Review Anonymous Referee #1

The paper by Bonan et al., describes the application of an ensemble Kalman filter (EnKF) for the joint assimilation of surface soil moisture and leaf area index over the Euro-Mediterranean region using the LDAS-Monde land surface data assimilation framework. The authors compare the skill of the EnKF with the skill of the well-tested simplified extended Kalman filter (SEKF). This is done by assimilating surface soil moisture and leaf area index, and compare the corresponding effect of the analysis on unobserved variables, such as evapotranspiration (from GLEAM) and gross primary production (from FLUXCOM). In addition, the authors investigate how the ensemble from the EnKF can transfer information between variables by the ensemble covariances and compare this to the SEKF Jacobians.

The paper provides a good starting point for future work on joint assimilation of surface soil moisture and leaf area index. The paper is well organized and it discusses the current difficulties in assimilating leaf area index using the EnKF. It would have been interesting to see if the joint assimilation (using the EnKF) could have improved over the open loop when compared to in situ soil moisture data, however, I leave this for the authors to include in future work. I recommend that the paper is published in HESS after a careful proofread by the authors and after taking care of the following minor comments.

We thank the referee for her/his positive comments about our work and for her/his detailed review that has helped us to improve the quality of our manuscript. Responses to comments and subsequent changes are detailed below in blue. Please note that, following a suggestion from referee #2, section 5.1 has been merged with section 4. Nevertheless, every correction/suggestion from referee #1 for section 5.1 has been issued when remaining applicable.

Minor comments:

Title: "moiture" to "moisture"

Correction done.

L1: Skip "a deterministic ensemble Kalman filter"

Correction done.

L7: Missing "filter"

Correction done.

L7: Unclear sentence, suggested change: ", which has been well studied within the LDAS-Monde framework over the Euro-Med. region, see for example..."

We have replaced the sentence by "..., the EnSRF is compared with the Simplified Extended Kalman Filter, which has been well studied within the LDAS-Monde framework. The comparison is carried out over the Euro-Mediterranean region at a 0.25^o spatial resolution between 2008 and 2017."

L9: "The SEKF"

Correction done.

L11: How is this assessed, there is no validation of the root-zone analysis?

The sentence is misleading. We just meant here to compare the influence of the EnSRF and the SEKF on unobserved control variables, i.e. root-zone soil moisture, which is done in section 4.4. The SEKF has been widely used in LDASs and showing if results obtained with the EnSRF are close or not to SEKF estimates is in itself interesting. As acknowledged by the referee, validating (root-zone) soil moisture with in situ measurements is left for future work. We have replaced the sentence by the following one: "The comparison between the two data assimilation approaches is also carried out on unobserved soil moisture in the other layers of soil."

L14: Please change wording "exhibited" to "is found" or something along those lines.

Correction done.

L15: Is this correlation/anti-correlation seasonally dependent?

The correlation/anti-correlation between LAI and soil moisture is indeed seasonally dependent with their absolute values peaking in summer while being close to zero during winter. The sentence in the abstract has been rephrased to reflect that point:"Moderate correlation and anti-correlations are also noticed between LAI and soil moisture, varying in space and time. Their absolute value, reaching their maximum in summer and their minimum in winter, tends to be larger for soil moisture in root-zone areas ..."

L18-19: Please change wording "and GPP, but a highly positive...". Also what skill metric is considered for the river discharge.

We consider root-mean square difference and correlation as skill metrics for evapotranspiration and GPP and the Nash-Sutcliffe efficiency score for river discharge. The corresponding sentence has been rephrased as follows: "The EnSRF shows a systematic albeit moderate improvement of root-mean square differences and correlations for evapotranspiration and GPP products, but its main improvement is observed on river discharges with a high positive impact on Nash-Sutcliffe efficiency scores. Compared to the EnSRF, the SEKF displays a more contrasting performance."

L23: ".. the earth's water and carbon cycles"

The sentence has been rephrased as: "Land surface variables (LSVs) are key components of the Earth's water, vegetation and carbon cycles."

L26: Please specify what goals?

The sentence containing "these goals" has been rephrased as: "Land surface models (LSMs) play an important role in improving our knowledge of land surface processes and their interactions with the other components of the climate system such as the atmosphere."

L28: "initialization" and please find other wording than "misspecified" forcing.

We have replaced "initialisation" by "initialization" (both being correct in British English) and "misspecified forcing" by "flawed forcing".

L30: "Generally provide sparse spatial coverage..."

Correction done.

L31: "ranging from the km scale to the meter scale"

Correction done.

L32: Please specify why this is the case.

The sentence line 32 is not accurate. We meant that not all key LSVs are observed directly from space. Passive satellite sensors used in the case of soil moisture are sensible only to the near-surface (0–2 cm depth) moisture content. Methodologies such as the exponential filter (Albergel et al., 2008) have been developed to provide estimates of root-zone soil moisture from satellite data but they are indirect approaches. We have rephrased the sentence line 32 has follows:"Not all key LSVs are also observed directly from space. For example, passive microwave satellite sensors used traditionally to estimate soil moisture are sensible only to the near-surface (0–2 cm depth) moisture content (Schmugge, 1983) leading to the development of indirect approaches to estimate root-zone soil moisture from satellite data (see e.g. Albergel et al., 2008)"

References:

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Piguet, B. and Martin, E.: From near-surface to root-zone soil moisture using an exponential filter: An assessment of the method based on in-situ observations and model simulations, Hydrol. Earth Syst. Sci. 12, 1323–1337, 10.5194/hess-12-1323-2008, 2008.

Schmugge, T. J.: Remote Sensing of Soil Moisture: Recent Advances, IEEE T. Geosci. Remote, GE21, 145–146, 10.1109/TGRS.1983.350563, 1983.

L33: "flaws in both approaches"

Correction done.

L35: "passive microwave brightness temperatures, microwave backscatter coeff..." and "obtained from the aforementioned satellite observations".

Correction done.

L39: "systems of both the . . .. and the UK Met office".

Correction done.

L41: "rapidly extended"? Please clarify. It has undergone development from assimilating SSM only to now also include LAI assimilation.

The sentence has been rephrased as: "The SEKF has also been applied to the sole assimilation of soil moisture retrievals (Draper et al., 2009) then to the joint assimilation of soil moisture retrievals and leaf area indices (Albergel et al., 2010; Barbu et al., 2011)."

L42: "the SEKF approach..."

Correction done.

L43-44: Please clarify this sentence, "thus limited their number"?

Indeed the sentence is unclear. It has been rephrased as follows: "It relies on a climatological background error covariance matrix assuming uncorrelated variables between grid points and involves the computation of a Jacobian matrix to build covariances between control variables at the same location. This Jacobian matrix is computed with finite differences, meaning that one model run is required per control variable, thus limiting the size of the control vector."

L45: ".., such as"

Correction done.

L47: In what context?

We meant in the context of assimilating solely surface soil moisture retrievals. This has been added in the manuscript.

L49: "Recently, ..."

Correction done.

L50: Please rephrase, "LAI is a key land biophysical variable, it is defined. . ..."

Correction done.

L51: Please rephrase, "One way to monitor LAI is to assimilate observations indirectly linked to LAI, such as"

Correction done.

L56: Please rephrase: "Another way to constrain LAI is through the assimilation of direct LAI observations."

Correction done.

L56-57: "...products benefit from..."

Correction done.

L58: "and at high-res..."

Correction done.

L58: Please rephrase: "..other studies have assimilated LAI in crop models and at a more local scale.."

Correction done.

L60: "Succeeded in introducing such an approach. . .."

Correction done.

L64: CNRM is already introduced.

Agreed, we have removed the definition of CNRM from the sentence.

L65: "allows for..."

Correction done.

L65: "Building on that work..."

Correction done.

L66: Remove "have"

Correction done.

L67: On a site? Please specify

The site is the SMOSREX site located in South West France. This has been added in the revised version of the manuscript.

L67: "Their study..."

Correction done.

L68: "Leading to the development of LDAS-Monde."

Correction done.

L68-69: "The LDAS-Monde is available through the CNRM modelling platform. . . and it has been successfully..."

Correction done.

L72: "For example, ..."

Correction done.

L72: Drop "has"

Correction done.

L73-74: "while Ling et al (2019) compared..."

Correction done.

L74: Drop "has"

Correction done.

L76: Drop "water and carbon cycles".

Correction done.

L77-78: Rephrase: "These studies did not update both SM and LAI, as we will do in this study".

Correction done.

L79: "in the LDAS-Monde..."

Correction done.

L79-82: Very long sentence, please consider to rephrase.

The sentence has been rewritten for clarity: "To that end, it will build upon the work of Fairbairn et al. (2015), that introduced an Ensemble Square Root Filter (EnSRF, Whitaker and Hamill, 2002) in the LDAS-Monde in the context of assimilating SSM solely. The EnSRF is one of the many deterministic formulations of the EnKF (see e.g. Tippett et al., 2003; Livings et al., 2008; Sakov and Oke, 2008). Fairbairn et al. (2015) compared the performance of the EnSRF with the SEKF, routinely used in the LDAS-Monde, over 12 sites in South-West France ..."

L82: Please make it clear that this is not the current study but the study of Fairbairn et al. (2015).

See response above.

L85: "used"

Correction done.

L85: SMAP acronym not defined.

The acronym SMAP (for Soil Moisture Active Passive) has been defined in the revised version of the manuscript.

L89: "the LDAS-Monde..."

Correction done.

L89: Replace "to" with "on".

Correction done.

L90: "..and its ability to..."

Correction done.

L90: "To achieve these..."

Correction done.

L105: "such as"

Correction done.

L105: "or for evaluation"? Please rephrase.

The sentence has been rephrased as follows: "... such as atmospheric forcing or assimilated observations. Sect. 3 also details the datasets used to assess the performance of the EnSRF and the SEKF. ..."

L106: "Finally,..."

Correction done.

L107: ". . ..prospects for future work."

Correction done.

L110: "by the Meteo-France research centre CNRM".

Correction done.

L110-L113: Very long sentence, please rephrase.

The sentence has been rephrased as follows: "Embedded within the open-access SURFEX surface modelling platform (Masson et al., 2013, https://www.umr-cnrm.fr/surfex/), LDAS-Monde involves the ISBA land surface model coupled with the CTRIP river routing system and data assimilation routines. Those routines assimilate routinely satellite-based products of SSM and LAI to analyse and update soil moisture and LAI modelled by ISBA."

L114: Replace "conduct" with "in"

Correction done.

L119-L121: Please consider to rephrase to make this sentence easier to read. For example: "In this paper we use the ISBA multilayer diffusion scheme (ref) which solves Richards equations (ref) for water transport and the one-dimensional Fourier equation for heat. The soil is discretized in 14 layers over a depth of 12 m."

We have rewritten L.119–123 as follows: "We use in this paper the ISBA multilayer diffusion scheme which solves the mixed form of Richards equations (Richards, 1931) for water and the one-dimensional Fourier law for heat (Boone et al., 2000; Decharme et al., 2011). The soil is discretized in 14 layers over a depth of 12m. The lower boundary of each layer is 0.01, 0.04, 0.1, 0.2, 0.4, 0.6, 0.8, 1.0, 1.5, 2.0, 3.0, 5.0, 8.0, and 12 m depth (see Fig. 1 of Decharme et al., 2013). The chosen discretization minimizes the errors from the numerical approximation of the diffusion equations."

L123: "..minimize the errors from the numerical approximation of the diffusion equations".

See previous response.

L124: "the water and carbon..."

Correction done.

L141: "simulated runoff into simulated river discharge".

Correction done.

L144: "The coupling between ISBA. ..."

Correction done.

L145: "groudwater" to "groundwater"

Correction done.

L146 "to CTRIP, while the. ..."

Correction done.

L149: ". . .assimilation system with a 24 h assimilation window".

Correction done.

L155: "...covariances are considered)."

Correction done.

L156: "..from a time t to t + 24 h."

Correction done.

L156-L157: Please rephrase, maybe: "The update of patch p is denoted..."

We have rewritten L.156–157 for clarity: "The forecast step consists of propagating the state of the system from a time t to $t+24$h using ISBA. Patches in each ISBA grid cell do not interact between each other. This implies that, for a patch p, the forecast of $\mathbf{x}_{[p]}$, denoted $\mathbf{x}_{[p]}^{f}(t+24$h), only depends on the analysis at time t, $\mathbf{x}_{[p]}^{a}(t)$, and the ISBA LSM using the parametrization for patch p, denoted by $\mathcal{M}_{[p]}$. It gives: ..."

L161-L162: Please describe what y0 is, and also what you mean by available at the grid cell level (instead of at the individual patches?).

To improve readability, we have rewritten the whole paragraph L.161–165 as follows: "LDAS-Monde uses routinely a Simplified Extended Kalman Filter for the analysis step (Mahfouf et al., 2009). Observations (SSM and/or LAI) are interpolated on the ISBA grid for assimilation (see Sect. 3.2 for more information). For each ISBA grid cell, we consider the vector $\mathbf{y}^{o}$ containing all the observations available for that grid cell at the time of assimilation. The SEKF analysis step is in two-step. First we calculate the model equivalent, denoted by $\mathbf{y}^{f}$, at the ISBA grid cell level. This is performed by aggregating control variables from each patch of the ISBA grid cell using a weighted average: ..."

L162: Suggested change: ".., we first calculate model equivalents of the observations. This is done separately for each individual grid cell."

See previous response.

L168: ".., it replaces the forecast error covariance matrix (B) with a fixed error covariance matrix. Please rephrase from and uses as..." Product of what? The model state evolution and?

The whole paragraph L.168–171 has been rewritten as follows:"Then, the SEKF analysis step is performed for each ISBA grid cell. We further assume that there are no covariances between the patches. Therefore, each patch is updated separately. For each patch, the SEKF analysis follows the traditional Kalman update. It replaces the forecast error covariance matrix with a fixed prescribed error covariance matrix $\mathbf{B}$. The observation operator is the product of the model state evolution from t to $t+24$h and the conversion of the model state into the observation equivalent. Thus, the Jacobian of the observation operator involves $\mathbf{H}$ and $\mathbf{M}_{[p]}$, the Jacobian matrix of $\mathcal{M}_{[p]}$. In the end, for each patch p, we have:"

7

In addition to the previous modification, we have added the full detail on how to obtain eq. (4) and (5) of the manuscript following a suggestion from referee #2.

L170: "..that there are no covariances between the patches."

See previous response.

L177: "column of can..." of what?

We meant columns of $\mathbf{M}_{[p]}$. Correction done in the manuscript.

L182: "The EnKF approximates the classical Kalman Filter equations using the ..."

Correction done.

L187: "..where X_p = [] is the ensemble perturbation matrix."

Correction done.

L188: Please change "The forecast step is simple..." to "In the forecast step we propagate. ..." and "...from time t to t + 24 h using the ISBA LSM."

Correction done.

L189: "The analysis step then updates. ..."

Correction done.

L190: "of the observations..."

Correction done.

Equation (8): Missing punctuation.

Correction done.

L193-194: Given certain conditions?

The EnsRF analysis indeed produces an analysed ensemble whose mean and ensemble covariance matrix matches the Kalman filter analysis. The only condition is the linearity of the observation operator. This has been added in the manuscript. Contrary to other deterministic EnKFs, such as the ETKF, the formulation of the EnSRF ensures to produce automatically ensemble perturbations for the analysis that have a zero mean (see Sakov and Oke, 2008).

Reference:

Sakov, P. and Oke, P. E.: Implications of the Form of the Ensemble Transformation in the Ensemble Square Root Filters, Mon. Weather Rev., 136, 1042–1053, 10.1175/2007MWR2021.1, 2008

L195: "We choose to neglect the ensemble. . .."

Correction done.

L195-L196: "This assumption is in line with the SEKF method and it ensures a fair comparison between the two approaches."

Correction done.

L196-L198: Please rephrase from "It also allows. . .." This is already given since you are working on a 1D EnKF. Maybe change to: "The approach outlined here is in line with other studies (ref) showing

that the 1D-EnKF can achieve promising results with around 20 ensemble members."

Correction done.

L201: "... patch p ..."

Correction done.

Equation (11): is subscript k defined?

Equation (11) should be:

$$\mathbf{C}_{\mathrm{EnSRF}} = \sum_{k=1}^{12} \alpha_{[k]}^2 \mathbf{H} \mathbf{P}_{[k]}^f \mathbf{H}^T + \mathbf{R} \qquad (1)$$

The error has been corrected in the updated version of the manuscript.

L210: Please rephrase, maybe: "This ensures that information from the analysis is stored in the ensemble and is propagated forward in time."

Correction done.

L213-L214: This should be placed in the "Experimental setup" section.

The sentence has been moved to Sec. 3.4.

L218-L219: "The ERA-5. . .." and ". . .. 31 km horizontal spatial resolution."

Correction done.

L219: Please change "To be used,..."

To be used has been removed.

L221: "or wind speed" is wind speed optional? "...interpolated to the ISBA 0.25 spatial resolution using bilinear interpolation."

Wind speed is not optional, we meant here an "and" rather than an "or". Correction done.

L222-L223: Please rephrase this sentence "...reanalysis improves the quality of LSVs reanalyses."

The sentence has been replaced by the following: "Replacing ECMWF's atmospheric ERA-Interim reanalysis by ERA5 has been shown beneficial in the context of LSVs reanalyses with LDASs (Albergel et al. 2018a,b)".

L226: Remove space between end of sentence and punctuation.

Correction done.

L226: "These satellite-derived products have already been successfully assimilated in the LDAS-Monde..."

Correction done.

L230: "..in order to measure.."? Please clarify what you mean by "measure" in this context.

We have replaced "measure" by "estimate".

L232: "Prior to the assimilation, the SWI. ..."

Correction done.

L234: "..the soil..."

We are sorry but we have not been able to understand what referee #1 meant here.

L235: Please clarify, you say that you use a linear rescaling but you also use a CDF method? The linear rescaling corrects the mean and variance while the CDF matching corrects all modes of the distribution?

We use in this paper a seasonal linear rescaling. Linear rescaling was introduced by Scipal et al. (2008) and has been shown giving results that are very similar to an exact CDF matching. Nevertheless, to avoid any confusion, we have rewritten the sentence as follows: "Introduced by Scipal et al. (2008), this rescaling gives in practice very similar results to CDF (cumulative distribution function) matching. The linear rescaling is performed on a seasonal basis (with a 3-month moving window)." Further mentions of CDF matching in the manuscript have been replaced by "seasonal linear rescaling".

L238: Please define GEOV1.

GEOV1 stands for GEOLAND2 Version 1. This has been added to the manuscript.

L239: "10 days.. with the finest spatial resolution being 1 km."

Correction done.

L251: Pleas define GPP at first occurrence (L245).

Correction done.

L252: "...from eddy-covariance flux towers..."

Correction done.

L253: "The FLUXCOM data are available. . .."

Correction done.

L256: Please make clear that this is model output data. "River discharge output from the CTRIP is. ..." and "data obtained from the Global. ..."

Correction done.

L261: Change "efficiency" to "skill"?

Correction done.

L269: "...for the soil moisture variable." Who are "They", the perturbations?

Indeed "They" means soil moisture perturbations. This has been corrected in the updated version of the manuscript.

L270: Please consider moving this sentence for after the SM perturbations.

Modification done.

L271: "covariance ," remove space

Correction done.

L272: "..dynamic range of soil moisture."

Correction done.

L274: Drop "successful"

Correction done.

L276: "..and using B for the covariance matrix." What does this mean?

We meant here that we sample the initial ensemble of the EnSRF from a multivariate Gaussian distribution using the prescribed **B** matrix of the SEKF as covariance matrix for that multivariate Gaussian distribution. We have rewritten L.275–276 as follows:"About the EnSRF configuration, the initial ensemble is obtained by perturbing the initial state using perturbations sampled from a multivariate Gaussian distribution with a zero-mean and using the prescribed **B** covariance matrix used in the SEKF as the covariance matrix of that multivariate Gaussian distribution."

L276: What do you mean by "underestimate ensembles" ensemble spread?

We meant indeed ensemble spread. Correction done.

L277: "..artificially small ensemble spread..."

Correction done.

L278: "have"?

"have" is correct if we consider the authors separately but "has" is also correct if we consider Whitaker and Hamill (2005) as one scientific publication.

L280: White noise of what?

We meant here a Gaussian noise. The manuscript has been corrected accordingly.

L281: Please define and use the SM acronym earlier.

The SM acronym is defined in the introduction. We have replaced "soil moisture" by "SM".

L283: White noise of what?

We meant here a Gaussian noise. The manuscript has been corrected accordingly.

L284: "This is similar to the work of. ..."

Correction done.

L287: Reference to these studies?

References to Albergel et al.(2017), Leroux et al. (2018) and Tall et al. (2019) have been added.

L290: Please drop "sanity"

Correction done.

L290: "open loop counterparts"

Correction done.

L291: Please clarify, what is "those two LSVs"?

Correction done.

L295: "the ensemble..."

Correction done.

L299: "response to SM6"? Do you mean they have the same behavior?

We meant indeed that SM5, SM6 and SM7 have the same behaviour. This has been clarified in the manuscript.

L300: "Potential improvements in EnSRF and SEKF estimates of evapotranspiration. ..."

Correction done.

L307: What do you mean by "accuracy"?

We meant here that, when NSE $= 0$, it means that the simulated or analysed river discharges Q_t^s provide a similar NSE as the observed averaged river discharge $\overline{Q}^o$ (see Eq. 13). We have rewritten the sentence as follows: "A NSE value of 0 means that the model/analysis has the same NSE as the observed averaged river discharge."

L308: "open loop run"

Correction done.

L312: Suggest you change this to something like this: "Figure 2 displays the open loop, SEKF, EnSRF and observed LAI 10-day time series..."

Correction done.

L319: Please use "open loop" for the rest of the text.

We have replaced the expression "model run" by "open loop" in the whole manuscript (including Tables and Figures) when it was appropriate.

L319-L320: The numbers show that the SEKF is closer than the EnSRF, this is the opposite of what you say in L317?

The expression L. 317 "with EnSRF estimates getting closer to observations than SEKF ones" was only valid for autumn (while not being obvious from Figure 2). This has been removed from the manuscript to avoid any confusion.

L320: Please make it clear that this is no longer independent validation data, as you compute the skill between the analyses and the observations assimilated in the analyses.

The following sentence has been added in the paragraph: "As expected, both DA approaches produce estimates that are closer to the assimilated LAI observations than their open loop counterpart."

Figure 3: Please provide masked regions with different color coding than zero values.

Figure 3 has been modified to change color coding for masked regions.

L344: "Strongest" to "Largest", "... occur for both cases. ..."

Correction done.

Figure 6: Does this figure not also show the spatial std, since it is averaged over the whole domain?

Indeed, that is why we have written "Figure 6 displays the seasonal evolution of ensemble standard deviations averaged over the whole domain and for grid cells dominated by one type of vegetation."

L355: Linear rescaling not CDF matching?

"CDF matching" has been replace by "seasonal linear rescaling".

L362: "southeast" and "northern"

Correction done.

Figure 7 (a), change title to "RMSD open loop".

Correction done.

L369: Note that this is for a single open loop run? An ensemble open loop run might improve more?

Correlations are indeed for a single open loop run. An ensemble open loop run might improve or degrade correlations with observed SSM depending on how model perturbations are generated. However, this remains out of scope of the current paper.

Figure 8: Please provide different color coding for masked regions vs zero value regions.

Figure 8 has been modified accordingly.

L376: "..the Jacobian is replaced by correlations sampled. ..."

Correction done.

L382: "..and correlations with SM2."?

Expression removed, it did not mean anything.

L283: "extend" to "extent"

Correction done.

L385: "western", "spring" and "summer". Please double check this spelling throughout the text e.g., L396-L398.

The spelling has been corrected throughout the manuscript.

L402: "Nevertheless we discern seasonal tendencies"? Please explain this sentence.

Section 4.4 has been fully rewritten and the sentence removed.

L402: What about SM6 and the abrupt change close to the Arctic circle?

We thank the referee for her/his excellent question. The abrupt change close to the Arctic circle is due to modified hydraulic and thermal soil properties in ISBA for arctic regions. This module has been developed by Decharme et al. (2016) in order to include a dependency on soil organic carbon content for ISBA's hydraulic and thermal soil properties. We have added a comment on this subject in section 2.1 dedicated to ISBA and another comment in section 4.4.

Reference:

Decharme, B., Brun, E., Boone, A., Delire, C., Le Moigne, P., and Morin, S.: Impacts of snow and organic soils parameterization on northern Eurasian soil temperature profiles simulated by the ISBA land surface model, The Cryosphere, 10, 853877, 10.5194/tc-10-853-2016, 2016.

L403: "... increments in SM4 ..."

Correction done.

L404: "tends" to "tend"

Correction done.

L406: "..disparity over arid regions is..."

Correction done.

L408: "estimates" to "increments"?

We meant here SM4 estimates.

L410: Please rephrase this sentence, start with: "The SM4 estimates and analyses increments for the SEKF and EnSRF tend to be similar, except for arid regions."

Section 4.4 has been fully rewritten to avoid that confusion.

L413-L418: Please rephrase this paragraph. What do you mean by "cycling"? "...does not modify directly estimates as correlations. ..." estimates of what?

Section 4.4 has been completely re-organised to issue those questions.

L423: Please define ET acronym earlier and GPP is already defined.

Correction done.

L428-L429: "..for almost all grid cells."

Correction done.

L436: "best" to "biggest" or "largest" and remove "on"

Correction done.

L441: Please rephrase "..., thus validating our approach."

The expression has been removed.

L445: How do you decide on this 3% limit?

We agree this 3% limit is arbitrary. It is just similar to what has been used previously for NIC in others of our publications, see e.g. Albergel et al. (2018b). This limit also allows a better visualisation of the NIC improvement or degradation in Figure 14.

L446: Maybe state that the rest of the stations (20) showed a neutral impact?

Agreed. A sentence has been added in the paragraph.

L457: What are these favorable atmospheric conditions?

We meant here that LAI dynamics depends more on atmospheric forcing than on initial conditions during the growing phase. This implies that, while assimilating observed LAI can add LAI and biomass, the effect of assimilation will fade quickly. On the contrary, during the senescence, LAI dynamics is driven by the rate of mortality, thus making assimilation more efficient. We have modified the sentence L. 456–458 as follows: "During the growing phase, modelled LAI is more sensible to atmospheric conditions than to initial LAI conditions. This implies that, while DA can artificially add LAI and biomass, its impact can be limited by the atmospheric forcing."

L461: "on" to "in" and what do you mean by "LAI dynamics is weak in those places"?

We meant there that the amplitude of the LAI annual cycle is smaller in those places than for places dominated by deciduous trees. This has been clarified in the updated version of the manuscript.

L464: ".., model perturbations can introduce. ..." remove "thus showing its influence."

Correction done.

L469: Please rephrase, for example: "Model perturbations can lead to LAI values below this threshold..."

Correction done.

L470: "when this is the case it can lead to. ..."

Correction done.

L475-L478: What does "more uncertainty in the additive model error" mean? It increases the size of the increments and therefore the EnSRF is closer to the observations than the SEKF? Please clarify this section.

We meant here that the prescribed model error in the EnSRF leads to ensembles with bigger standard deviations (SD) than the prescribed SD for soil moisture in the second layer of soil (1–4 cm depth). This implies that observations of SSM have a bigger weight in the EnSRF than in the SEKF. Thus, EnSRF estimates are closer to SSM observations than SEKF estimates. This has been clarified in the updated version of the manuscript.

L479: "CDF match" to "CDF matching", are you sure that this is the approach you are using?

We have replaced "CDF match" by "seasonal linear rescaling" (see previous comment on that subject).

L479: "This shows that the short-term variability of the observations is different from what we model with ISBA in this region."

Correction done.

L480: Check CDF matching.

We have replaced "CDF matching" by "seasonal linear rescaling" (see previous comment on that subject).

L482: Maybe change to: "Further studies of such aspects are beyond the scope of this paper."

Correction done.

L483: "...both DA approaches..."

Correction done.

L484: "model error" to "model perturbations"

Correction done.

L486: Change "fatally" to "could potentially..."

Correction done.

L488: "..by the model perturbations..."

Correction done.

L491: "summer"

Correction done.

L493: "southwest"

Correction done.

L494: Please rephrase to make it clear that covariances between the ensemble members explain the relationship between e.g., soil moisture in different soil layers in ISBA.

This sentence has been rewritten following the merge of Section 4 and Section 5.1.

L494: Maybe change to: "Another type of model error..." and "...different characteristics of the covariances between the ISBA variables."

Correction done.

L496: "also provides"

Correction done.

L497: "their" to "the"?

Correction done.

L500: Remove "Considered out of scope for this paper."

Correction done.

L501: "...dams, . . .) can potentially modify soil moisture, streamflow and river discharge." Maybe provide a reference for this?

Correction done. The following reference has been added:

Milano, M., Ruelland, D., Dezetter, A., Fabre, J., Ardoin-Bardoin, S. and Servat, E.: Modeling the current and future capacity of water resources to meet water demands in the Ebro basin, J. Hydrol., 500, 114-126, 10.1016/j.jhydrol.2013.07.010, 2013.

L506: "more physical states"? Do you mean "...ensemble of land surface states"?

We meant here that by using perturbed atmospheric forcings, it would lead to more physical model perturbations and to an ensemble with covariance that are more physically-based. We have modified the sentence to reflect that point.

L514: "the model error."

Correction done.

L517: Include a reference to the original Desroziers paper.

Reference added.

L521: "...processes, etc).."

Correction done.

L524: "such ideas have"

Correction done.

L526: Please simplify section heading. For example, "The question of 1D or 3D filtering"

Correction done.

16

L529: "to" to "in correlated..."

Correction done.

L530: Provide reference for this statement. Same for line L531.

A reference to the upcoming paper on ERA5 reanalysis from Hersbach et al. (2019) has been added. The paper will include a description of associated uncertainties.

L532: I dont understand, could you please clarify this section? You say that the SEKF cannot include covariances, but it relies on ISBA to calculate covariances. Why does the SEKF need covariances from ISBA when it cannot include them?

Indeed the SEKF relies on ISBA to calculate covariances. But, since patches in ISBA do not interact between each other, the Jacobian cannot build those covariances between patches from the model. Therefore to include covariances between patches, they have to be prescribed in the fixed background error covariance matrix in the SEKF. The same problem occur for spatial covariances as ISBA grid cells do not interact with each others. This has been clarified in the updated version of the manuscript.

L537: "...12 times the size..."?

If someone wants to consider covariances between patches, she/he has to consider in the control vector LAI and soil moisture from each patch. Since each ISBA grid cell is divided into 12 patches, it means that the control vector would have to be 12 times bigger than the one we used in this paper.

L549: Please rephrase, for example: "approach, because of the 1D nature of the ISBA LSM."

Correction done.

L550: "applications"

Correction done.

L553: "based on spatial characteristics and it..."

Correction done.

L559: "Results show"

Correction done.

L563: "..the model error perturbations."

Correction done.

L565 "surface. The EnSRF..." Please rephrase after this, what are those estimates and what are those layers?

We have replaced "those estimates" by "soil moisture estimates" and "those layers" by "soil layers either near the surface or in the root zone".

L569: "for the two previous"? Please make it clear what the two previous are.

This has been clarfied in the updated version of the manuscript.

L569: Please clarify "While involving a crude model error", the sentence feels a bit out of context.

We have removed the expression for clarity.

L572: Maybe change to: "(for CGLS products). This only allows for an update of LAI every 10-days,

17

as the assimilation of surface soil moisture is found to have negligible impact on the LAI analyses."

Correction done.

L573: "for the radar backscatter..."

Correction done.

L574: What do Livens (2017) and Shamambo (2019) show?

These two papers show how radar backscatter coefficients can be linked to LAI or vegetation optical depth through a water cloud model. Assimilating radar backscatter coefficients in LDAS-Monde would imply developing an observation operator linking modelled LAI and surface soil moisture to radar backscatter coefficients and the water cloud model seems to be a good candidate for the observation operator. This has been clarfied in the updated version of the manuscript.