Response to Reviewer 4 are structured as follow: (1) 4.X: comments from Reviewer 4, (2) Response to 4.X: author's response and author's changes in manuscript when any. For sake of clarity, line and page numbering from the first submission is used.

[...]I think the paper can be an important contribution and can eventually be suitable for publication in HESS, but at this point I recommend MAJOR revisions with consideration of the comments below. […]

Dear Reviewer#4 many thanks for reviewing the manuscript and for highlighting its relevance and interest. Your comments and suggestions led to an improved version of the manuscript. Below is a point by point answer to your specific comments, all your editorial and technical comments were accounted for in the revised version of the manuscript.

Major

4.1 [1) Six of the thirteen figures that show results (i.e., not counting "data & methods" figures 1 and 2) are about evaluating the skill of the assimilation estimates *exclusively* against the assimilated observations (Figs 3 and 9-13). Comparisons against the assimilated observations are also included in Figs 4-6 (along with other variables) and Figs 14-15 (along with forecast estimates of SSM and LAI). While I agree that it is important to verify that the assimilation system works as intended, the authors overemphasize the comparison against the assimilated observations.]

Response to 4.1

We agree with Reviewer #4, verifying that the assimilation system works as intended is an important task. Part of the figures mentioned are indeed dedicated entirely (Fig. 3) or partially (Figs. 4-6) to that validation. The other aforementioned figures play a different role. Fig. 9 allows us to identify potential hotspots for droughts and heat waves. Figs. 10-11 study the behaviour of LDAS-ERA5 in the context of droughts for the WEUR (Western Europe) and the MUDA (Murray-Darling) areas. Figs. 12-15 focus on the capacity of our system to forecast the evolution of land surface variables depending on how it is initialized.

Comment 4.6 on using SSM and LAI as an independent source of information to evaluate the forecast has been further discussed and added in the revised version of the manuscript. While LAI remains an independent source of information (although constrained by the assimilation as explained in Rewiewer#4 4.6), ASCAT SWI has been rescaled to match the model climatology. The seasonal rescaling impacts both bias and correlation. In an attempt to have a more independent evaluation an additional figure has been put in the revised version of the manuscript. It presents maps of correlations, between soil moisture (1-4 cm) from the four experiments (LDAS-HRES openloop, analysis, LDAS_fc4 and LDAS_fc8) and ASCAT SWI (i.e. ASCAT data prior rescaling) for the WEUR domain. Correlations are applied to both absolute values and to anomalies (to assess the short term variability of soil moisture).

End of section 3.2.2

P.22, Lines 703-724: "Similarly to Figures 13(a, b, c, d) panels of Figure 15 illustrates the impact of the analysis on SSM using correlations. This time, ASCAT SWI (i.e. no rescaling) has been used. Figure 15 (top panels) shows map of R values based on absolute values while Figure 15 (bottom panels) shows R values on anomalies (short term variability) as defined in Albergel et al. (2018a). Figure 15 (a) and (e) represents R values and anomaly R values for LDAS_HRES, respectively. As expected R values are higher than anomaly R values. Maps of differences (panels b and f) of Figure 15 suggest that after assimilation, both scores are improved rather equally. While the 4 day and 8-day forecast still show an improvement from the initial condition on R values (panels c and d of

Figure 15 dominated by positive differences, analysis-openloop), maps of anomaly R values forecast do not display any negative or positive impact (panels g and h of Figure 15).”

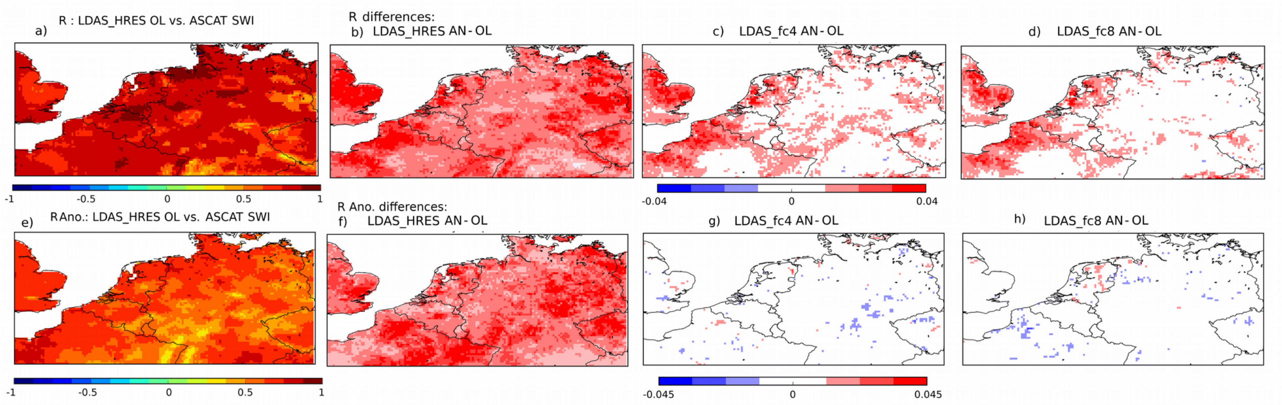


Figure 14: Top row, (a) R values between LDAS_HRES open-loop and ASCAT SWI estimates from the Copernicus Global Land Service (CGLS) over 2017-2018 for the WEUR domain, (b) R differences between LDAS_HRES analysis (open-loop) and ASCAT SWI. (c) and (d) same as (b) between LDAS_fc4 initialised by the analysis (open-loop) and LDAS_fc8. Bottom row, same as top row for R values based on anomaly time-series.

Discussion and conclusion sections
P.23, Lines 749-754: “For SSM, the assimilation is done after a rescaling to the model climatology (see section 2.3), which removes bias. For LAI, however, this is not the case and the assimilation process removes bias in the modelled LAI (w.r.t. the observation). This technical difference between SSM and LAI assimilation, combined with the longer memory of LAI compared to SSM, contributes to the results presented in this section”

4.2 [2) The two figures about snow (Figs 7 & 8) could be simplified considerably because there is no meaningful difference between the assimilation estimates and the open-loop estimates, which is a rather trivial result (as the authors discuss).]

Response to 4.2

Agreed, both figures have been moved to the supplementary document (Figures S1 and S2) and it has been further emphasized that there is no snow data assimilation yet. Those results are presented to highlight areas of improvements in LDAS-Monde:
P.15, Lines 487-492: “As expected, the analysis has an almost neutral impact on snow as both SSM and LAI observations are filtered out from frozen/snow condition and as there is no snow data assimilation in LDAS_ERA5 (Figure S2 and panels (j), (k) and (l) of Figure S1). This clearly shows, however, an area of potential improvement of data assimilation within LDAS-Monde using satellite data such as the IMS one (as in e.g. de Rosnay et al., 2014).”

4.3 [3) There are no graphics in the main text (only in the supplement) about the validation of the results against *independent* in situ measurements (section 3.1.2). This independent validation should be reflected more prominently in the main paper.]

Response to 4.3

Most of the in situ evaluation datasets involved in this study are available over North America and (western) Europe and two regional-scale studies assessing LDAS-Monde analysis impact have already been published (Albergel et al., 2017 over Europe, Albergel et al., 2018a over North America). To avoid redundancy with these previous studies, we preferred not to put too much of those results in the main part of the study. However to better reflect the findings of this evaluation, last paragraph of section 3.1.2 on ground based dataset has been modified and is now (P.18, L.583-587): "For evapotranspiration, river discharge and surface soil moisture it can be stated that there is a slight advantage from LDAS_ERA5 analysis with respect to its open-loop counterpart. Even if the distribution of the averaged statistical metrics can be rather similar for both (particularly true for surface soil moisture evaluation), there are regional significant differences for some sites, which shows the added value of the analysis with respect to the open-loop. Note that for fewer sites, a negative impact from the analysis can also be observed."

Also, the whole evaluation against in situ measurements has been revisited and now includes such a figure, see response to comment 4.4.

4.4 [4) The claim about "improvement" of the assimilation estimates vs. the open-loop estimates from the independent validation against in situ soil moisture estimates in section 3.1.2 (~line 460) is on shaky footing. For none of the networks listed in Table S3 is there a difference of more then 0.02 in the R values between the assimilation and the open loop. In some cases, the 0.02 difference is negative (ie., degradation). For most networks the R difference is 0 or 0.01, that is, there really isn't a meaningful change. Here, and also for at least the other in-situ based results, it is imperative that the authors provide some estimates of whether the differences are meaningful (e.g., by including statistical confidence intervals), and then honestly discuss the results. The claim in line 460 about significant improvements at some sites may be true, but given the network-average neutral results there must then also be sites with a significant degradation, which is not mentioned in the paper.]

Response to 4.4

Thank your for your highly relevant comment. Following it and similar comments from the other Reviewers, it has been decided to revisit the soil moisture evaluation part of the study:
(1) we have added an evaluation of soil moisture from LDAS-Monde fourth layer of soil (10 to 20 cm) against in situ measurements of soil moisture at 20 cm depth when available (10 networks and 685 stations),
(2) for surface soil moisture (SSM), correlation values (R) were calculated for both absolute and anomaly time-series in order to remove the strong impact from the SSM seasonal cycle on this specific metric,
(3) a 95% Confidence Interval (CI) has been added to R values.
(4) we have added the number of stations for which correlations differences are significant (significant improvement or degradation from the analysis) as well as a map over North America for illustration.

It involves several changes in the revised version of the manuscript, they are listed below.

Methodology section, 2.5 Evaluation datasets and metrics

P.11, Lines 358-365: "In situ measurements of surface soil moisture from 19 networks across 14 countries available from the ISMN are also used to evaluate the performance of the soil moisture analysis. They represent 782 stations with at least 2 years of daily data over 2010-2018. Sensors at 5 cm depth (SSM) are compared with soil moisture from LDAS_ERA5 third layer of soil (4-10 cm), sensors at 20 cm depth with the fourth layer of soil (10-20 cm, 685 stations from 10 networks). Beside 11 stations located in 4 countries of Western Africa (Benin, Mali, Sénégal and Niger) and 21 stations in Australia, most stations are located in North America and Europe, see Table S3."

P.12, Lines 374-377: "For global estimates, Normalized RMSD (NRMSD, Eq.(2)) was used, also. Finally, for surface soil moisture, 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., 2018a, 2018b)."

Result section, 3..1.2 Ground-based datasets

P.17-18, Lines 548-582: "The statistical scores for soil moisture from LDAS_ERA5 open-loop and analysis (third and fourth layers of soil, 4-10 cm depth, 10-20 cm depth, respectively) over 2010-2018 when compared with ground measurements from the ISMN (5 cm depth and 20 cm depth) are presented in Table S2 for each individual network. Averaged statistical metrics (ubRMSD, R,$R_{anomaly}$ and bias) are similar for both LDAS_ERA5 analysis and open-loop even if local differences exist. For the analysis, averaged R ($R_{anomaly}$) values along with its 95% Confidence Interval (CI) using in situ measurements at 5 cm (782 stations from 19 networks) are 0.68±0.03 (0.53±0.04) (0.67±0.03(0.53±0.04) for the open-loop) with averaged-network values going up to 0.88±0.01 (0.58±0.04) for the analysis (SOILSCAPE network, 49 stations in the USA) and always higher than 0.55 except for one network, ARM (10 stations in the USA) presenting an averaged R value of 0.29±0.05. Averaged ubRMSD and bias (LDAS_ERA5 minus in situ) are 0.060 m^3m^{-3} and 0.077 m^3m^{-3} for the analysis, 0.060 m^3m^{-3} and 0.076 m^3m^{-3} for the open-loop, respectively. NIC (Eq.1) has also been applied to R values, 65% of the pool of stations present a neutral impact from the analysis (511 stations at NIC ranging between -3 and +3), 12% present a negative impact (91 stations at NIC < -3) and 23% present a positive impact at (180 stations at NIC > +3).
The number of stations where R differences between the analysis and the openloop are significant (i.e. their 95% CI are not overlapping) is 186 out of 782 (about 26%). There is an improvement from the analysis w.r.t. the openloop for 128 stations (out of 186, i.e. about 69%) and a degradation for 58 stations (about 31%). Figure 7 illustrates R differences between the analysis and the openloop runs. When differences (analysis minus openloop) are not significant stations are represented by a small dot. When they are significant, large circles have been used, blue for positive differences (an improvement from the analsysis) and red for negative differences (a degradation from the analysis). For most of the stations where a significant difference is obtained, it represent an improvement from the analysis.
Averaged analysis R (95%CI), bias and ubRMSD for the fourth layer of soil (685 stations from 10 networks) are 0.65±0.03, 0.049 m^3m^{-3} and 0.055 m^3m^{-3}, respectively. For the open-loop, they are 0.064±0.03, 0.048 m^3m^{-3} and 0.056 m^3m^{-3}, respectively. For soil moisture at that depth, about 60% of the stations present a neutral impact from the analysis (410 stations at NIC ranging between -3 and +3), 28% a positive impact (189 stations at NIC > +3) and 12% a negative impact (86 stations at NIC < -3). Although differences between the openloop run and the analysis are rather small, these results underline the added value of the analysis with respect to the model run. Figure S6 represents the distribution of the scores values for LDAS_ERA5 open-loop and analysis using boxplots centred on the median value. They look very similar and from this figure, it is difficult to see either improvement or degradation from the analysis."

Figure 7: Map of correlations (R) differences (analysis minus openloop) for stations available over North America. Small dots represent stations where R differences are not significant (i.e. 95% confidence intervals are overlapping), large circles where differences are significant.

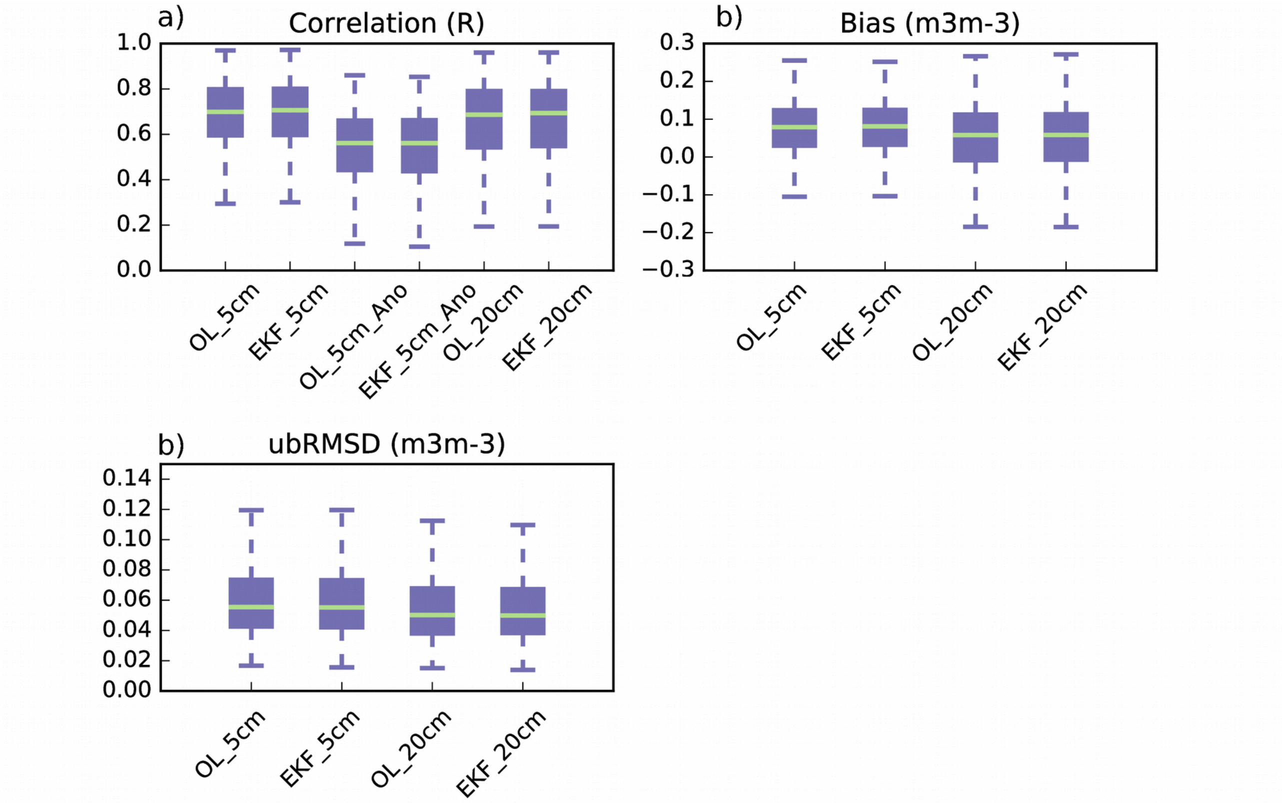


Figure S6: a) Boxplots representing the distribution of the correlation values on absolute time-series and anomaly time-series ("Ano") between the stations with in situ measurements of soil moisture either 5cm depth or 20 cm depth and soil moisture from LDAS_ERA5 openloop and analysis over 2010-2018 (third and forth layer of soil, respectively). Correlation values are presented for surface soil moisture (5 cm depth measurements against third layer of soil), only. Distribution are centred on the median values. b) Distribution of the Bias values between the stations with in situ measurements of soil moisture either 5cm depth or 20 cm depth and soil moisture from LDAS_ERA5 openloop and analysis over 2010-2018 (third and forth layer of soil, respectively).c) Same as b) for ubRMSD.

The following text has been added to the revised version of the manuscript: "Figure S6 represents the distribution of the scores values for LDAS_ERA5 open-loop and analysis using boxplots centred on the median value. They look very similar and from this figure, it is difficult to see either improvement or degradation from the analysis."

4.5 [5) The editing of the paper is rather careless. There are many small mistakes, and the organization of the text is lacking.]

The 4 Reviewers have provided many editorial comments, corrected several mistakes. Thanks to their work we have an improved version of the manuscript.

4.5a [a) The Introduction lacks a clear statement of the paper's objectives. The text in Lines 107-121 simply states what will be presented (with lots of references and details). It's hard to tell what the objectives might be.]

Response to 4.5a

Agreed. In order to make the paper's objectives clearer, the following paragraph in the introduction has been revisited:

"In this study, stemming from previous works referenced above, this global, offline, joint integration of Surface Soil Moisture (SSM) and Leaf Area Index (LAI) EOs into the ISBA (Interaction between Soil Biosphere and Atmosphere) LSM (Noilhan and Planton, 1989, Noilhan and Mahfouf, 1996) are presented: [...]"
is now (P.4, Lines 108-114):
"In this study, stemming from previous works referenced above, it is shown that LDAS-Monde global, offline, joint integration of Surface Soil Moisture (SSM) and Leaf Area Index (LAI) EOs into the ISBA (Interaction between Soil Biosphere and Atmosphere) LSM (Noilhan and Planton, 1989, Noilhan and Mahfouf, 1996) can be used to detect, monitor and forecast the impact on extreme events on LSVs. Are presented in this study: [...]"

4.5b [b) There are several instances in the Results section of text that belongs in the Methods section, incl: Lines 384-387 - IMS snow cover product description Lines 405-409 -Fluxnet description Lines 440-447 - ISMN description]

Response to 4.5b

Agreed. When appropriate, those instances were moved to the section dedicated to methodology (description of IMS data; ISMN and FLUXNET-2015 networks, river discharge).

Response to 4.5c [c) Section 3.2.2 is a *single* paragraph that stretches over nearly two pages. Really? There are several other paragraphs of excessive length.]

Response to 4.5c

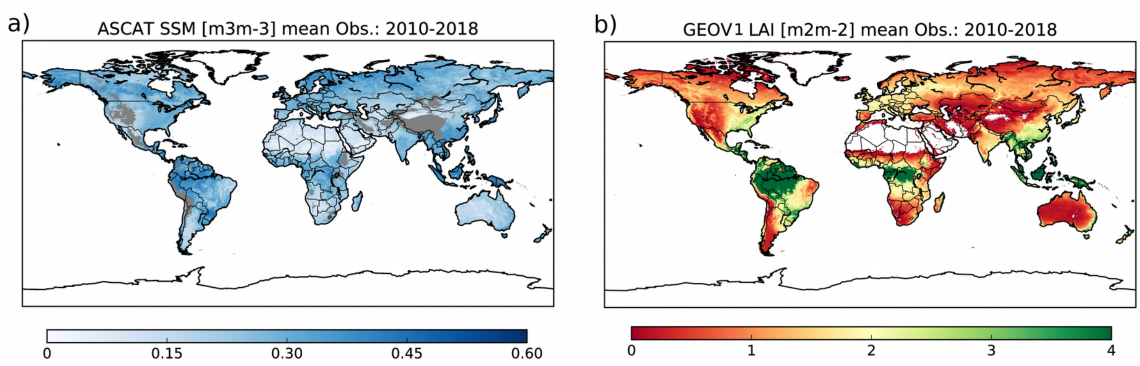
Section 3.2.2 has now been reshuffled with one paragraph per group of 2 figures.

d) Graphics:

4.5d_f1 [Figure 1a: Use different color for zero values and no-data value. (currently, both are white, making it unclear whether there are data in, e.g., the western US, or whether those are screened, perhaps because of topography.]

Response to 4.5d_f1

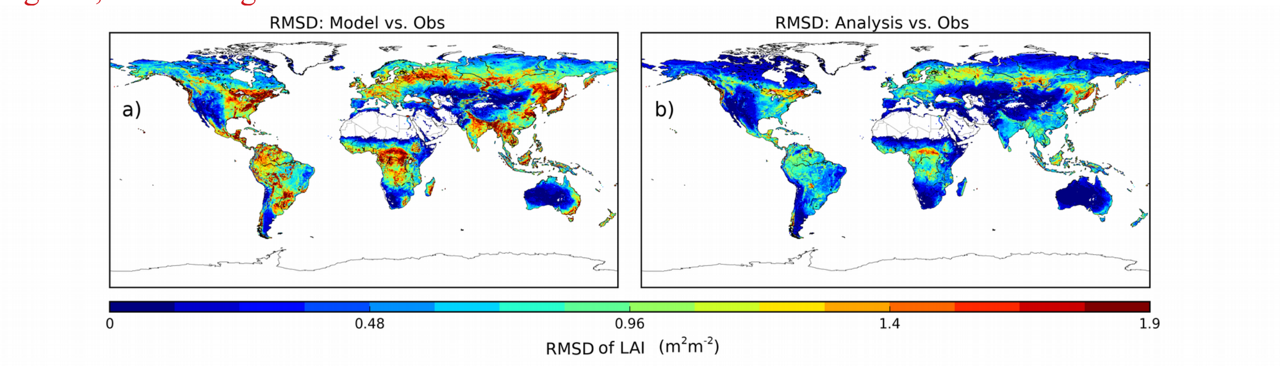
Agreed, see new figure below.



4.5d_f2 [Figure 3: The label of the colorbar should read "RMSD of LAI [m2 m-2]", not just "LAI [m2 m-2]"]

Response to 4.5f_f2

Agreed, see new figure below



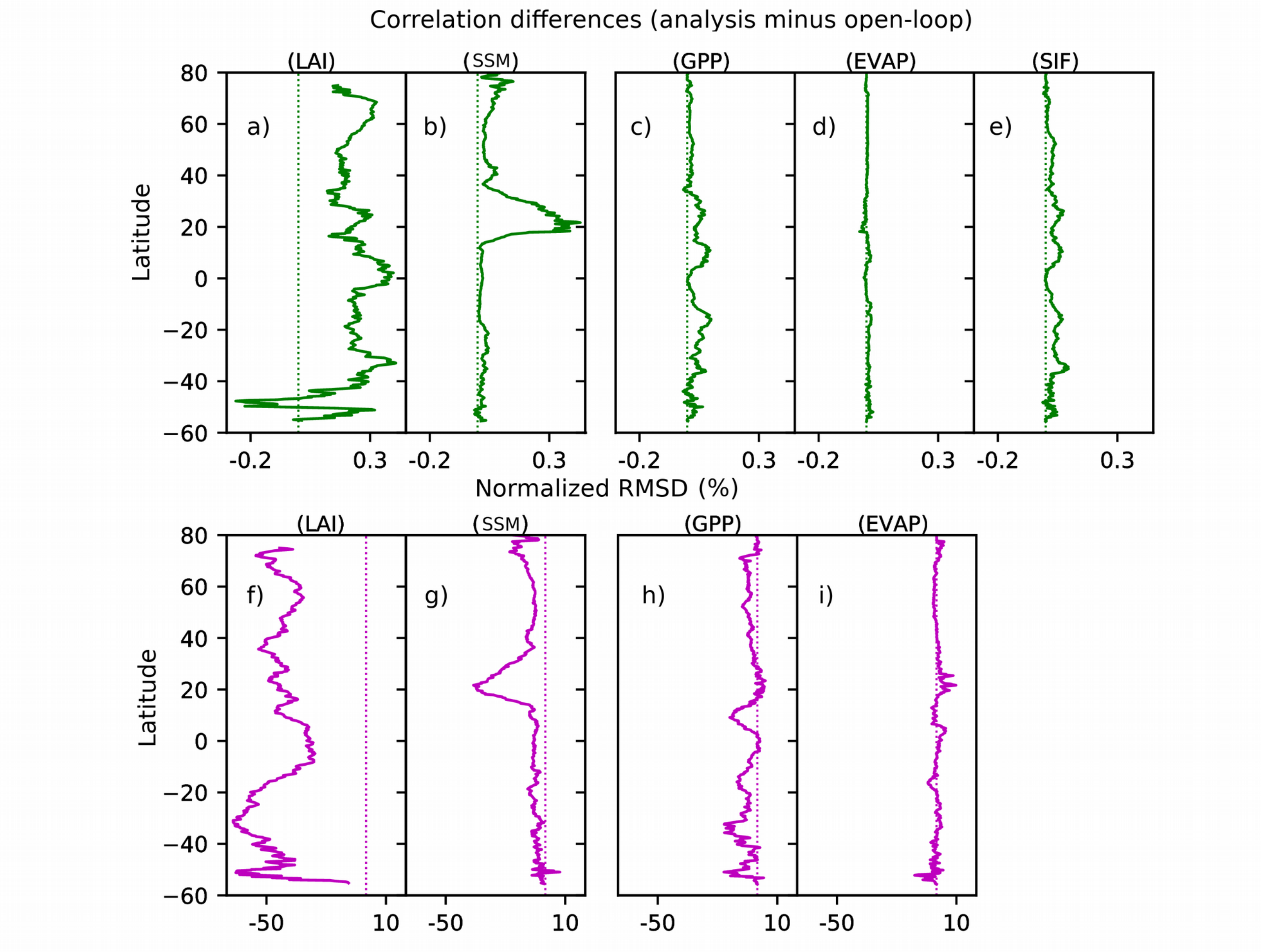
4.5d_f5 [Figure 5: Units are missing for RMSD panels. (This is particularly important because this information is needed to judge whether the differences are in fact meaningful.)]

Response to 4.5d_f5

Thank you for this suggestion, for RMSD panels it has been decided to use normalized RMSD (% of improvement and/or degradation) so one can really see the impact on each evaluated variable, it also echoes Reviewer 4's comment 4.7 on analysis impact on GPP. Using similar x-axis limits provides a better information at a glance. For instance it minimizes the previous visual impact of the analysis on GPP, and as such addressing your comment 4.7. Also panels of new Figure 5 separate the assimilated and independent variables, see new figure below.

Also, in section 3.1.1 on gridded dataset:the following sentence "For SSM a noticeable improvement in both correlation and RMSD is found around 20°N corresponding mainly to an improvement in the Sahara desert (not shown). GPP is also improved across almost all latitude with a particularly positive impact below 20°N which is also true for EVAP. This variable is less impacted by the analysis and some parts of the world show a decrease in e.g. RMSD values." is now (P.14, Lines 436-441):

"For SSM a noticeable improvement in both correlation and RMSD is found around 20°N corresponding mainly to an improvement in the Sahara desert (not shown). Being linked to LAI, GPP is also improved across almost all latitudes (to a lesser extend than LAI) with a particularly positive impact below 20°N. As seen on Figure 5 d) and i), there is little impact on variable EVAP which can be considered negligible. It highlights the difficulty of land surface data assimilation to impact model fluxes by modifying model states."



4.5d_f6 [Figure 6: Three panels only have a single tick & tick label on the y-axis. At least two are required to interpret the axis scale.]

Response to 4.5d_f6

Agreed, it has been added in the revised version of the manuscript.

4.5d_f7 [Figure 7: The color choices should be made consistent with Fig 4.]

Response to 4.5d_f7

Agreed, Figure 7 is now in the supplementary.

4.5d_f9 [Figure 9: I could not find out what the thin cyan lines depict.]

Response to 4.5d_f9

The following sentence has been added to the caption of the considered figure's caption: "Solid red line, dashed red line and solid green line represent regions MUDA, WEUR and EAFR. Solid cyan lines represent all other boxes (see Table 1 and Figure 2)."

4.5d_f10+11 [Figures 10+11: add "LAI" to plot title of c) and d); add "SSM" to plot title of g) and h)]

Response to 4.5d_f10+11

Agreed, it has been added in the revised version of the manuscript.

4.5d_fS2 [Figure S2: NSE should vary from -infinity to 1. The colorbar is from -20 to 20, and darker blue values would clearly be greater than 1. Either the colorbar is wrong or the values show something other than NSE.]

Response to 4.5d_fS2

Thanks for spotting this issue resulting from a wrong call in a python script, it has been corrected in the revised version of the manuscript.

4.5d_ts3 [Table S3: The column headings on the 2nd page of the table still include French words.]

Response to 4.5d_ts3

Corrected, thanks for spotting this issue.

4.6 [6) In section 3.2.2, the authors no longer make it clear that the verification is against the assimilated datasets. While verification of forecast data against the assimilated dataset can be viewed as independent validation because the verification data have not (yet) been assimilated, there is an important distinction here between SSM and LAI. For SSM, the assimilation is done after rescaling (cdf-matching), which removes bias. For LAI, however, the assimilation uses the raw LAI observations (I think). That is, the assimilation removes bias in the modeled LAI (w.r.t. the observed LAI). This technical difference between SSM and LAI assimilation, combined with the longer memory of LAI compared to SSM, should contribute to the results in section 3.1.2. Put differently, the LAI results of section 3.1.2 are not likely to hold if an independent LAI dataset had been used for validation that is itself biased against the assimilated LAI observations. (Different LAI datasets may not be as biased against each other as typical satellite SSM datasets, but there are considerable biases between LAI products.)]

Response to 4.6

Verifying that the assimilation system works as intended is an important task. This is why several figures have been included for “sanity check”. We have emphasized in the manuscript that several presented evaluations are carried out to check if the assimilation system is working properly.

Also, using SSM and LAI as an independent source of information to evaluate the forecast has been further discussed and added in the revised version of the manuscript. While LAI remains an independent source of information for the forecast evaluation (although constrained by the assimilation), ASCAT SWI has been rescaled to match the model climatology. The seasonal rescaling impacts both bias and correlation. In an attempt to have a more independent evaluation, an additional figure has been put in the revised version of the manuscript. It displays maps of correlations between modelled soil moisture (1-4 cm) from the four experiments (LDAS-HRES openloop, analysis, LDAS_fc4 and LDAS_fc8) and ASCAT SWI (i.e. ASCAT data prior rescaling) for the WEUR domain. Correlations are applied to both absolute values and to anomalies (to assess the short term variability of soil moisture).

End of section 3.2.2

P.22, Lines 703-724: "Similarly to Figures 13(a, b, c, d), panels of Figure 15 illustrate the impact of the analysis on SSM using correlations., To that end, ASCAT SWI (i.e. no rescaling) has been used. Figure 14 (top panels) shows map of R values based on absolute values while Figure 14 (bottom panels) shows R values on anomalies (short term variability) as defined in Albergel et al., 2018a. Figure 15 (a) and (e) represents R values and anomaly R values for LDAS_HRES, respectively. As expected R values are higher than anomaly R values. Maps of differences (panels b and f) of Figure 15 suggest that after assimilation, both scores are improved rather equally. While the 4 day and 8-day forecast still show an improvement from the initial condition on R values (panels c and d of Figure 15 dominated by positive differences, analysis minus openloop), maps of anomaly R values forecast don't show any negative or positive impact (panels g and h of Figure 15)."

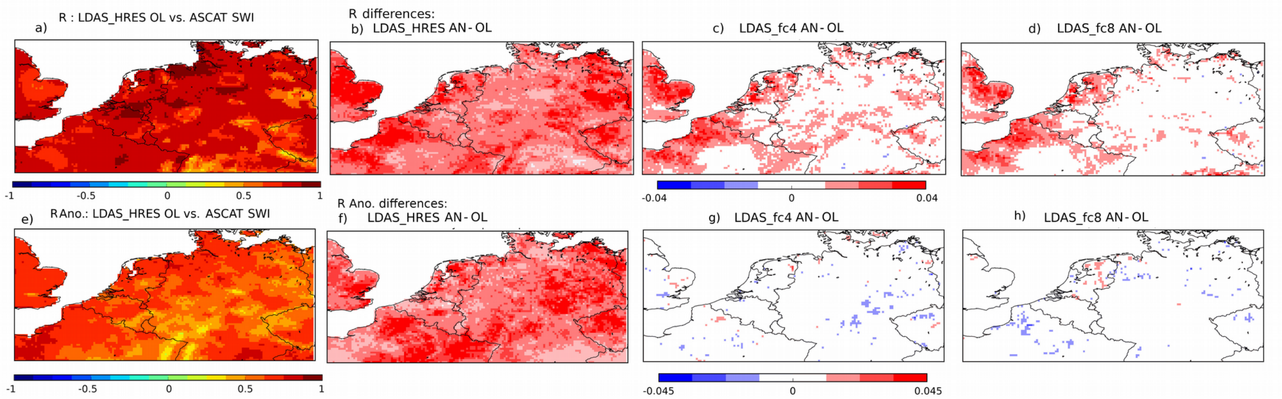


Figure 15: Top row, (a) R values between LDAS_HRES open-loop and ASCAT SWI estimates from the Copernicus Global Land Service (CGLS) over 2017-2018 for the WEUR domain, (b) R differences between LDAS_HRES analysis (open-loop) and ASCAT SWI. (c) and (d) same as (b) between LDAS_fc4 initialised by the analysis (open-loop) and LDAS_fc8. Bottom row, same as top row for R values based on anomaly time-series.

<u>Discussion and conclusion sections</u>

P.23, Lines 749-754: "For SSM, the assimilation is done after a rescaling to the model climatology (see section 2.3), which removes bias. For LAI, however it is not the case and the assimilation process removes bias in the modelled LAI (w.r.t. to the observation). This technical difference between SSM and LAI assimilation, combined with the longer memory of LAI compared to SSM, contributes to the results presented in this section"

4.7 [7) Figure 3c suggests that the change in GPP is negligible, at least in the zonal mean sense although Figure 4f suggests that GPP does change in terms of RMSD. Given the considerable change in the (zonal mean) LAI (Fig 3a), I would have expected a lot more change in the mean GPP. I suspect that the disconnect between the LAI and GPP changes is rooted in how these variables are connected in ISBA and how exactly the assimilation system goes about updating LAI. This rather counter-intuitive result requires clarification in the paper.]

Response to 4.7

We believe that Figure 5 was rather confusing and that the new Figure proposed (see Response 4.5d_f5, also) permits to clarify this point. In section 2.1.1 on ISBA land surface model, the following sentence is now "In the CO_2-responsive versions of ISBA, photosynthesis is in control of the evolution of vegetation variables." <u>is now</u> (P.5, Lines 157-160) "In the CO_2-responsive versions of ISBA, ISBA-A-gs, the model can simulate the CO_2 net assimilation and GPP by considering the functional relationship between the photosynthesis rate (A) and the stomatal aperture (gs) based on the biochemical A-gs model proposed by Jacob et al. (1996). Photosynthesis is in control of the evolution of vegetation variables."

References:

Jacobs, C.M.J.; van den Hurk, B.J.J.M.; de Bruin, H.A.R. Stomatal behaviour and photosynthetic rate of unstressed grapevines in semi-arid conditions. Agric. For. Meteorol. 80, 111–134, 1996.

4.8 [8) Fig 5h: The changes in EVAP are with +/- 0.02 (mm/d???). If my guess about the units is correct, this would amount to only a few mm per year, which is well within the uncertainty of in situ measurements. That is, the EVAP changes are not likely to be meaningful in a practical sense. This should be discussed more explicitly.]

Response to 4.8

Agreed, the new figure 5 also helps to clarify that the impact on variable EVAP is rather negligible. See also Responses to 4.5d_f5, 4.13

Minor

4.9 [9) Line 167: typo "bale" –> "able"]

Response to 4.9

Typo corrected in the revised version of the manuscript, thanks.

4.10 [10) Line 209: "fifth generation of European reanalyses produced by ECMWF" I recommend phrasing this differently to avoid the misunderstanding that the reanalyses are just for the European domain. E.g.,: "fifth generation of global reanalyses produced by ECWMF"]

Response to 4.10

Rephrased in the revised version of the manuscript, thanks.

4.11 [11) Lines 293-295: How did you address the heterogeneity within the 0.25-deg grid cells during spin-up? It is not obvious that the short spin-up period from April 2016 suffices for properly spinning up grid cells with strong heterogeneity at the sub-0.25-degree scale.]

Response to 4.11

The global LDAS-ERA5 runs were spun-up by running 20 times the first year (2010). For LDAS_HRES, nine months can be perceived as a too short period to spin up the system. Unfortunately, HRES atmospheric forcing is only available from April 2016 and the LDAS-HRES experiment ends in December 2018. We have considered this 9 months period for the spin up in order to have the longest possible time series for land surface variables, thus giving more strength to statistics. We could have considered a longer period for spin up (April 2016 to December 2017) and studied only 2018. This gives very similar results on surface soil moisture and LAI (not shown). While not being fully spun-up, results obtained with LDAS-HRES can be considered as representative of the system response to data assimilation. Note that most initial values of the LDAS-HRES run are taken from the ECOCLIMAP-II database. For instance, initial LAI is set from a 1999-2005 MODIS climatology.

Another possibility to initialise LDAS-HRES could have been to downscale the state of LDAS-ERA5 run in April 2016 to 0.10°x0.10° spatial resolution. LDAS-ERA5 runs have been set to an equilibrium spinning up 20 times the first year (2010).

The following sentence: “The period 2017-2018 is presented, HRES is available at this spatial resolution from April 2016, only, and the time period from April to December 2016 is used as a short spinup." has been modified and is now (P.10, Lines 327-332): "HRES is available at a 0.1° x 0.1° resolution only from April 2016. April to December 2016 is used as a short period for spinup and results are presented for the period 2017-2018. Although a 9-month spinup period can be seen as rather short, evaluating LDAS-HRES on either 2017-2018 or 2018 (using instead a 21-month spinup) leads to similar results on surface soil moisture and LAI (not shown). While the system is not fully spun-up, it can be considered as representative of the system response to data assimilation."

4.12 [12) Line 379: Do you mean a decrease in RMSD or a decrease in skill?]

Response to 4.12
This sentence has been revised and "[...] shows a degradation" is now "[...] shows a decrease in skill"

13) Line 412: If I'm reading this correctly RMSD decreases while both bias and ubRMSD increase. This is quite counter-intuitive and requires a rather odd distribution of the metrics across the sites or networks included in the average. In any case, since bias and ubRMSD get worse, I do not think that the statement about "a small advantage of the analysis over the open-loop" is justified.

Response to 4.13
Agreed, the considered sentence has been reformulated and is now: "If these numbers depict a small advantage of the analysis over the open-loop configuration, it is worth mentioning that differences are rather small and likely to fall within the uncertainty of the in situ measurements."

14) Line 429: "NSE values below -2 were discarded" requires a justification, otherwise it reads like cherry-picking.

Response to 4.14
Agreed, this threshold has also been used for previous studies at CNRM as we did not want to look at river discharges we do not represent well. The pool of stations we have used are monitoring all types of rivers and streams including those where human impacts (dams and reservoirs, irrigation, water uptake, not represented in ISBA yet) is affecting the natural flow of rivers. As we expect the impact of the analysis on river discharge to be small (based on previous work), we did not find necessary to include stations we badly represent in ISBA, possibly for known reasons. Futur work will focus on preparing a more robust in situ pool of station, separating e.g. managed and unmanaged rivers and stream.

The following paragraph has been added to the methodology section (P.12-13, Lines 394-399): "Stations with NSE values lesser that -2 were discarded. A similar threshold has already been used in previous studies evaluating LDAS-Monde (e.g. Albergel et al., 2017, 2018a). Many processes, most of them linked to water management such as the presence of dams and reservoirs, irrigation, water uptake in urban areas, are not yet represented in ISBA leading to a poor representation of river discharges. As previous evaluations studies have suggested a neutral to positive impact from the assimilation, only, it has been decided to focus on stations with reasonable NSE values."

4.15 [15) Line 535: "the analysis is of better quality" Given the numbers, I see at best "slightly better quality"]

Response to 4.15
Emphasized in the revised version of the manuscript, "Note however that for the MUDA area, a 4-d forecast of surface soil moisture initialised by the analysis is of better quality than a 4-d forecast

initialised by the open-loop" is now (P.21, Lines 664-666): "Note, however, that, for the MUDA area, there is a small positive impact of the initialisation on the 4-d and 8-d forecast of surface soil moisture (blue areas on Figure 13 c) and d))."

4.16 [4.16 [16) Line 592: "surface (0-1 cm)" In section 3.2.2 the discussion was about the "(1-4cm)" layer. Which is it?]

Response to 4.16
Thanks, it should read 1-4cm, it is now corrected in the revised version of the manuscript