Review Anonymous Referee #2

General Comments:

This paper presents the LDAS-Monde EnSRF adapted to multivariate soil moisture and LAI assimilation. Results are presented and compared to the SEKF and to the model in the Euro-Mediterranean region for 2008-2017. The paper provides substantial contribution to scientific progress in the field of land surface data assimilation which is relevant for HESS. The analysis is very thorough, the paper is very well written and presented. I suggest it is published after the following comments are taken into account.

We thank the referee for her/his positive comments about our work and for her/his insightful 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.

Specific comments:

line 52, sentence starting by "Both brightness temperature..." is too vague: not all brightness temperature are influenced by vegetation dynamics. The authors should indicate specify that it is for low microwave frequencies

We thank the referee for mentioning that point, this has been added in the manuscript.

line 91-92: I find it too detailed to give the latitude and longitude min and max of the studied area in the introduction. These details are given in Section 3 and this is enough.

We have removed these details from the introduction in the revised version of the manuscript.

line 97: You should perhaps add the reference to the peer reviewed ERA5 paper submitted to by Hersbach et al. in 2019. Same comment line 218.

Agreed, the reference has been added.

line 170-175: it would be very useful to give more details on the patch formulation in equations 3-4 as it was not provided in any of the previous papers describing the SURFEX SEKF. It could be added as an annex.

To that purpose, we have decided to add the full details on the patch formulation in supplementary material. The justification is too long to be put in appendix. For information, the supplementary material has been added at the end of this response.

line 317: It is not very clear on this figure that the EnSRF estimates get closer to observation than the SEKF ones. Please revise the sentence.

"with EnSRF estimates getting closer to observations than SEKF ones" has been removed from the sentence.

line 320-323: The authors should refer to Table 1 at this stage of the results presentations. Table 1 is only used in support of the results presentation in section 4.5 line 430). It would be very useful to refer to it everywhere its statistics are discussed. Same comment applies for example line 360, lines 368-369, .

Following referee #2's suggestion, Section 4 has been modified to include references to Table 1 when statistics are discussed.

Figure 3, caption is not clear. Replace "and difference between nRMSD for SEKF (b) and EnSRF (c) vs nRMSD Model." by "and nRMSD difference between assimilation experiments (SEKF in a, and EnSRF in b) and Model".

The caption for Figures 3, 7, 8, 12 and 13 has been modified following referee #2's suggestion.

lines 332-345: This analysis and corresponding figures are interesting to understand the performance of the assimilation systems for the different vegetation types. It does not help to understand the EnSRF degradation in NW Spain and in the Alps shown in Figure 3. The authors should investigate further and present in the paper results in these areas.

This analysis explains indirectly the EnSRF degradation in NW Spain and in the Alps. In those places, coniferous trees represent around 40% of the vegetation and grasslands around 30%. They are the two types of vegetations for which the EnSRF performs poorly. In practice, in both places, the ensemble collapses for the coniferous trees patch and the ensemble for grasslands cannot compensate for such collapse due to a too-small ensemble spread. Few sentences have been added in section 4.1 to explain the issue.

line 321-323: The authors should comment on the fact that the model bias is the lowest also because the winter and summer negative bias is partly compensated by positive bias in autumn, whereas DA experiments correct for the autumn positive bias only.

We thank the referee for her/his interesting remark. We have added few lines on that subject in section 4.1 of the manuscript.

line 371: It would be useful to comment on the negative correlations shown between model output and observed SSM in arid areas. The explication (short term variability) is given later and discussed in Section 5.1 (lines 479-480). It should be briefly mentioned already when the correlation map are shown.

Instead and as suggested, we have merged Section 5.1 with Section 4. The following sentences have been added to Section 4.2: " Finally we observe negative correlations between the open loop and observed SSM (even with the seasonal linear rescaling) in arid places such as deserts in Sahara and the Arabian Peninsula. This shows that the short-term variability of the observations is different from what we model with ISBA in this region. It raises the question of the quality of ISBA and/or SSM observations (after seasonal linear rescaling) in arid places. Stoffelen et al. (2017) has shown that observed SSM derived from scatterometers can have a poor quality in arid places."

line 388: the authors should explain or clarify in the text why SM2 and SM6 are uncorrelated in summer over Spain and Northern Africa?

We have added in Section 4.3 the following sentence for clarification: "This decorrelation between surface and root-zone soil moisture occurs during very dry conditions such as occured in Spain and northern Africa during summer."

line 396: explain here the meaning of larger LAI leading to drier soil. It is pretty obvious that it is related to more evaporation, meaning that LAI influences soil moisture in this case, but it would be interesting to discuss here as the previous sentence is the other way around, with positive correlation and soil moisture influencing LAI. As commented above, it is not optimal to have the results presented here, but only partly explained, with the full explanation later in section 5.1. When reading Section 5.1, we have to go back in the paper to match the figure description and the figure interpretation given several pages later in Section 5.1. Please revise the text by merging section 5.1 with the presentation of the results in Section 4.

We have, as suggested, merged Section 5.1 with Section 4. Negative correlations between LAI and soil moisture occurs for dry soils due to evaporation. But for wetter areas, evaporation is far less important and correlations between LAI and soil moisture are instead positive.

line 400-411: This paragraph starts with Figure 10, but then the second and third sentences "We observe that the SEKF has the same averaged SM4 as the model. Nevertheless we discern seasonal tendencies." are clearly not related to Figure 10. Then it discussed Figure 11, but line 404-405 ("EnSRF estimates..") content does not match Figure 11, but it is more adapted to Figure 10. So, this paragraph needs to be slightly reorganized.

Section 4.4 was poorly written. It has been re-organised as follows to improve clarity:

- Averaged estimates of SM4 and SM6 depicted in Figure 10 are first studied. We have highlighted the wet bias introduced by EnSRF model perturbations in places where assimilation of SSM and LAI plays no role (no correlation with SM4 or SM6).
- Then analysis increments from the SEKF and EnSRF are compared for SM4 using Figure 11.
- Finally we write few lines on increments from the SEKF and EnSRF for SM6 (no figure associated).

line 413-415: "...in Figure 10. We identify these patterns for every month without any seasonality (not shown). For SEKF drier estimates are obtained through cycling as analysis increments are close to zero. For EnSRF, cycling is also responsible to this drying but analysis increments are not negligible (-0.01 $m^3.m^{-3}$ for biggest values) and compensate the wet bias from model error in SM6 (not shown)." This suggest that SM6 negative increments have a larger amplitude in EnSRF than in EKF, however this is not obvious from Figure 10.

Indeed, Figure 10 shows roughly similar SM6 estimates for the SEKF and the EnSRF. But they are not entirely obtained through the same process. Differences between SEKF and open loop SM6 estimates are solely due to the joint effect of the ISBA LSM and the updated LAI and soil moisture near the surface (SEKF increments are close to zero). In the EnSRF case, increments, which are not shown for SM6, have a bigger size (in absolute value) because DA compensates the wet bias introduced by model perturbation. So in the EnSRF, both analysis and the joint effect of the ISBA LSM and the updated LAI and soil moisture near the surface play a role for SM6. This point has been hopefully clarified in Section 4.4.

Section 5: The discussion provided in Sections 5.2 and 5.3 is excellent, it discusses the limits of the proposed EnSRF approach and perspectives to improve the system. Section 5.1 is less relevant to the discussion as it mainly supports of the results description and it provides information that was actually missing when reading Section 4 (see comments above). So, Section 4.1 (or most of it) should be merged with corresponding paragraphs Section 4.

We thank the referee for her/his kind comments. Following her/his advice, we have merged Section 5.1 with Section 4 in the updated version of the manuscript.

line 576-577: the last sentence of the conclusion, starting with "Once fully tested, it should, hopefully, provide daily..." sounds technical and hazardous. Replace by something like "It will open the possibility to have access to daily..."

Correction done.

Technical corrections:

- line 49: "Recently," (add a comma)

Correction done.

- line 71: replace "has" by "have". Also line 74 twice.

Referee #1 suggested to drop "have" and "has" in those sentences. We followed her/his advice.

- line 109: move the reference to Albergel at al. and the end of the sentence. Same comment for the references given line 118 and line 119.

Modification done.

- line 121: "The lower boundary of the 14 soil layers (0.01...) .. was chosen to" is not clear. A more accurate language would perhaps be "The vertical soil discretization into 14 layers (0.01...) .. was chosen to"

Following suggestions from both referees, 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."

-line 177: remove "of"

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

-line 222: remove "ISBA"

Correction done.

- line 404: add "particularly" as follow: "..in July, particularly in Northern Europe..."

Correction done.

-line 413: "For the SEKF, ..."

Section 4.4 has been fully rewritten instead.

-line 462: "introduce a larger negative bias"

Section 5.1 has been merged with Section 4 instead.

-line 539-540: the sentence starting by "However, if we take ..." sounds familiar, reformulate it.

The sentence has been replaced by the following one:"However, including covariances between patches or between grid cells would make undersampling and spurious covariances more likely to occur due to the increased size of the state vector."

-line 541: what caveats?

We meant by "caveats" undersampling and spurious covariances. We have replaced the word "caveats" by the expression "potential issues".

In this supplementary material, we detail how the equations of the SEKF are derived in the context of patches and the ISBA land surface model.

1 Simplified Extended Kalman Filter

We first recall in this section the equations of the Simplified Extended Kalman Filter. Introduced by [1], the SEKF is a simplified version of the Extended Kalman Filter. It is a sequential approach aiming to give the estimation of the state $\mathbf{x}$ of a system at various times. We denote by $n_{\mathbf{x}}$ the size of the state vector.

The SEKF is a two-steps algorithm. For a given time t_k, it provides a first estimate $\mathbf{x}_k^f$ called the *forecast*

$$\mathbf{x}_k^f = \mathcal{M}_{k-1}\left(\mathbf{x}_{k-1}^a\right) \tag{1}$$

with $\mathcal{M}_{k-1}$ a (nonlinear) model. The forecast step just aims to propagate the estimate $\mathbf{x}_{k-1}^a$ at the last previous time t_{k-1} to the new time t_k.

This forecast is then corrected by using observations $\mathbf{y}_k^o$ of the system with $\mathbf{R}_k$ its associated error covariance matrix. We denote by $n_{\mathbf{y}}$ the size of $\mathbf{y}_k^o$. The observations are linked to the state through the (possibly nonlinear) observation operator $\mathcal{H}_k$. This correction step is called the *analysis* and provides a new estimate $\mathbf{x}_k^a$ with

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k\left(\mathbf{y}_k^o - \mathcal{H}_k\left(\mathbf{x}_k^f\right)\right) \tag{2}$$

$$\mathbf{K}_k = \mathbf{B}\,\mathbf{J}_k^T\left(\mathbf{J}_k\,\mathbf{B}\,\mathbf{J}_k^T + \mathbf{R}_k\right)^{-1} \tag{3}$$

$\mathbf{B}$ is a prescribed background error covariance matrix of size $n_{\mathbf{x}} \times n_{\mathbf{x}}$ and $\mathbf{J}_k$ is a Jacobian matrix of size $n_{\mathbf{y}} \times n_{\mathbf{x}}$ defined as

$$\mathbf{J}_k = \frac{\partial\left(\mathcal{H}_k\left(\mathbf{x}_k^f\right)\right)}{\partial\,\mathbf{x}_{k-1}^a} = \frac{\partial\left(\mathcal{H}_k\left(\mathcal{M}_{k-1}\left(\mathbf{x}_{k-1}^a\right)\right)\right)}{\partial\,\mathbf{x}_{k-1}^a} \tag{4}$$

This Jacobian can be estimated using finite differences. In that case, we would need to run $n_{\mathbf{x}}$ perturbed model runs in addition to the model run used in the forecast step. If $n_{\mathbf{x}}$ is too big, computing $\mathbf{J}_k$ with finite differences is unaffordable.

2 First assumption: linearity of the observation operator

We now assume that the observation operator meaning that $\mathcal{H}_k = \mathbf{H}_k$. This implies that the Jacobian matrix $\mathbf{J}_k$ can be rewritten as

$$\mathbf{J}_k = \frac{\partial\left(\mathbf{H}_k\mathbf{x}_k^f\right)}{\partial\,\mathbf{x}_{k-1}^a} = \mathbf{H}_k\frac{\partial\,\mathbf{x}_k^f}{\partial\,\mathbf{x}_{k-1}^a} = \mathbf{H}_k\frac{\partial\left(\mathcal{M}_{k-1}\left(\mathbf{x}_{k-1}^a\right)\right)}{\partial\,\mathbf{x}_{k-1}^a} = \mathbf{H}_k\,\mathbf{M}_{k-1} \tag{5}$$

with $\mathbf{M}_{k-1}$ the tangent linear operator of $\mathcal{M}_{k-1}$ at $\mathbf{x}_{k-1}^a$.

Following this assumption, the analysis step of the SEKF is now:

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k\left(\mathbf{y}_k^o - \mathbf{H}_k\,\mathbf{x}_k^f\right) \tag{6}$$

$$\mathbf{K}_k = \mathbf{B}\,\left(\mathbf{H}_k\,\mathbf{M}_{k-1}\right)^T\left(\left(\mathbf{H}_k\,\mathbf{M}_{k-1}\right)\,\mathbf{B}\,\left(\mathbf{H}_k\,\mathbf{M}_{k-1}\right)^T + \mathbf{R}_k\right)^{-1} \tag{7}$$

3 The case of LDAS-Monde, ISBA and patches

Until now, we have not assumed anything regarding the spatial distribution of state variables and observations.

The ISBA land surface model involved in LDAS-Monde owns features that can help to simplify the SEKF. They are:

- At grid cell level: ISBA only consider vertical diffusion for soil moisture and temperature and vegetations variables of different grid cells do not interact with each other.
- Each grid cell of ISBA is divided into 12 different patches representing different types of vegetation. To each patch p is associated a patch fraction $\alpha_{[p]}$ representing the proportion of the type of vegetation associated to patch p in the grid cell.
- At patch level: variables (vegetation, soil moisture, soil temperature, ...) of different patches do not interact with each other.

Second assumption: Observations are available at ISBA grid cell level and no spatial covariances are taken into account in LDAS-Monde.

Following this second assumption, equations (6) and (7) can be applied directly at a grid cell level. This allows an easy parallelisation of the SEKF analysis using domain decomposition.

Now we split the control vector $\mathbf{x}$ into 12 vectors $\mathbf{x}_{[p]}$, $p = 1, \ldots, 12$, each containing only control variables relative to that particular patch. It means we have 12 LAI variables (one for each patch), 12 SM2 variables (soil moisture in layer 2, 1-4 cm depth), etc. $\mathbf{x}$ can be written as the concatenation of these 12 vectors:

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_{[1]} \\ \mathbf{x}_{[2]} \\ \vdots \\ \mathbf{x}_{[12]} \end{pmatrix} \tag{8}$$

While control variables are available at patch level, observations are available at grid cell level. It means that variables at patch level need to be aggregated to grid cell level to obtain observation equivalents.

Third assumption: The observation operator $\mathbf{H}_k$ aggregates control variables at patch level averaging them with patch fractions as weights:

$$\mathbf{H}_k \mathbf{x} = \sum_{j=1}^{12} \alpha_{[j]} \, \widetilde{\mathbf{H}}_k \, \mathbf{x}_{[j]} \tag{9}$$

$\widetilde{\mathbf{H}}_k$ is a matrix selecting directly the observed variable (either LAI and/or SM2) meaning that $\widetilde{\mathbf{H}}$ is full of 0 and 1.

Following the third assumption, the observation operator $\mathbf{H}_k$ can also rewritten as:

$$\mathbf{H}_k = \begin{pmatrix} \alpha_{[1]} \, \widetilde{\mathbf{H}}_k & \alpha_{[2]} \, \widetilde{\mathbf{H}}_k & \ldots & \alpha_{[12]} \, \widetilde{\mathbf{H}}_k \end{pmatrix} \tag{10}$$

Since variables of different patches do not interact with each other in ISBA, it also simplifies the Jacobian matrix $\mathbf{M}_{k-1}$ making it block-diagonal as follows:

$$\mathbf{M}_{k-1} = \begin{pmatrix} \mathbf{M}_{[1],k-1} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{[2],k-1} & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{M}_{[12],k-1} \end{pmatrix} \tag{11}$$

It leads that $\mathbf{H}_k\,\mathbf{M}_{k-1}$ can be now written as:

$$\mathbf{H}_k\,\mathbf{M}_{k-1} = \begin{pmatrix} \alpha_{[1]}\,\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[1],k-1} & \alpha_{[2]}\,\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[2],k-1} & \dots & \alpha_{[12]}\,\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[12],k-1} \end{pmatrix} \tag{12}$$

Fourth assumption: No covariances between patches are taken into account in LDAS-Monde.

This assumption leads to a block diagonal $\mathbf{B}$ matrix that can be defined as:

$$\mathbf{B} = \begin{pmatrix} \widetilde{\mathbf{B}} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \widetilde{\mathbf{B}} & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \widetilde{\mathbf{B}} \end{pmatrix} \tag{13}$$

with $\widetilde{\mathbf{B}}$ the prescribed covariance matrix for control variables within a patch. In practice $\widetilde{\mathbf{B}}$ is taken diagonal.

Using this new definition of $\mathbf{B}$ and equation (12), $\mathbf{B}\,(\mathbf{H}_k\,\mathbf{M}_{k-1})^T$ can be written as:

$$\mathbf{B}\,(\mathbf{H}_k\,\mathbf{M}_{k-1})^T = \begin{pmatrix} \alpha_{[1]}\,\widetilde{\mathbf{B}}\left(\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[1],k-1}\right)^T \\ \alpha_{[2]}\,\widetilde{\mathbf{B}}\left(\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[2],k-1}\right)^T \\ \vdots \\ \alpha_{[12]}\,\widetilde{\mathbf{B}}\left(\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[12],k-1}\right)^T \end{pmatrix} \tag{14}$$

and

$$(\mathbf{H}_k\,\mathbf{M}_{k-1})\,\mathbf{B}\,(\mathbf{H}_k\,\mathbf{M}_{k-1})^T = \sum_{j=1}^{12} \alpha_{[j]}^2 \left(\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[j],k-1}\right)\widetilde{\mathbf{B}}\left(\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[j],k-1}\right)^T \tag{15}$$

Using equations (14) and (15) into (7), it gives for the gain matrix:

$$\mathbf{K}_k = \begin{pmatrix} \alpha_{[1]}\,\widetilde{\mathbf{B}}\left(\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[1],k-1}\right)^T \left(\sum_{j=1}^{12} \alpha_{[j]}^2 \left(\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[j],k-1}\right)\widetilde{\mathbf{B}}\left(\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[j],k-1}\right)^T + \mathbf{R}_k\right)^{-1} \\ \alpha_{[2]}\,\widetilde{\mathbf{B}}\left(\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[2],k-1}\right)^T \left(\sum_{j=1}^{12} \alpha_{[j]}^2 \left(\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[j],k-1}\right)\widetilde{\mathbf{B}}\left(\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[j],k-1}\right)^T + \mathbf{R}_k\right)^{-1} \\ \vdots \\ \alpha_{[12]}\,\widetilde{\mathbf{B}}\left(\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[12],k-1}\right)^T \left(\sum_{j=1}^{12} \alpha_{[j]}^2 \left(\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[j],k-1}\right)\widetilde{\mathbf{B}}\left(\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[j],k-1}\right)^T + \mathbf{R}_k\right)^{-1} \end{pmatrix} \tag{16}$$

Using this formulation of the gain matrix and equation (10) into equation (6), it leads to the following equations for the analysis in each patch p:

$$\mathbf{x}^a_{[p],k} = \mathbf{x}^f_{[p],k} + \mathbf{K}_{[p],k}\left(\mathbf{y}^o_k - \sum_{j=1}^{12} \alpha_{[j]}\widetilde{\mathbf{H}}\,\mathbf{x}^f_{[j],k}\right) \tag{17}$$

$$\mathbf{K}_{[p],k} = \alpha_{[p]}\,\widetilde{\mathbf{B}}\left(\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[p],k-1}\right)^T \left(\sum_{j=1}^{12} \alpha_{[j]}^2 \left(\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[j],k-1}\right)\widetilde{\mathbf{B}}\left(\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[j],k-1}\right)^T + \mathbf{R}_k\right)^{-1} \tag{18}$$

These two equations are equivalent to equations (3) and (4) of the manuscript.

In practice, we do not compute $\mathbf{M}_{[p],k-1}$ but directly $\widetilde{\mathbf{H}}_k\,\mathbf{M}_{[p],k-1}$ using finite differences.

References

[1] J.-F. Mahfouf, K. Bergaoui, C. Draper, C. Bouyssel, F. Taillefer and L. Taseva, A comparison of two offline soil analysis schemes for assimilation of screen-level observations, *J. Geophys. Res.*, 114, D08105, doi: 10.1029/2008JD011077, 2009.