j.rse.2012.03.014, 2012.

Lievens, H., Al Bitar, A., Verhoest, N. E. C., Cabot, F., De Lannoy, G. J. M., Drusch, M., Dumedah, G., Hendricks Franssen, H.J., Kerr, Y. H., Kumar Tomer, S., Martens, B., Merlin, O., Pan, M., van den Berg, M. J., Vereecken, H., Walker, J. P., Wood, E. F., and Pauwels, V. R. N.: Optimization of a radiative transfer forward operator for simulating SMOS brightness temperatures over the Upper Mississippi Basin, USA, *J. Hydrometeorol.*, **16**, 1109–1134, 2015.

Lievens, H., Martens, B., Verhoest, N. E. C., Hahn, S., Reichle, R. H., Miralles, D. G.: Assimilation of global radar backscatter and radiometer brightness temperature observations to improve soil moisture and land evaporation estimates, *Remote Sens. Environ.*, **189**, 194–210, <https://doi.org/10.1016/j.rse.2016.11.022>, 2016.

Mätzler, C. and Standley, A.: Relief effects for passive microwave remote sensing, *Int. J. Remote Sens.*, **21**, 2403–2412, <https://doi.org/10.1080/01431160050030538>, 2000.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gobelin, A.-L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J.-F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V., and Voldoire, A.: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes, *Geosci. Model Dev.*, **990** 6, 929–960, doi: 10.5194/gmd-6-929-2013, 2013.

McNally, A., Arsenault, K., Kumar, S., Shukla, S., Peterson, P., Wang, S., Funk, C., Peters-Lidard, C. D. and Verdin, J. P.: A land data assimilation system for sub-Saharan Africa food and water security applications. *Scientific Data*, **4**, 170012, :10.1038/sdata.2017.12, 2017.

Meesters, A.G.C.A., Jeu, R.A.M.D., Owe, M., 2005. Analytical derivation of the vegetation optical depth from the microwave polarization difference index. *IEEE Geosci. Remote Sens. Lett.* **2** (2), 121–123. <http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=1420287>

Miralles, D.G., De Jeu, R.A.M., Gash, J.H., Holmes, T.R.H., Dolman, A.J., 2011a. Magnitude and variability of land evaporation and its components at the global scale. *Hydrol. Earth Syst. Sci.* **15** (3), 967–981. <http://dx.doi.org/10.5194/hess-15-967-2011>.

Miralles, D.G., Holmes, T.R.H., De Jeu, R.A.M., Gash, J.H., Meesters, A.G.C.A., Dolman, A.J., 2011b. Global land-surface evaporation estimated from satellite-based observations. *Hydrol. Earth Syst. Sci.* **15** (2), 453–469. <http://dx.doi.org/10.5194/hess-15-453-2011>.

Mitchell, K. E., et al. The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, *J. Geophys. Res.*, 109, D07S90, 2004. doi:10.1029/2003JD003823

Momen, M., Wood, J. D., Novick, K. A., Pangle, R., Pockman, W. T., McDowell, N. G., & Konings, A. G.: Interacting effects of leaf water potential and biomass on vegetation optical depth. *Journal of Geophysical Research: Biogeosciences*, 122, 3031–3046, 2017. <https://doi.org/10.1002/2017JG004145>

Noilhan, J., and J.-F. Mahfouf, The ISBA land surface parameterisation scheme, *Global Planet. Change*, 13, 145–159, 1996.

Nash, J.E., Sutcliff, V.: River forecasting through conceptual models, *Journal of Hydrology*, 10:282-290, 1970.

Oki T., and Y. C. Sud: Design of Total Runoff Integrating Pathways (TRIP). A global river channel network, *Earth Interaction*, 2, 1-36, 1998.

Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L., and Morel, S.: Analysis of near surface atmospheric variables: validation of the SAFRAN analysis over France, *J. Appl. Meteorol. Clim.*, 47, 92–107, 2008.

Quintana-Seguí, P., Turco, M., Herrera, S., and Miguez-Macho, G.: Validation of a new SAFRAN-based gridded precipitation product for Spain and comparisons to Spain02 and ERA-Interim, *Hydrol. Earth Syst. Sci.*, 21, 2187-2201, <https://doi.org/10.5194/hess-21-2187-2017>, 2017.

Reichle, R. H., R. D. Koster, P. Liu, S. P. P. Mahanama, E. G. Njoku and M. Owe: Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *J. Geophys. Res.* 112 (D9): D09108 (10.1029/2006JD008033), 2007.

Rodell, M., P. R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J. K. Entin, J. P. Walker, D. Lohmann, and D. Toll, The Global Land Data Assimilation System, *Bull. Amer. Meteor. Soc.*, 85(3), 381–394, 2004.

Rodríguez-Fernández, N. J., Muñoz Sabater, J., Richaume, P., de Rosnay, P., Kerr, Y. H., Albergel, C., Drusch, M., and Mecklenburg, S.: SMOS near-real-time soil moisture product: processor overview and first validation results, *Hydrol. Earth Syst. Sci.*, 21, 5201-5216, <https://doi.org/10.5194/hess-21-5201-2017>, 2017.

Sawada, Y., and T. Koike, Simultaneous estimation of both hydrological and ecological parameters in an ecohydrological model by assimilating microwave signal, *J. Geophys. Res. Atmos.*, 119, doi:10.1002/2014JD021536, 2014.

Sawada, Y., T. Koike, and J. P. Walker, A land data assimilation system for simultaneous simulation of soil moisture and vegetation dynamics, *J. Geophys. Res. Atmos.*, 120, doi: 10.1002/2014JD022895, 2015.

Sawada, Y., H. Tsutsui, T. Koike, M. Rasmy, R. Seto and H. Fujii, "A Field Verification of an Algorithm for Retrieving Vegetation Water Content From Passive Microwave Observations," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 4, pp. 2082-2095, April 2016. doi: 10.1109/TGRS.2015.2495365

Schellekens, J., Dutra, E., Martínez-de la Torre, A., Balsamo, G., van Dijk, A., Sperna Weiland, F., Minvielle, M., Calvet, J.-C., Decharme, B., Eisner, S., Fink, G., Flörke, M., Peßenteiner, S., van Beek, R., Polcher, J., Beck, H., Orth, R., Calton, B., Burke, S., Dorigo, W., and Weedon, G. P.: A global water resources ensemble of hydrological models: the earth2Observe Tier-1 dataset, *Earth Syst. Sci. Data*, 9, 389-413, <https://doi.org/10.5194/essd-9-389-2017>, 2017.

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., and

Papale, D.: Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms, *Biogeosciences*, **13**, 4291–4313, <https://doi.org/10.5194/bg-13-4291-2016>, 2016.

Scipal, K., Drusch, M., and Wagner, W.: Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system, *Adv. Water Resour.*, **31**, 1101–1112, <https://doi.org/10.1016/j.advwatres.2008.04.013>, 2008.

Tian, F., Brandt, M., Liu, Y.Y., Verger, A., Tagesson, T., Diouf, A.A., Rasmussen, K., Mbow, C., Wang, Y., Fensholt, R., 2016. Remote sensing of vegetation dynamics in drylands: evaluating vegetation optical depth (VOD) using AVHRR NDVI and in situ green biomass data over West African Sahel. *Remote Sens. Environ.* **177**, 265 – 276. <http://linkinghub.elsevier.com/retrieve/pii/S0034425716300852>.

Vreugdenhil, M., Dorigo, W.A., Wagner, W., Jeu, R.A.M.d., Hahn, D., Marle, M.J.E.v., 2016a. Analyzing the vegetation parameterization in the TU-Wien ASCAT soil moisture retrieval. *IEEE Trans. Geosci. Remote Sens.* **54** (6), 3513–3531. <http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=7410033>.

Vreugdenhil, M., Hahn, S., Melzer, T., Bauer-Marschallinger, B., Reimer, C., Dorigo, W.A., Wagner, W., 2016b. Assessing vegetation dynamics over Mainland Australia With Metop ASCAT. *IEEE J. Select. Top. Appl. Earth Observ. Remote Sens.* **PP** (99), 1–9.

Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century, *J. Hydrometeorol.*, **12**, 823–848, <https://doi.org/10.1175/2011JHM1369.1>, 2011.

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing data methodology applied to ERA- interim reanalysis data, *Water Resour. Res.*, **50**, 7505–7514, <https://doi.org/10.1002/2014WR015638>, 2014.

Xia, Y., et al. 2012, Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, *J. Geophys. Res.*, **117**, D03109, doi:10.1029/2011JD016048, 2012.

Zribi, M., Chahbi, A., Shabou, M., Lili-Chabaane, Z., Duchemin, B., Baghdadi, N., Amri, R., Chehbouni, A., 2011. Soil surface moisture estimation over a semi-arid region using ENVISAT ASAR radar data for soil evaporation evaluation. *Hydrol. Earth Syst. Sci. Discuss.* **15** (1), 345–358. <http://hal.ird.fr/ird-00610514/document>

Tables**Table I:** Set up of the experiments used in this study

Experiments (time period)	Model	Domain & spatial resolution	Atmospheric forcing	DA method	Assimilated observations	Observation s operators	Control variables	Additional options
Model or Openloop (2010-2016)	ISBA Multi-layer soil model CO2-responsive version (Interactive vegetation)	CONtiguous US (CONUS), 0.25°x0.25°	ERA-5	N/A	N/A	N/A	N/A	Coupling with CTRIP (0.5°)
Analysis (2010-2016)	ISBA Multi-layer soil model CO2-responsive version (Interactive vegetation)	CONtiguous US (CONUS), 0.25°x0.25°	ERA-5	SEKF	SSM (ESA-CCI) LAI (GEOV1)	Rescaled WG2 (Second layer of soil (1-4cm)) LAI	Layers of soil 2 to 8 (WG2 to WG8, 1-100cm) LAI	Coupling with CTRIP (0.5°)
Ini_Model (2016)	ISBA Multi-layer soil model CO2-responsive version (Interactive vegetation)	CONtiguous US (CONUS), 0.25°x0.25°	ERA-5	12-month model run starting on 1 January 2016 (initialised by the model simulation run from 2010 to 2015)				Coupling with CTRIP (0.5°)
Ini_Analysis (2016)	ISBA Multi-layer soil model CO2-responsive version (Interactive vegetation)	CONtiguous US (CONUS), 0.25°x0.25°	ERA-5	12-month model run starting on 1 January 2016 (initialised by the analysis run from 2010 to 2015)				Coupling with CTRIP (0.5°)

Table II: Analysis impact evaluation against in situ measurements of soil moisture from the USCRN network. In situ measurements at a depth of 5cm are used to evaluate soil moisture from the third layer of soil (4-10 cm) from either the model or analysis experiment over 2010-2016. The Normalized Information Contribution (NIC) is applied to the correlation (anomaly correlations) values. NIC scores are classified according to three categories: (i) negative impact from the analysis with respect to the model with values smaller than -3 %, (ii) positive impact from the analysis with respect to the model with values greater than +3 % and (iii) neutral impact from the analysis with respect to the model with values between -3 % and 3 %.

110 (110) stations with significant R (Anomaly R)	Median R (Anomaly R)	Median ubRMSD	Positive impact: >+3	↔-3 Negative impact: <-3	Neutral impact [-3;+3]
Model	0.72±0.02* (0.60±0.02*)	0.049±0.004*	N/A	N/A	N/A
Analysis	0.74±0.02* (0.60±0.02*)	0.048±0.004*	46%(18%)	8%(1%)	46%(81%)

*95% confidence interval of the median derived from a 10000 samples bootstrapping

Table III: Analysis impact evaluation against daily streamflow over 2010-2016. The impact from the analysis with respect to the model is assessed through the Normalized Information Contribution (NIC) applied to the Kling-Gupta efficiency (KGE) score as well as using normalized relative error of simulated or analysed mean (RE_{μ}) and standard deviation (RE_{σ}). Scores are classified according to three categories: (i) negative impact from the analysis with respect to the model with values smaller than -3, (ii) positive impact from the analysis with respect to the model with values greater than +3 and (iii) neutral impact from the analysis with respect to the model with values between -3 and 3.

258 out of 531 stations with KGE greater than 0	Positive impact: >+3	Negative impact: <-3	Neutral impact [-3;+3]
NIC_{KGE}	26%	12%	62%
$N_{RE_{\sigma}}$	22%	1%	77%
$N_{RE_{\mu}}$	34%	1%	65%

Figures

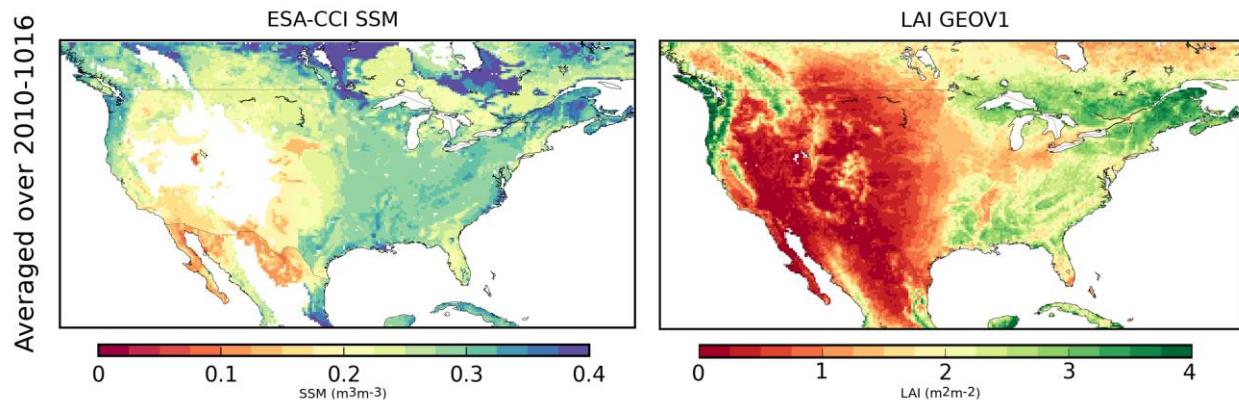


Figure 1: Averaged (a) surface soil moisture from the Climate Change Initiative project of ESA (for pixels with less than 15% of urban areas and with an elevation of less than 1500 m above sea level), (b) GEOV1 leaf area index from the Copernicus Global Land Service project (for pixels covered by more than 90 % of vegetation) from 2010 to 2016.

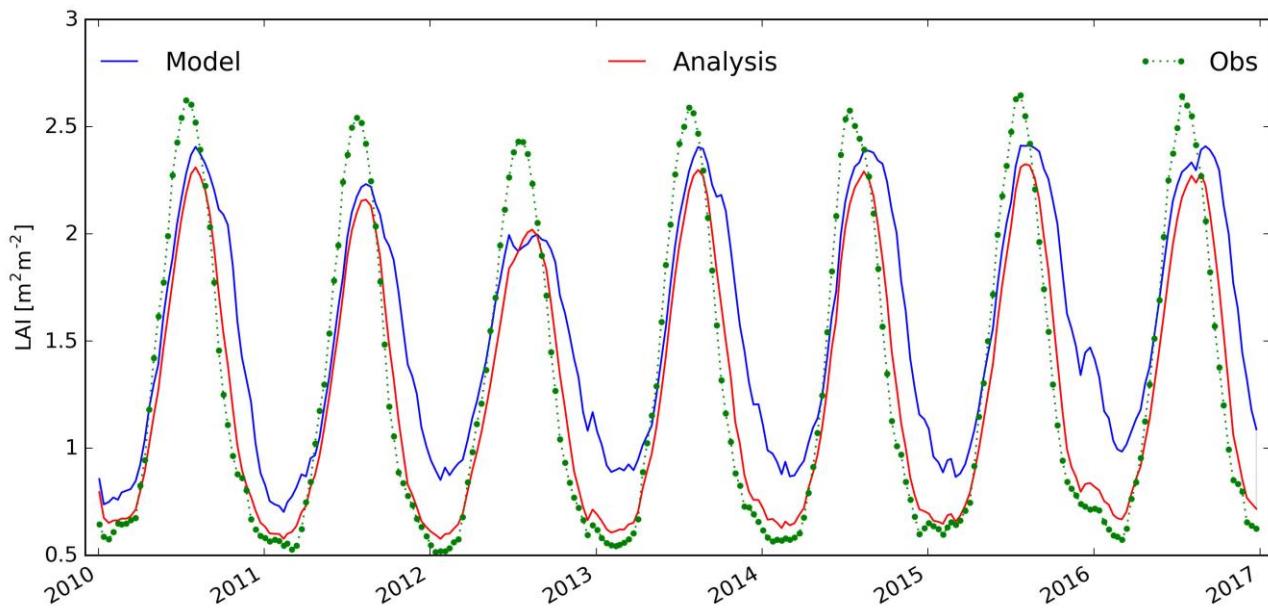


Figure 2: Leaf Area Index time series from the model (blue line), the observations (green dots and dashed line) and the analysis (red line) averaged over the whole domain from 2010 to 2017.

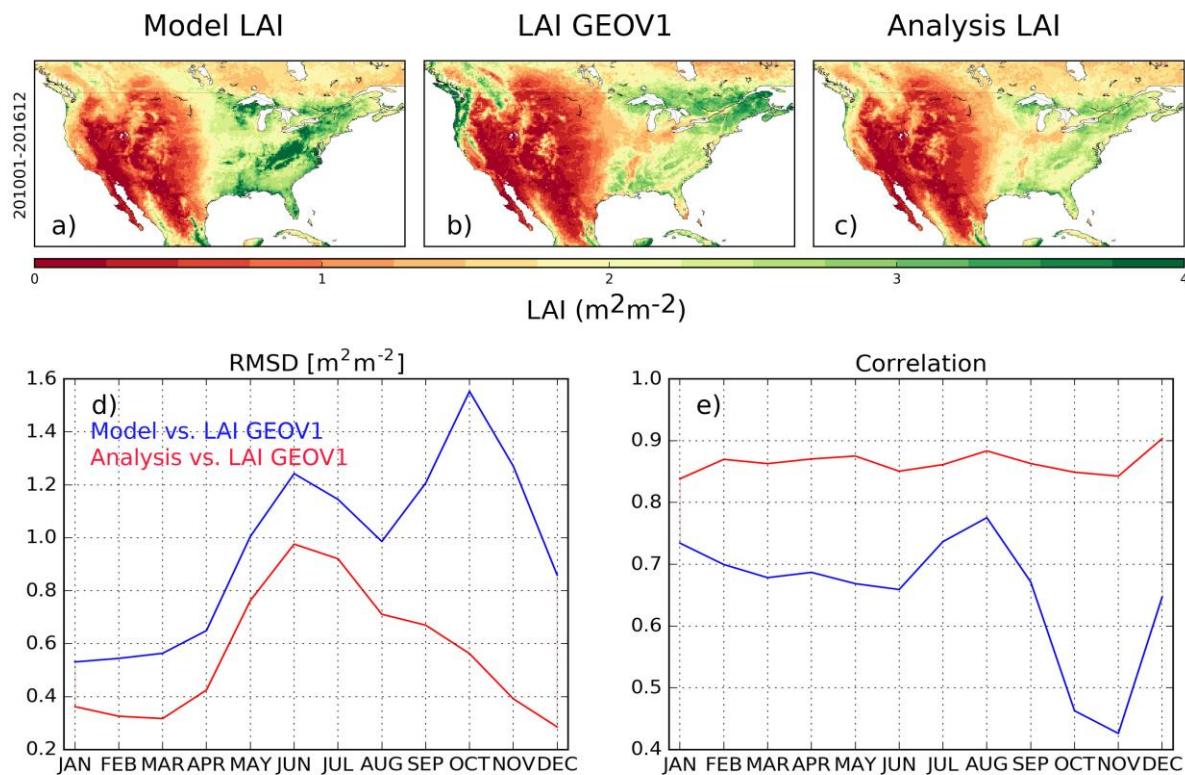


Figure 3: Top row; Leaf Area Index from (a) the model, (b) the observations and (c) the analysis averaged over the 2010-2016 time period. Bottom row: Seasonal (d) RMSD and (e) correlation values between leaf area index (LAI) from the model (in blue), the analysis (in red) and GEOV1 LAI estimates from the Copernicus Global Land Service project from 2010 to 2016.

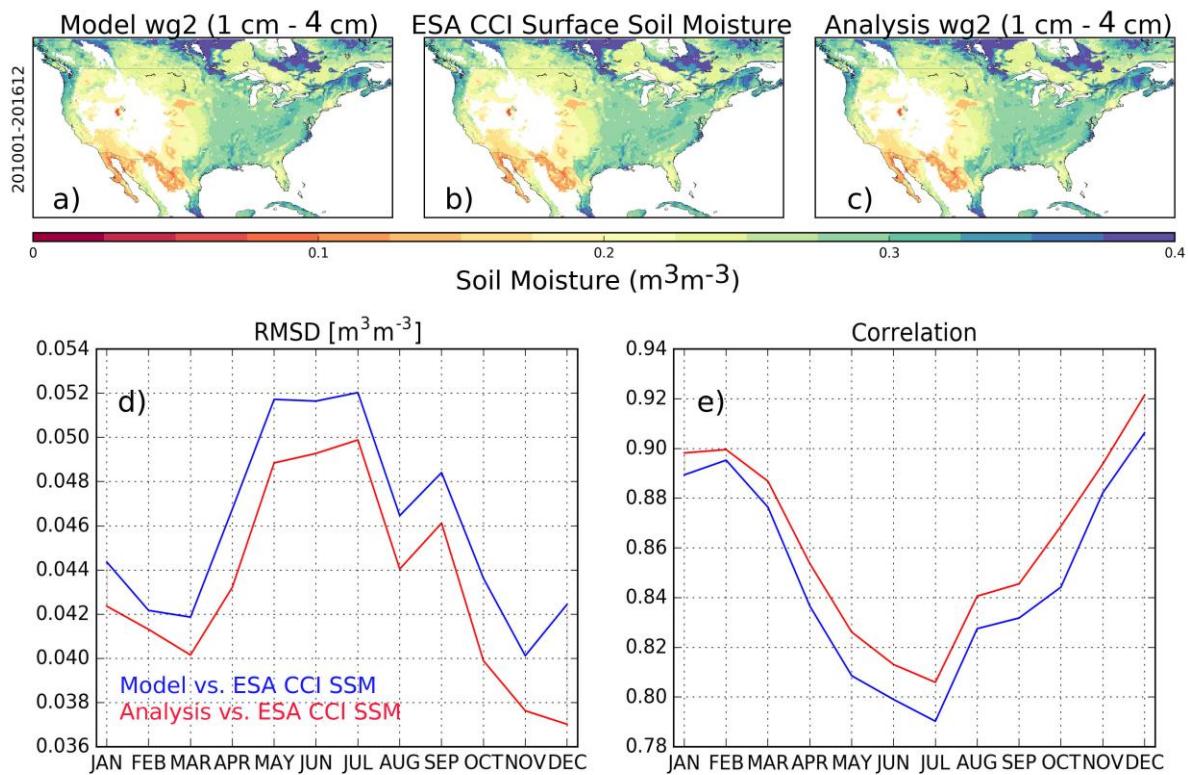


Figure 4: Same as Figure 3 for soil moisture.

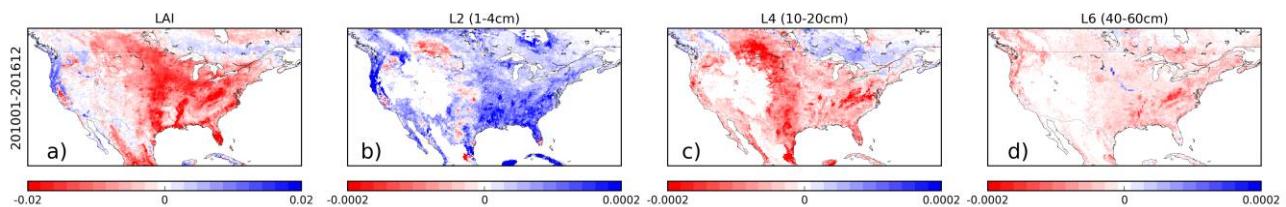


Figure 5: Analysis increments averaged over the 2010-2016 time period for (a) LAI in $\text{m}^2 \text{m}^{-2}$, (b) second (c) fourth and (d) sixth layer of soil moisture in $\text{m}^3 \text{m}^{-3}$.

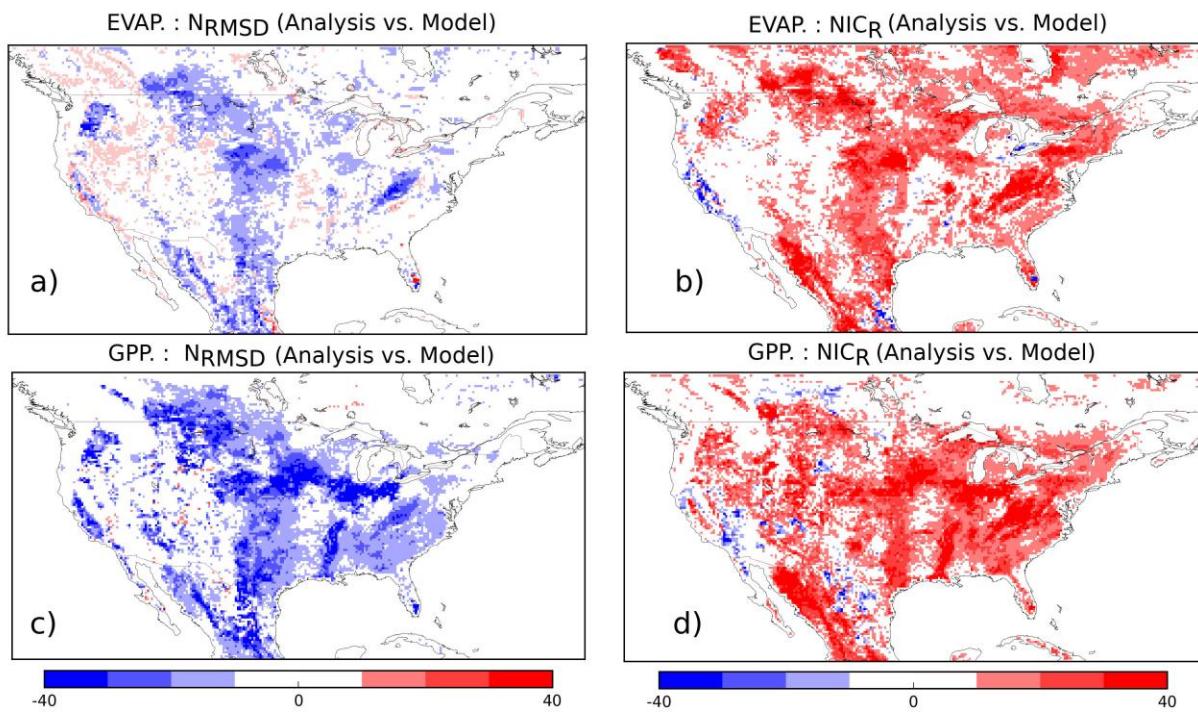


Figure 6: Top row: (a) normalized RMSD (blue colours indicate an improvement) and (b) NIC applied on Correlations values (red colours indicate an improvement) for evapotranspiration from the analysis with respect to the model. Bottom row: same as top row for Gross Primary Production. Units are percent.

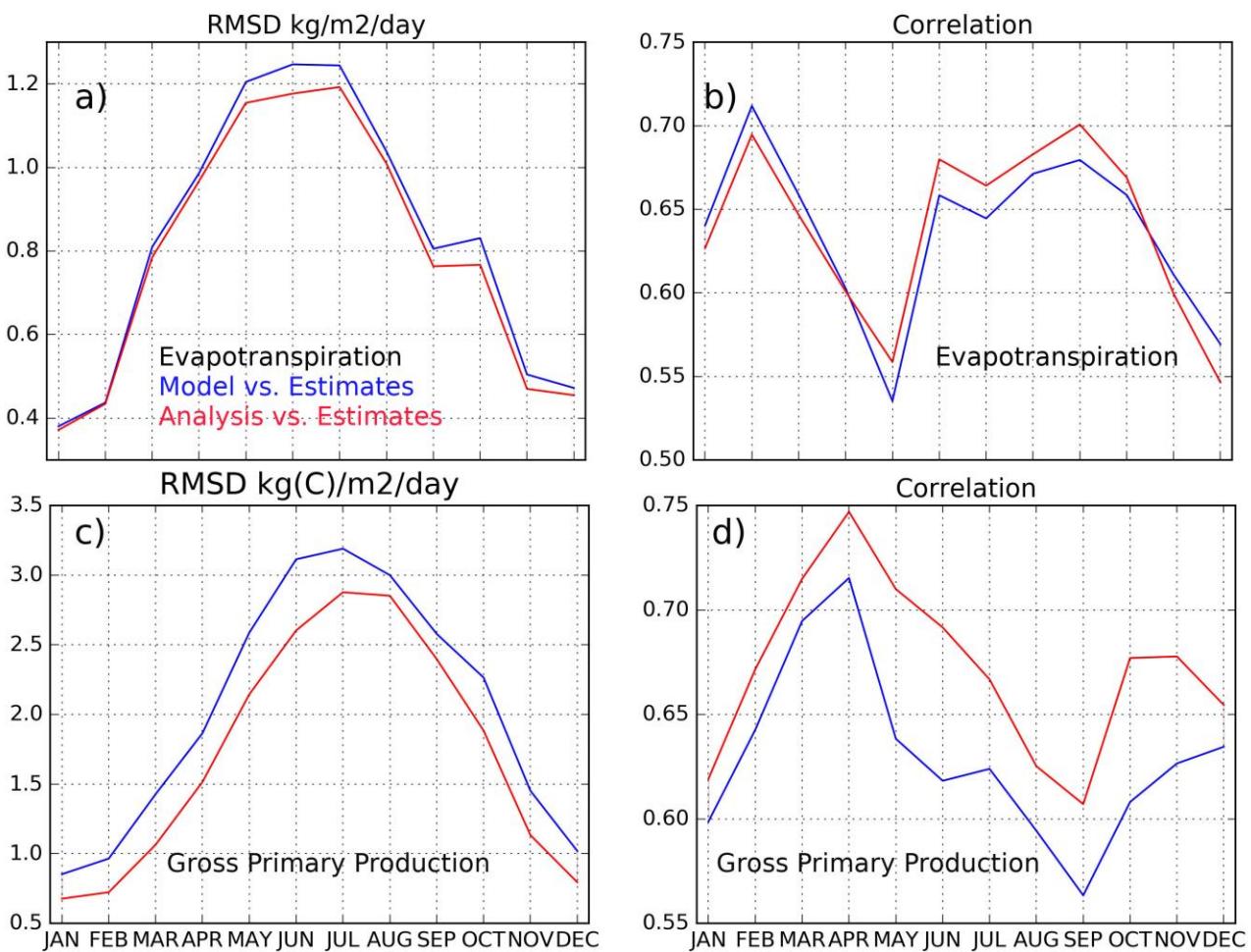


Figure 7: Top row: seasonal (a) RMSD and (b) correlation values between Evapotranspiration from the model (in blue), the analysis (in red) and estimates from the GLEAM project over 2010–2016. Bottom row: (c) and (d), same as (a) and (b) for Gross Primary Production from either the model or the analysis and estimates from the FLUXCOM project.

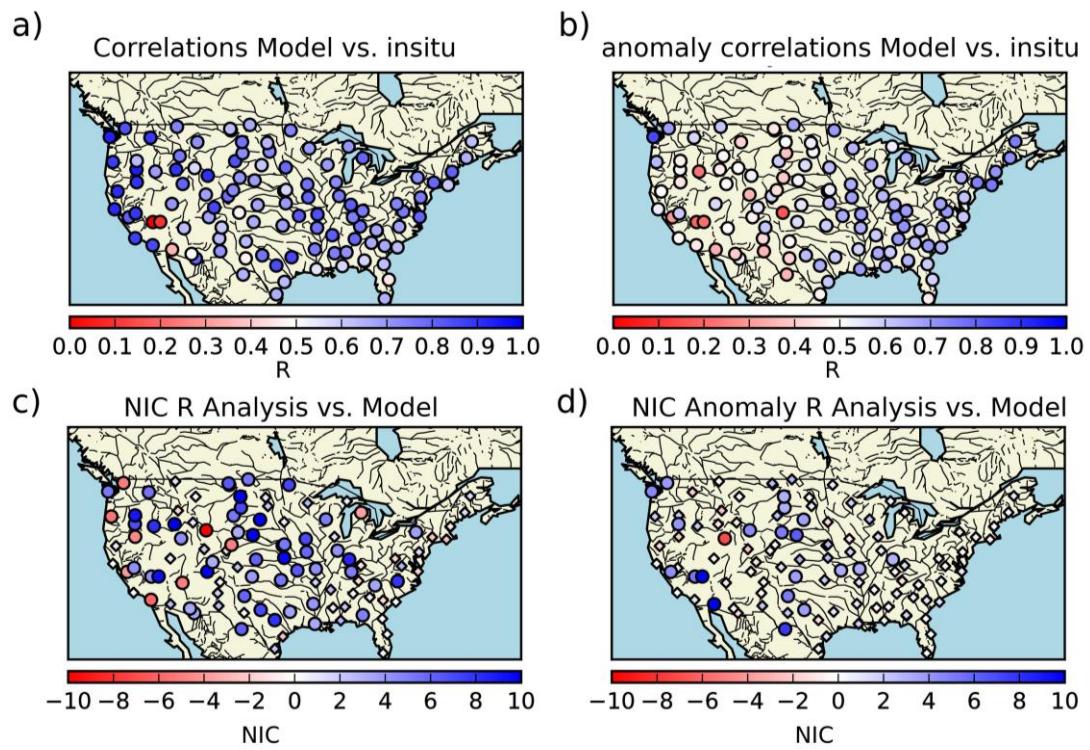


Figure 8: Maps of correlation (R) on volumetric time-series (a) and anomaly time-series (b) between in situ measurements at 5 cm depth from the USCRN network and soil moisture from the model (third layer of soil between 4 cm and 10 cm) from 2010 to 2016. NIC applied on R (anomaly R) values (c and d); analysis with respect to the model. NIC scores are classified according to three categories: (i) negative impact from the analysis with respect to the model with values smaller than -3 % (red circles), (ii) positive impact from the analysis with respect to the model with values greater than +3 % (blue circles) and (iii) neutral impact from the analysis with respect to the model with values between -3 % and 3 % (diamonds).

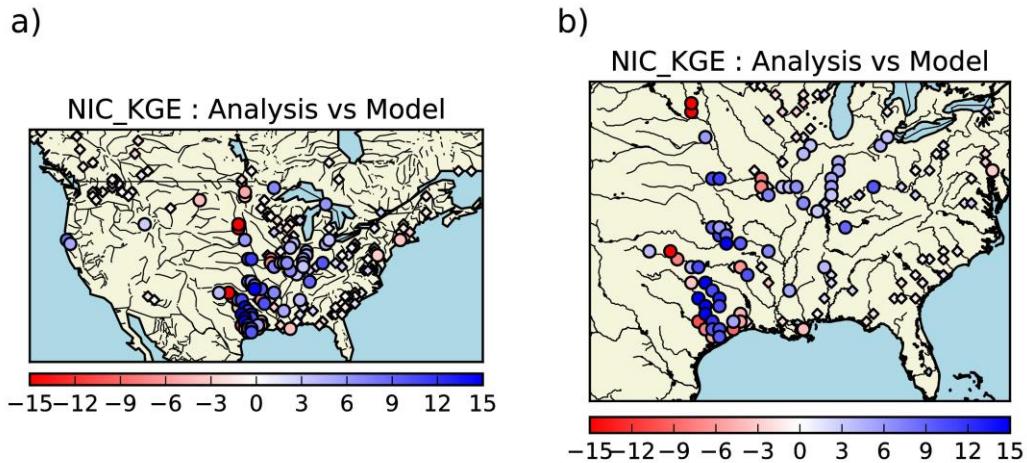


Figure 9: Normalized information contribution scores based on KGE scores (NIC_{KGE}) (a) analysis with respect to the model, (b) zoom over the eastern part of the domain. Small diamonds represent stations for which NIC_{KGE} are between $[-3;+3]$. NIC_{KGE} greater than 3 (blue large circles) suggest an improvement from the analysis over the model, values smaller than -3 (red large circles) a degradation. For sake of clarity, a factor of 100 has been applied to NIC.

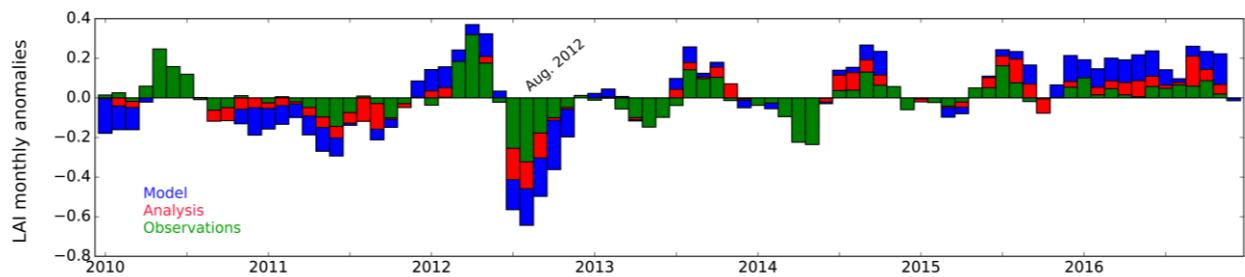


Figure 10: Leaf Area Index monthly anomalies over the 2010-2016 time period for the model (blue bars), the analysis (red bars) and the CGLS GEOVI observations (in green) over the corn belt drought defined as a box from 110°W to 70°W and 30°N to 50°N .

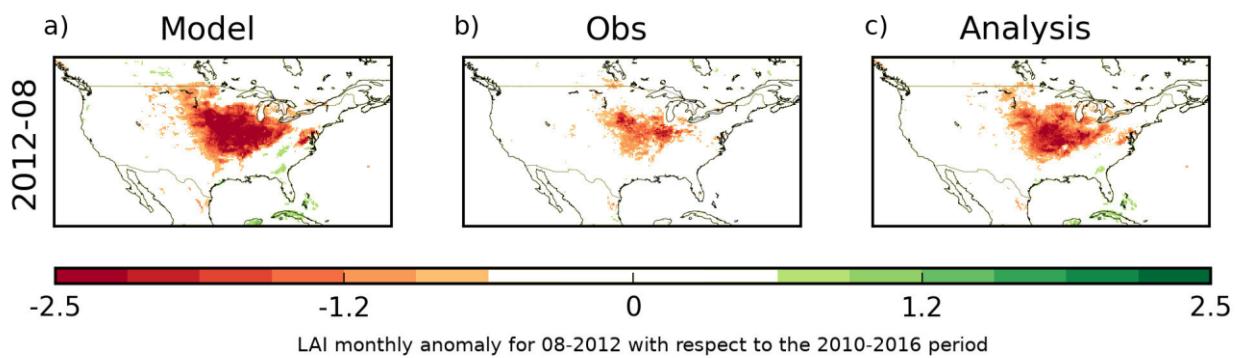


Figure 11: (a) Monthly anomaly of Leaf Area Index for August 2012 with respect to the 2010-2016 period, (b) same as (a) for observed Leaf Area Index, (c) same as (a) for analysed Leaf Area Index.

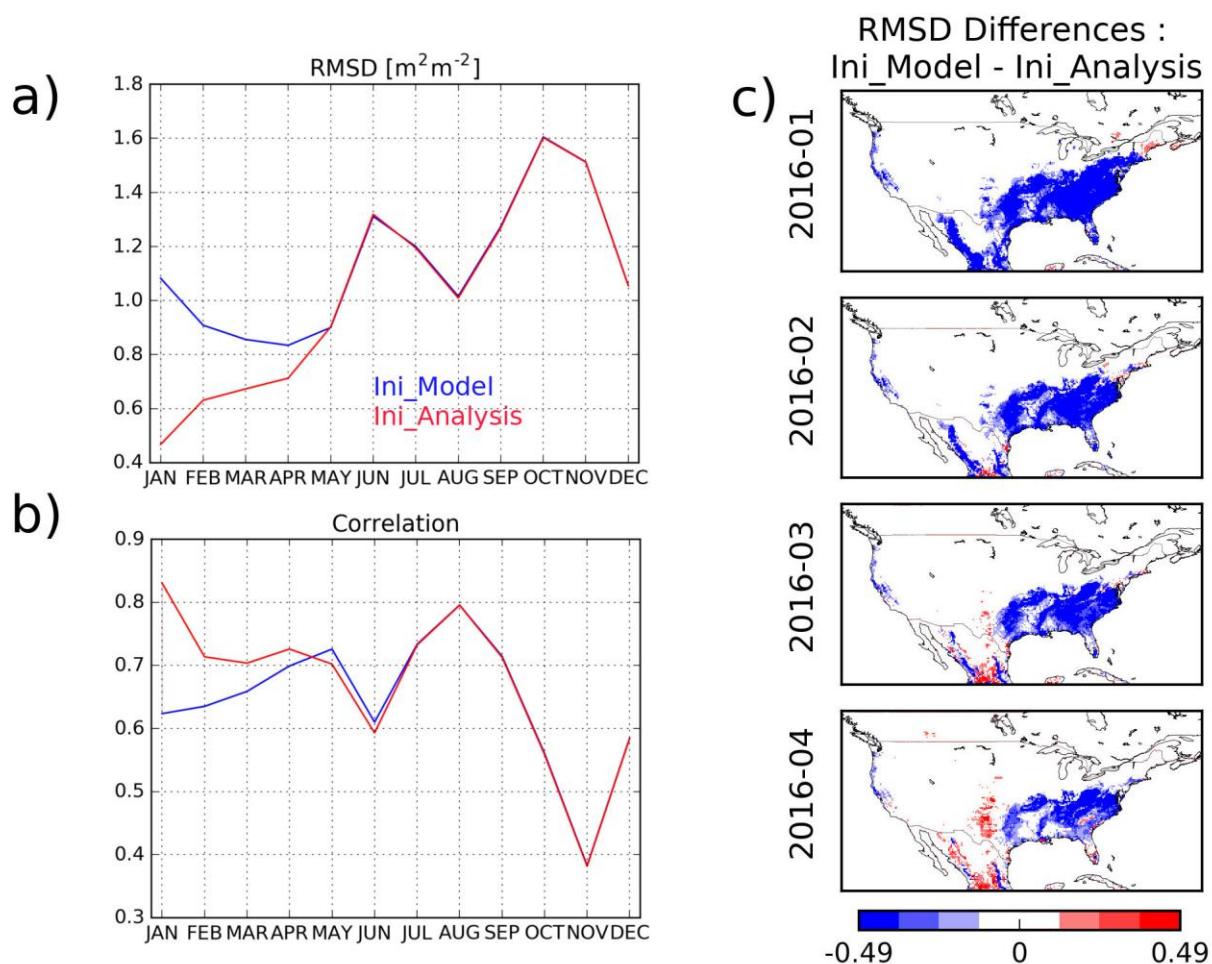


Figure 12: Seasonal (a) RMSD and (b) correlation values between observed Leaf Area Index (LAI) and (in blue) a 12-month model run, (in red) a 12-month model run initialised by analysed conditions from LDAS-Monde. (c) RMSD differences values between a 12-month model run and a 12-month model run initialised by analysed conditions from LDAS-Monde.

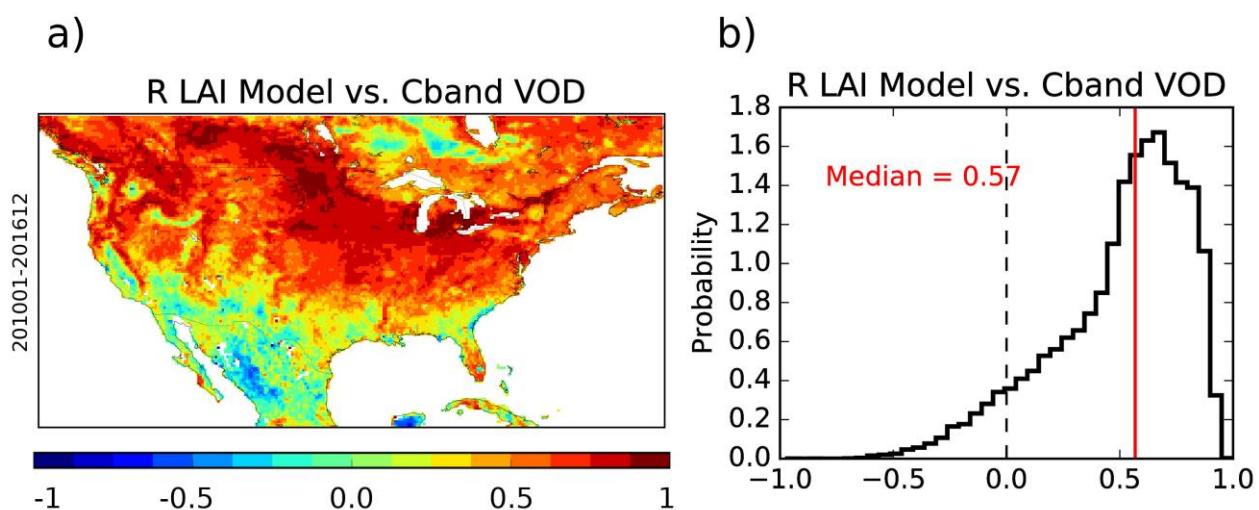


Figure 13: (a) Correlation coefficient values between modelled LAI and C-band VOD over 2010-2016, (b) probability distribution of the correlation coefficient values over the same period