Towards a Long-Term Reanalysis of Land Surface Variables over Western Africa: LDAS-Monde Applied over Burkina Faso from 2001 to 2018

Moustapha Tall (1,2,3), Clément Albergel (3*), Bertrand Bonan (2), Yongjun Zheng (3), Françoise Guichard (3), Mamadou S. Dramé (2), Amadou T. Gaye (2), Luc O. Sintondji (5), Fabien C. C. Hountondji (6), Pinghouinde Michel Nikiema (7), Jean-Christophe Calvet (3)

(1) Climate Change and Water Resources, West African Science Service Centre on Climate Change and Adapted Land Use (WASCAL), Université D’abomey Calavi, 03 BP 526 Cotonou, Benin
(2) Laboratoire de Physique de l’Atmosphère et de l’Océan, Ecole Supérieure Polytechnique, Université Cheikh Anta Diop, Dakar, Senegal
(3) CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France
(4) Département de Physique, Faculté des Sciences et Techniques, Université Cheikh Anta Diop, Dakar, Senegal
(5) Laboratory of Hydraulics and Water Control, National Institute of Water, University of Abomey-Calavi (LHME/INE/UAC), Abomey-Calavi, Benin
(6) Faculté d’Agronomie, Université de Parakou, Parakou BP 123, Benin
(7) ANAM, Agence Nationale de la Météorologie, BP 576 Ouagadougou, Burkina-Faso

Correspondence: clement.albergel@meteo.fr

Abstract. This study focuses on the ability of the global land data assimilation system LDAS-Monde to improve the representation of land surface variables (LSVs) over Burkina Faso through the joint assimilation of satellite derived Surface Soil Moisture (SSM) and Leaf Area Index (LAI) from January 2001 to June 2018. The LDAS-Monde offline system is forced by the latest European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis ERA5, leading to a 0.25° x 0.25° spatial resolution reanalysis of the LSVs. Within LDAS-Monde, SSM and LAI observations from the Copernicus Global Land Service (CGLS) are assimilated using the CO2-responsive version of the ISBA (Interactions between Soil, Biosphere and Atmosphere) land surface model (LSM). First, it is shown that ERA5 better represents precipitation and incoming solar radiation than ERA-Interim former reanalysis from ECMWF. Results of two experiments are compared: open-loop simulation (i.e. no assimilation) and analysis (i.e. joint assimilation of SSM
and LAI). After jointly assimilating SSM and LAI, it is noticed that the assimilation is able to impact soil moisture in the first top soil layers (the first 20 cm), and also in deeper soil layers (from 20 cm to 60 cm and below). The assimilation is able to improve the simulation of both SSM and LAI. For LAI in particular, the southern region of the domain (dominated by a Sudan-Guinean climate) highlights a strong impact of the assimilation compared to the other two sub-regions of Burkina Faso (dominated by Sahelian and Sudan-Sahelian climates). In the southern part of the domain, differences between the model and the observations are the largest, prior to any assimilation. These differences are linked to the model failing to represent the behavior of some specific vegetation species, which are known to put on leaves before the first rains of the season. The LDAS-Monde analysis is very efficient at compensating for this model weakness. Evapotranspiration estimates from the Global Land Evaporation Amsterdam Model (GLEAM) project as well as upscaled carbon uptake from the FLUXCOM project are used in the evaluation process, again demonstrating improvements in the representation of evapotranspiration and gross primary production after assimilation.

1. Introduction

An accurate representation of land surface variables (LSVs) such as soil moisture or vegetation cover is critical in climate science as well as environmental monitoring and prediction (e.g. in order to cope with drought, flood or other extreme events). To that end, land surface models (LSMs) have been widely used to simulate and predict the Earth’s water storage energy budgets over a broad range of time scales [1–4]. For instance, the AMMA (African Monsoon Multidisciplinary Analysis) Land Surface Model Intercomparison Project (ALMIP) used a set of LSMs forced in offline mode by a combination of satellite products and high quality in situ measurements in order to better apprehend LSV processes and their representation [5,6]. These LSMs are intended to reproduce LSVs such as surface and root zone soil moisture (SSM and RZSM, respectively), vegetation biomass and leaf area index (LAI), together with surface energy fluxes and streamflow simulations.
Over the last two decades, much progress has been made on the degree to which realistic land surface initialization contributes to the skill and performance of subseasonal land-related predictability as documented by [7,8] in the Global Land-Atmosphere Coupling Experiment (GLACE). LSMs have subsequently benefited from the growing development of observational networks. Unfortunately those are not evenly spaced and data sparse regions remain very difficult to model with accuracy. This is the case of West Africa ([5] and [9]), where LSVs are of primary importance, as emphasized by many studies, see e.g. [10] and [11]. In this context, the present study aims to evaluate reanalyses of LSVs obtained with LDAS-Monde over Burkina Faso (domain shown in Figure 1). This country exhibits three distinctive climates: Sahel, Sudan-Sahel and Sudan-Guinea that cover most part of West Africa, making Burkina Faso an area of interest for such study. Land Data Assimilation Systems (LDASs) combine LSMs with satellite observations in order to produce reanalyses of LSVs. LDAS-Monde [12] is based on the CO₂-responsive version of the Interactions between Soil, Biosphere and Atmosphere (ISBA) LSM [13–16] available through the SURFEX (SURFace EXternalisée; [17]) modelling system of Météo-France. The reanalysis is performed by assimilating jointly satellite-derived SSM and LAI using a Simplified Extended Kalman Filter (SEKF). For that purpose, the most recent SURFEX_v8.1 Offline Data Assimilation (SODA) implementation has been utilized considering a long-term period (January 2001 to June 2018) along with the joint assimilation of both satellite-derived SSM (from 2007, [18]) and LAI (GEOV2, from 2001, http://land.copernicus.eu/global/, last access November 2018).

The quality of LSV reanalyses depends on the quality of LSMs and observations used but also on the quality atmospheric forcings used by LSMs. Numerous improvements were made in the generation of long-term (1979-onwards) global atmospheric reanalyses, leading to more advancements in land surface modeling fields and their applications (e.g. water resources monitoring [2,19,20]. In line with those improvements, NASA’s Modern Era Retrospective analysis for Research and Applications (MERRA; [21], and MERRA2; [22]) as well as ECMWF’s (European Centre for Medium-Range Weather Forecasts) Interim reanalysis (ERA-Interim; [23])
were the most investigated. We take advantage of the recent development of ERA5, which was released in 2017 as the fifth generation of ECMWF global atmospheric reanalyses. At the time of this study, a time-slice of the ERA5 database was available from 2001 within 3 months of real time. ERA5 brings extensive changes compared to ERA-Interim including higher spatial and temporal resolutions as well as a generally improved representation of e.g. precipitation and incoming solar radiation (SW\text{in}) [4,24,25]. The performance of ERA5 and ERA-Interim precipitation and SW\text{in} is first investigated using in situ measurements over BF before studying the quality of LDAS-Monde reanalysis of LSVs.

**Figure 1:** Averaged (a) surface soil moisture from ASCAT. (b) GEOV2 leaf area index (for pixels covered by more than 90 % of vegetation) from the Copernicus Global Land Service project. The symbols SH (Sahel), SS (Sudan-Sahel) and SG (Sudan-Guinea) represent the three agroclimatic regions across Burkina Faso (BF).

Section 2 provides (i) a description of LDAS-Monde including details on the CO\textsubscript{2}-responsive version of the ISBA LSM and the data assimilation system, (ii) information on the atmospheric reanalyses used to force the system, the assimilated remotely sensed observations along with the in-situ and satellite datasets used to assess the sensitivity of the results to the atmospheric reanalysis, and (iii) the experimental set up and the evaluation strategies. Section 3 presents the results of a
performance assessment of the reanalyses against in situ measurements, the assimilation impact of
the assimilated variables as well as independent satellite datasets and section 4 provides
perspectives and future directions.

2. Materials and Methods

2.1 LDAS-Monde

The LDAS [26–30] developed at Météo France’s Centre National de Recherches Météorologiques
(CNRM) allows the integration of satellite-derived products into the ISBA LSM using a data
assimilation scheme. It has been recently expanded to global scale (i.e. LDAS-Monde) [12,31,32].
LDAS-Monde is accessible through the open-access SURFEX modelling platform [17]
(https://www.umr-cnrm.fr/surfex, last accessed January 2019). The following subsections quickly
recall the main components of LDAS-Monde. More details can be found in [12].

2.1.1 ISBA land surface model

LDAS-Monde uses in this paper the CO\textsubscript{2}-responsive [14–16], the multi-layer soil water and heat
transfer [33,34] version of ISBA. In this configuration, ISBA describes the water and carbon fluxes
on a subdaily frequency, plant growth and crucial variables of vegetation components such as LAI
and above ground-biomass.

In the ISBA model, the evolution of vegetation variables is controlled by photosynthesis, which
enables vegetation growth resulting from the CO\textsubscript{2} uptake. A minimum LAI threshold is prescribed
as 1 m\textsuperscript{2} m\textsuperscript{−2} for coniferous forest or 0.3 m\textsuperscript{2} m\textsuperscript{−2} for other vegetation types. Conversely, a lack of
photosynthesis triggers higher mortality rates. The carbon uptake related to photosynthesis
represents the gross primary production (GPP) while the CO\textsubscript{2} released by the soil-plant system
constitutes the ecosystem respiration (RECO). The difference between these two quantities corresponds to the net ecosystem CO$_2$ exchange (NEE).

In the model version used in this study, ISBA parameters are prescribed for 12 generic land surface types, which consist of (i) 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), (ii) bare soil, (iii) rocks, and (iv) permanent snow and ice surfaces. Those parameters are derived from the ECOCLIMAP land cover database [35].

2.1.2 Data assimilation

LDAS-Monde routinely uses a Simplified Extended Kalman Filter (SEKF, [27]) to assimilate observations of SSM and LAI. This approach considers the observation operator as the product of the model propagation of control variables from one analysis time step to the next one with the projection of those variables to observation equivalents. The analysis requires the calculation of the Jacobian of the observation operator. It is computed using finite differences obtained by perturbed model runs. In this study, the analysis step updates modeled LAI and soil moisture from layer 2 (1–4 cm) to layer 7 (60–80 cm). The approach is fully detailed in [12].

2.2 Datasets and data processing

2.2.1 In situ measurements

In this study, in situ data of precipitation for the 2010-2016 period were provided by the General Directorate of Meteorology (GDM) of Burkina Faso (BF), as previously used by [36]. It consists of 134 stations which are relatively well spread over the country except for less density in the north of BF also called Sahel zone (SH) and eastern part of Sudan-Sahel zone (SS) of the domain (Figure 2). All stations include a daily time series of good quality (with few missing data) over the considered period. The in situ measurements of SW$_{in}$ are also from the GDM with data available every 15
minutes for 4 stations (Figure 3). In the present study, 24h-mean values of these radiative fluxes are used.

2.2.2 ERA-Interim and ERA5 atmospheric reanalyses

ERA-Interim is a global atmospheric reanalysis produced by the ECMWF [23]. Reanalyses provide a numerical description of the recent climate by combining models with observations using data assimilation systems. ERA-Interim overlays the period from 1 January 1979 onward and continues to be extended forward in near-real time. It is based on the integrated forecast system (IFS) version 31r1 (more informations at https://www.ecmwf.int/en/forecasts/documentation-and-support/, last access: November 2018) using approximately an 80 km (T255) spatial resolution and with analyses available for 00:00, 06:00, 12:00 and 18:00 UTC. A detailed explanation of the ERA-Interim product archive is provided in [23,37].

Recently, ERA5 [38] the latest version of ECMWF reanalyses was released as the fifth generation produced. It is envisioned that ERA5 will replace the release of the current ERA-Interim reanalysis, from 1979 to the near real time period (on a regular basis). Regarding climate information, ERA5 has numerous improved characteristics compared to ERA-Interim reanalysis. It presents one of the most updated version of the Earth System Model and data assimilation techniques used at ECMWF which enables the use of more sophisticated parametrization of geophysical processes in comparison to the previous versions used in ERA-Interim. Moreover, ERA5 has two other important features which are the improved temporal sampling and spatial resolution: from 6-hourly in ERA-Interim to hourly in ERA5, and from 79 km in the horizontal dimension and 60 vertical levels to 31 km and 137 levels in ERA5.

At the time of this study, ERA5 is a new product and to the best of our knowledge, only three other studies compared the performance of ERA5 and ERA-Interim. In [4], the authors assessed the two reanalysis ERA5 and ERA-Interim using them to force the ISBA LSM over North America. Better performances in the representation of evaporation, snow depth, soil moisture and river discharge estimates were observed in the simulations forced by ERA5. They were attributed by the authors to
the improved precipitation estimates. Urraca et al. [25] compared $SW_{in}$ estimates from ERA5 and ERA-interim at a global scale, and observed a better performance with ERA5. Finally, Beck et al. [24] highlighted good performance of ERA5 precipitation with respect to 26 gridded subdaily precipitation datasets using Stage-IV gauge-radar data for the evaluation over the continental United States of America.

### 2.2.3 ASCAT soil water index and GEOV2 leaf area index

This study uses the ASCAT Soil Water Index (SWI) product distributed by the CGLS through its third version, i.e. SWI-001 Version3.0. The SWI refers to the soil moisture content in relative units between 0 (dry) and 100 (saturated). It is computed based on a recursive exponential filtering method [39] using the backscatter observations from the ASCAT C-band radar on board MetOP satellites [40,41]. The SWI retrieved from the exponential filter using a T-value (characteristic time length; the higher the T-values, the smoother the SWI) of one day is used. It represents the SWI in the top soil layer [39]. It is used in the present study as a proxy for SSM. During the period considered in the experiment, the amount of soil moisture data increases in 2015 because the data from MetOP-B (launched in 2012) are used in addition to those from MetOP-A (launched in 2006) (see Table 1). Figure 1a presents the map of the average ASCAT SSM estimate for the whole January 2007 – June 2018 period over the study area. For more details on the ASCAT SSM, readers are referred to [18]. Consistently with previous studies, e.g. [42], SSM displays a typical spatial structure which is dominated by a strong meridional gradient (with wetter soil to the South and drier to the North). Some smaller scale patterns also emerge, such as enhanced SSM spots along the Sourou river (13.04 °N, 3.04 °W) and around Niono in Mali (around 14.15 °N, 5.59 °W) in the north eastern part of the domain. These are consistent with existing mapping of water bodies in the region (e.g. [43]) and also probably related to the presence of irrigated rice puddles and crops [44].

In the purpose of assimilating the SSM product, a rescaling of observations into model climatology space is needed in order to avoid introducing any artificial bias in the system caused for example by a possible mis-specification of physiographic parameters related to soil texture types [41,45]. To
that end, the SWI product is transformed into model-equivalent SSM (from the model second layer of soil, 1-4 cm), based on the first two statistical moments (the mean and the variance) through a linear transformation [46]. The relevance of performing a seasonal rescaling was emphasized by several studies (e.g., [26,47]). In this study, the matching of SSM statistical distributions was made on a monthly basis by using a 3-month moving window over the January 2007 – June 2018 period after screening for the presence of urban areas (> 15 %) and complex terrains (1500m a.s.l.). Finally, the SWI observations are interpolated by an arithmetic average to the 0.25° model grid points (from their original 12.5 km spatial resolution).

Table 1: Yearly N points of SWI and LAI from 2007 to 2018-06.

<table>
<thead>
<tr>
<th>Year</th>
<th>SWI</th>
<th>LAI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>36154</td>
<td>37107</td>
</tr>
<tr>
<td>2008</td>
<td>39624</td>
<td>37500</td>
</tr>
<tr>
<td>2009</td>
<td>38761</td>
<td>37342</td>
</tr>
<tr>
<td>2010</td>
<td>41687</td>
<td>37425</td>
</tr>
<tr>
<td>2011</td>
<td>41699</td>
<td>37269</td>
</tr>
<tr>
<td>2012</td>
<td>41652</td>
<td>36706</td>
</tr>
<tr>
<td>2013</td>
<td>41086</td>
<td>37024</td>
</tr>
<tr>
<td>2014</td>
<td>41129</td>
<td>38778</td>
</tr>
<tr>
<td>2015</td>
<td>173818</td>
<td>38839</td>
</tr>
<tr>
<td>2016</td>
<td>319565</td>
<td>38757</td>
</tr>
<tr>
<td>2017</td>
<td>319097</td>
<td>38243</td>
</tr>
<tr>
<td>2018</td>
<td>157476</td>
<td>19908</td>
</tr>
</tbody>
</table>

The GEOV2 LAI observations are also distributed by the CGLS. They are retrieved from the SPOT-VGT and PROBA-V satellite data using the methodology prescribed in [48]. The 1 km x 1 km resolution observations are interpolated to 0.25° model grid points through an arithmetic average as in [12], so that at least 75 % of the grid points are observed. In terms of temporal resolution, LAI observations are available with a 10-day frequency (at best). Figure 1b illustrates the averaged LAI (January 2001- June 2018). As in Figure 1a, the spatial structure of LAI is dominated by a strong meridional gradient (from lower LAI to the North to higher LAI to the South). This correspond to three climatic regions: the Sahel (SH), the Sudan-Sahel (SS) and the Sudan-Guinea (SG). The country’s climate is characterized by two distinct seasons: a dry season and a rainy season (May to October) with growing seasons varying from six (SG region) to three (SH region) months [36].
2.2.4 Evapotranspiration and gross primary production

Independent datasets of evapotranspiration and gross primary production (GPP) are used to assess the quality of LDAS-Monde reanalysis of LSVs.

Terrestrial evapotranspiration estimates are from the GLEAM (Global Land Evaporation Amsterdam Model) v3.1. product [49]. They cover the period 1980-2016 and are available at a spatial resolution of 0.25° x 0.25°. The GLEAM dataset is widely used for investigating both trend and spatial variability in the terrestrial water cycle (e.g. [50–52]) as well as land atmosphere interactions (e.g. [53,54]). In short, the model computes the terrestrial evaporation and root-zone soil moisture [55] and is mainly driven by microwave remote sensing observations, potential evaporation amount being constrained by satellite-derived soil moisture.

For the evaluation of GPP, we use estimates derived from meteorological parameters through the use of machine learning algorithms within the FLUXCOM project [56]. This set of observations can be found at the Max Planck Institute for Biogeo-chemistry data portal (https://www.bgc-jena.mpg.de/geodb/projects/Home.php, last access: November 2018) and is available at a 0.5° x 0.5° spatial resolution with a monthly temporal frequency over the 1982-2013 period. In this study, GPP products were used over the 2001-2013 time period.

2.3 Experimental setup and evaluation strategies

In this study, we first evaluate both precipitation and SW_in variables from ERA-Interim and ERA5 reanalyses against in situ measurements. On the basis of this evaluation (section 3.1.), LDAS-Monde is driven by ERA5 reanalyses, with all atmospheric variables interpolated at a spatial resolution of 0.25° x 0.25°. In order to drive the model to the equilibrium state, the first year (2001) is spun-up 20 times. Then, a comprehensive application of LDAS-Monde is performed using the SEKF as well as its open-loop counterpart (model only without assimilation). It is the same framework as used in [12]. The experiment covers the period 2001 – June 2018.
Performance metrics are used (i) to assess the ability of LDAS-Monde to represent the land surface conditions as well as (ii) to evaluate ERA-Interim and ERA5 reanalyses. Metrics such as the correlation coefficient (R), mean bias, standard deviation of differences (SDD) and root mean squared differences (RMSD) are applied to LDAS-Monde reanalyses and to the open-loop. R and ubRMSD (unbiased RMSD) metrics were applied to precipitation and SW$_{in}$ reanalyses for the evaluation. A 10000 samples bootstrapping is used to determine the 95 % confidence interval of the median from the precipitation reanalyses (see Table 2). For the evaluation of both precipitation and SW$_{in}$ from ERA5 and ERA-Interim reanalyses, the 2010-2016 period and the year 2017 are considered, respectively.

3. Results

3.1 Evaluation of ERA5 and ERA-Interim reanalyses

The performance of ERA5 and ERA-Interim precipitation and SW$_{in}$ is assessed by comparing reanalyses with the in situ measurements described earlier.

The statistical scores for 2010-2016 daily precipitation from ERA5 and ERA-Interim with respect to 134 gauge stations spanning all over BF are shown in Table 2. Median R (resp. ubRMSD) value for total monthly precipitation time series along with their 95 % confidence interval is 0.82 ± 0.009 (52.02 ± 1.39 mm/month) for ERA5, and 0.77 ± 0.010 (58.44 ± 1.42 mm/month) for ERA-Interim. These results demonstrate the ability of ERA5 reanalysis to better represent precipitation variability than ERA-Interim. ERA5 performs better than ERA-Interim for 84 % of the precipitation gauging stations for R values and 89 % for ubRMSD values. This is also illustrated by maps in Figure 2 where triangle (circle) symbols indicate stations where ERA5 performs better (worse) than ERA-Interim in terms of R (Figure 2a) and ubRMSD (Figure 2b). Overall, triangle symbols dominate the two maps of Figure 2, which implies that ERA5 precipitation reanalyses are in better agreement with in situ observations than ERA-Interim over BF. Precipitation from ERA-interim or ERA5 are
both closer to in situ observations in the northern (lower ubRMSD values) than in the southern part of the domain. Overall, ERA5 performs better than ERA-Interim likely due to an improved representation of convective precipitation in the tropical region [57] and to the larger number of assimilated data; it is also possibly related to its higher spatial resolution.

**Table2**: Comparison of precipitation forcing with in situ observations for ERA5 and ERA-Interim over the period 2010–2016 (based on monthly sum). Scores are given for significant correlations with p-values < 0.05.

<table>
<thead>
<tr>
<th></th>
<th>Median R(^1) on precipitation time series, 95 % confidence interval(^2)</th>
<th>Median ubRMSD(^1) on precipitation time series (in mm/month), 95 % confidence interval(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERA5</td>
<td>0.82 ± 0.009 (84 %)</td>
<td>52.02 ± 1.39 (89 %)</td>
</tr>
<tr>
<td>ERA-Interim</td>
<td>0.77 ± 0.010 (16 %)</td>
<td>58.44 ± 1.42 (11 %)</td>
</tr>
</tbody>
</table>

\(^1\)Only for stations presenting significant R values on precipitation time series (p-value < 0.05): 134 stations;

\(^2\)95 % confidence interval of the median derived from a 10000 samples bootstrapping.
Figure 2: Maps of correlation (R) on precipitation time series (a) and ubRMSD (mm/month) on precipitation time series (b) between in situ measurements and both ERA-Interim and ERA5. For each station presenting significant R (p values < 0.05), the simulation that presents the better R values is represented. Triangle symbols indicate when ERA5 presents the best value and circles when it is ERA-Interim.

Figure 3 shows maps of R and ubRMSD values between ERA5 daily mean surface SW$_{in}$ and in situ measurements at 4 stations (Figures 3a and b) as well as their differences against R and ubRMSD values from ERA-Interim (Figures 3c and 3d) over 2017. From Figure 3, one can clearly observe higher correlations and lower ubRMSD values for ERA5 compared to ERA-Interim for the considered four stations. This is consistent with the observed positive correlation differences (ERA5 – ERA-Interim, Figure 3c) and negative ubRMSD differences (ERA5 – ERA-Interim, Figure 3d).

Figure 4 shows the 2017 time series of the daily SW$_{in}$ for ERA-Interim (blue), ERA5 (green) and the in situ observations (red) for Bobo (Figure 4a; 11.16 °N, 4.30 °W) and Dori (Figure 4b; 14.03 °N, 0.03 °W) stations belonging to SG and SS zones, respectively. At these subtropical sites, the temporal structure of the annual cycle of SW$_{in}$ is strongly shaped by the top-of-the atmosphere incoming radiation, which drives two well defined maxima of SW$_{in}$ [58,59]. At Bobo, they occur in April and October while further North in Dori, the second maximum is less pronounced. Both
reanalyses broadly capture these features, even though they tend to overestimate $SW_{in}$ (in particular during the monsoon when clouds induce sharp drops which can reach more than 100 W/m$^2$) - a similar bias also noted by [60] at the relatively close Sahelian site of Niamey. However, this bias is slightly reduced in ERA5. This implies that ERA5 performs better in representing $SW_{in}$ variations than ERA-Interim over BF. This better performance of ERA5 is also probably related to the implementation of an improved radiation scheme (REF) (see [61] for more details).

Figure 3: Maps of correlation $R$ (a) and ubRMSD (b) between incoming solar radiation time series from ERA5 and in situ measurements. (c) and (d) represent the difference between ERA5 and ERA-Interim in correlation and ubRMSD for 2017 respectively. For each station presenting significant $R$ ($p$ values < 0.05), the simulation that presents the better $R$ values is represented.
3.2 LDAS-Monde impact

3.2.1 Assimilation impact on LAI and soil moisture

In order to obtain the land surface analysis, satellite derived observations are combined to the model simulations through the above-mentioned data assimilation technique. Doing that, the resulting analysis is expected to be closer to the assimilated observations (LAI and SSM) than the open-loop (i.e. model with no assimilation). The largest impact is observed for LAI. Figure 5 shows maps of monthly average values of precipitation from ERA5 as well as LAI from the model, the
observations and the difference between the analysis and the modeled LAI. Observations indicate a sharp jump of LAI in April, very likely in response to the increase of soil moisture availability due to the start of the rainy season. However, an increase in LAI is also observed earlier in the SG region (southern part of the domain) from February to May; i.e. before the first rains. Interestingly, this peculiar behaviour would be consistent with findings from in situ studies from [62–64] which point to some tree species that put on leaves before the first rains of the season. This functioning does not appear to be linked to soil moisture [65]. It could involve rises of air temperature or humidity though the precise mechanisms at play are still unknown. In any case, this process is not represented in the ISBA LSM, leading to a temporal shift of two to three months in the leaf onset and an underestimation of observed LAI.

Figure 6 presents similar maps for the following months, from July to December. While the model underestimates LAI in Spring, it significantly overestimates the LAI in November, after the demise of the rainy season.

The analysis is efficient to compensate for these two model caveats, as shown in Figure 5 and 6. This is also clear in Figure 7 which shows LAI monthly mean time series averaged values over the whole domain for January 2001 to June 2018. Observations indicate interannual fluctuations in the yearly maximum of LAI, with for instance higher values in 2003, 2010 and 2012 and lower values in 2002, 2011 and 2017 (consistently with [66] and [67]). Beyond differences in the structure of their annual cycles, the model and the analysis both capture part of this interannual variability.
Figure 5: Seasonal average (from January to June) maps of precipitation \([\text{kg.m}^{-2}.\text{d}^{-1}]\) and LAI \([\text{m}^2\text{m}^{-2}]\) over January 2001 to June 2018. From left to right: precipitation, LAI model, LAI analysis, LAI analysis-model difference. The latter column shows the impact of assimilating LAI and SWI on the simulated LAI.
Figure 6:

Same as Figure 5, but from July to December
Figure 7: Monthly average values of LAI from 1\textsuperscript{st} January 2001 to June 2018: model (blue line), satellite product (green circles), analysis (red line).

The improved annual cycle of LAI provided by the analysis is associated with a higher correlation and a positive impact on the SDD, Bias and RMSD scores over the study region at all months (Figure 8).
Considering now the three climatic regions of BF, it appears that the correlation between the analyzed and observed LAI is lower during the rainy season in the SG region while it is higher in the SS and SH regions (Figure 9b). This reduction of R values in the analysis during the rainy season in the SG region could be related to a decrease in the number of observed LAI linked to cloud cover (see Table 3). Overall, the assimilation corrects the model seasonal responses by increasing LAI during the rainy season and decreasing LAI after the rainy season especially in the sudanian region (SS and SG regions).

Figure 8: Seasonal LAI scores of the model (blue) and analysis (red) when compared to the observations for January 2001 to June 2018.
Figure 9: Seasonal RMSD and correlation of LAI analysis when compared to the observations considering the three climatic regions (SH, SS and SG) over January 2001 to June 2018.

Table 3: Seasonal N values of LAI from 2001-01-01 to 2018-06-30.

<table>
<thead>
<tr>
<th></th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF</td>
<td>59724</td>
<td>59723</td>
<td>59724</td>
<td>59715</td>
<td>58914</td>
<td>53840</td>
<td>48140</td>
<td>48006</td>
<td>46022</td>
<td>54265</td>
<td>56406</td>
<td>56406</td>
</tr>
<tr>
<td>SH</td>
<td>18684</td>
<td>18683</td>
<td>18684</td>
<td>18679</td>
<td>18674</td>
<td>18678</td>
<td>17638</td>
<td>17644</td>
<td>17646</td>
<td>17646</td>
<td>17646</td>
<td>17646</td>
</tr>
<tr>
<td>SS</td>
<td>19440</td>
<td>19440</td>
<td>19440</td>
<td>19440</td>
<td>19440</td>
<td>19440</td>
<td>18334</td>
<td>18357</td>
<td>18319</td>
<td>18360</td>
<td>18360</td>
<td>18360</td>
</tr>
<tr>
<td>SD</td>
<td>21600</td>
<td>21600</td>
<td>21600</td>
<td>21596</td>
<td>20800</td>
<td>15722</td>
<td>12168</td>
<td>12005</td>
<td>10057</td>
<td>18259</td>
<td>20400</td>
<td>20400</td>
</tr>
</tbody>
</table>

Analysis increments for LAI and for soil moisture in (i) the second layer of soil ($w_{g_2}$, between 1 and 4 cm), (ii) the fourth layer ($w_{g_4}$, between 10 and 20 cm) and (iii) the sixth layer ($w_{g_6}$, between 40 and 60 cm) are averaged for the period January 2001 to June 2018. They are presented in Figure 10. Overall, the analysis removes LAI in the central part (SS) of the domain. However, small spots of increases are noticeable in both the south-western (SG) and north-eastern part (SS) of the domain. Regarding the $w_{g_2}$ soil moisture values, the analysis adds water in some parts of the domain (SH, SG and the eastern part of SS) and removes water in the central part. For $w_{g_4}$ and $w_{g_6}$ i.e. for deeper soil layers, the impact is less pronounced but a perceptible north-eastward oriented drying is still observed. It is worth mentioning that the dark blue point in the north-western part of Figure 10a (positive LAI increments) around Mopti in Mali corresponds to the Niger inner
delta including wetlands and irrigated areas. Irrigation is not accounted for in the version of ISBA used in this study and the analysis tends to compensate for this model weakness. This area is also visible in the map of observed LAI (Figure 1b).

**Figure 10:** Averaged analysis increments for January 2001 to June 2018. Four control variables are illustrated: (a) leaf area index and soil moisture in (b) the second (WG2, 1–4 cm), (c) fourth (WG4, 10–20 cm) and (d) sixth (WG6, 40–60 cm) layer of soil.

As mentioned above, the impact of assimilation on SSM is relatively weaker than on LAI. This can be seen for instance in Figure 11a which shows that the annual cycle of SSM in both the model and the analysis are well correlated with ASCAT SSM estimates (R always higher than 0.88). Figure 11a also indicate a positive impact of the assimilation especially from January to May and from October
to December with regard to the model, emphasizing the improvement in SWI values. Removing the seasonality effect through anomaly time series allows to assess the shorter-term variability of soil moisture. Better results are again observed with the assimilation. The impact of the assimilation on the representation of soil moisture shorter-term variability is further presented using maps of anomaly correlations between soil moisture before and after assimilation against ASCAT SSM estimates (Figure 11c and d); as well as using their correlation differences (Figure 11e). It is evident from Figure 11c and Figure 11d that the analysis improves soil moisture simulations over the whole domain with a more pronounced impact over the Sudanian region i.e. SH and SG (Figure 11d). Looking at correlation differences based on anomaly time series, positive values dominate especially in the SH and SS regions along with a rather neutral effect in the southern part of the domain.

Figure 11: Seasonal correlations for (a) volumetric time series and (b) anomaly time series between SSM estimates from the CGLS project and SSM from the second layer of soil of ISBA LSM (in blue) and the assimilation (in red) over the period 2007–06/2018. Maps of anomaly correlations...
between (c) soil moisture from the ISBA LSM and the SSM estimates and (d) soil moisture after assimilation and the SSM estimates. The map (e) expresses the difference between (d) and (c).

3.2.2 Evaluation using independent datasets

For evapotranspiration and GPP, seasonal scores (RMSD and R values in Figure 12) are computed for the model and for the analysis. For evapotranspiration, only vegetated grid points (> 90 %) are considered. After the joint assimilation of SSM and LAI, a small positive impact on evapotranspiration is observed, as found in [29]. However, there is a clear improvement in GPP in terms of RMSD and R scores especially from January to August (Figure 12c and Figure 12d). While the impact on R values of the analysis is equally distributed over the domain for both evapotranspiration and GPP, a larger (positive) impact on RMSD values is found on the southern part of the domain (not shown). This is in agreement with the analysis impact on LAI described above.
5. Discussion and conclusion

This study focused on the capability of LDAS-Monde to provide an improved reanalysis of the land surface conditions through the joint assimilation of satellite-derived soil moisture and vegetation products over BF in western Africa. The LDAS-Monde offline system was forced by the ERA5 atmospheric reanalysis, leading to a quarter degree spatial resolution reanalysis. The quality of ERA5 precipitation and SW$_{in}$ with respect to the former ERA-Interim reanalysis was evaluated using in situ measurements. The comparison of the use of the two atmospheric forcings yields to two key results: ERA5 provides substantial improvements compared to ERA-interim for precipitation (over 2010-2016), and also for the SW$_{in}$ variable (over 2017). Using ERA5 to force the
ISBA LSM (open-loop) provides a good first guess, e.g., on the LAI variable. However, comparing the open-loop with observed LAI has permitted to highlight missing processes in the representation of vegetation phenology. From the analysis, important improvements in the representation of LAI, SWI and GPP variables were obtained with better scores for the analysis than for the model equivalent (open-loop simulation). In particular the LAI analysis is very good in compensating for caveats such as the model’s failure in capturing leaf onset prior to the first rains for particular plant species.

The LAI representation could be enhanced even more using more efficient observation and assimilation systems. For example, one single LAI observation was used within a grid cell for all types of vegetation to perform the assimilation, making the Kalman gain and the increment disaggregation relying on the vegetation type as in [26]. This procedure could be improved by performing the assimilation using disaggregated values of LAI for each individual vegetation type. Such a LAI disaggregation method was recently proposed by [68]. Munier et al. [68] produced disaggregated global LAI maps available every 10 days for each vegetation type available from ECOCLIMAP-II over 1999-2015. Assimilating the disaggregated LAI could impact the analyzed LAI and other vegetation related variables, mainly evapotranspiration and GPP as emphasized in [68]. Also, LDAS-Monde analyses for SSM variable could be improved by implementing an observation operator for ASCAT radar (backscatter coefficients), instead of assimilating a retrieved soil moisture product. It would allow a direct assimilation of level 1 data and the use of all the information contained in these observations [69]. Based on these preliminary results, future works should be focused on the use of ERA5 and on the improvement of LDAS-Monde through the direct assimilation of satellite-based soil moisture and vegetation characteristics from remote sensing. This could lead to possible applications such as monitoring LSVs related to hydrometeorological and agrometeorological variables as well as starting to develop forecasting systems based on analyzed initial conditions. In addition, all ERA5 products include information about uncertainties which are provided for each parameter at 3-hourly intervals and at a horizontal resolution of 62 km by a 10-
member ensemble. Using this ensemble together with the whole available period (1979-present) would provide a long-term ensemble of LSV reanalyses driven by high quality meteorological data. This would help quantifying uncertainties in the description of land surface variables. LDAS-Monde could also be forced by ECMWF Integrated Forecasting System High Resolution operational analysis (~0.1° x 0.1° spatial resolution). Additionally to its higher spatial resolution, it offers the possibility to monitor and forecast LSVs, as shown in [70]. Our results over BF are very promising and pave the way toward large-scale long-term reanalyses of the land surface conditions in Western Africa.

Author Contributions: M.T. and C.A. conceived and designed the experiments; M.T. performed the experiments; all the authors analysed the results, M.T. wrote the paper.

Acknowledgements: We thank the German Federal Ministry of Education and Research (BMBF) for funding the West African Science Service Centre on Climate Change and Adapted Land Use (WASCAL) that funds this work, and the University of Abomey Calavi-Benin that hosts the Graduate Research Program of Climate Change and Water Resources. We are further thankful to the Laboratoire mixte international (LMI-ECLAIR2) and the Agence Nationale de la Météorologie du Burkina-Faso (ANAM). Finally, we are grateful to the VEGEO research group of the Centre National de Recherches Météorologiques (CNRM Météo-France) where this work was mainly conducted and for their various supports.

Conflicts of Interest: The authors declare no conflict of interest.
References


38. Hersbach, H.; Dee, D. ERA5 reanalysis is in production. *ECMWF newsletter* 2016, 147.


44. Tappan, G.G.; Cushing, W.M.; Cotillon, S.E.; Mathis, M.L.; Hutchinson, J.A.; Dalsted, K.J. West Africa land use land cover time series. **2016**.


64. Awessou, B.K.; Peugeot, C.; Agbossou, E.K.; Seghieri, J. *Consommation en eau d’une espèce agroforestière en zone soudanienne*; Quae, 2017;


