

RESPONSE TO THE EDITOR AND TO THE REFEREES FOR THE PAPER

“AN ENSEMBLE SQUARE ROOT FILTER FOR THE JOINT ASSIMILATION OF SURFACE SOIL MOISTURE AND LEAF AREA INDEX WITHIN LDAS-MONDE: APPLICATION OVER THE EURO-MEDITERRANEAN REGION”

We would like to thank the editor and the two anonymous Referees for their supportive appreciation of our work and for their thorough review. Their comments and suggestions have led to an improved version of the manuscript. Our response is organised as follows:

- We first issue a point-by-point response to both reviews. Comments and suggestions from the Referees are in black and responses are in blue. Please note that, following a suggestion from Referee #2, section 5.1 has been merged with section 4. Nevertheless, every correction/suggestion for section 5.1 has been issued when remaining applicable.
- Then, the updated manuscript with highlighted modifications is appended after the point-by-point response.
- Finally, we provide, as supplementary material, the justification of the patch formulation of the SEKF as suggested by Referee #2.

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.

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° 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 as 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., Pignatelli, 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 $\mathbf{x}_{[p]}^f(t + 24\text{h})$, 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 y_0 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 (\mathbf{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 $t + 24\text{h}$ 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:”

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 $\mathbf{X}_p = \square$ 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:

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

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 re-

analysis 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: Please 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 \mathbf{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 \mathbf{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 \bar{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 replaced 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.

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 clarified 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, 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 clarified in the updated version of the manuscript.

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.

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 occurred 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 ($-0.01 \text{ m}^3 \cdot \text{m}^{-3}$ 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”.

An Ensemble Square Root Filter for the joint assimilation of surface soil ~~moiture~~ moisture and leaf area index within LDAS-Monde: application over the Euro-Mediterranean region

Bertrand Bonan¹, Clément Albergel¹, Yongjun Zheng¹, Alina Lavinia Barbu¹, David Fairbairn², Simon Munier¹, and Jean-Christophe Calvet¹

¹CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France

²European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

Correspondence: Clément Albergel (clement.albergel@meteo.fr)

Abstract. This paper introduces an Ensemble Square Root Filter (EnSRF), ~~a deterministic Ensemble Kalman Filter~~, to the context of assimilating jointly observations of surface soil moisture (SSM) and leaf area index (LAI) in the Land Data Assimilation System LDAS-Monde. By ingesting those satellite-derived products, LDAS-Monde constrains the Interaction between Soil, Biosphere and Atmosphere (ISBA) land surface model (LSM), coupled with the CNRM (Centre National de Recherches

5 Météorologiques) version of the Total Runoff Integrating Pathways (CTRIP), to improve the reanalysis of land surface variables (LSVs). To evaluate its ability to produce improved LSVs reanalyses, the EnSRF is compared with the Simplified Extended Kalman Filter, which has been ~~routinely operated in well studied within the~~ LDAS-Monde ~~; in a real case over the well-studied~~ framework. The comparison is carried out over the Euro-Mediterranean region at a 0.25° spatial resolution between 2008 and 2017. Both data assimilation approaches provide a positive impact on SSM and LAI estimates with respect to the model alone,

10 putting them closer to assimilated observations. ~~SEKF and~~ The SEKF and the EnSRF have a similar behaviour for LAI showing performances that are influenced by the vegetation type. For SSM, EnSRF estimates tend to be closer to observations than SEKF. ~~The impact of assimilating SSM and LAI is also assessed~~ comparison between the two data assimilation approaches is also carried out on unobserved soil moisture in the other layers of soil. Unobserved control variables are updated in the EnSRF through covariances and correlations sampled from the ensemble linking them to observed control variables. In our

15 context, a strong correlation between SSM and soil moisture in deeper soil layers is ~~exhibited~~ found, as expected, showing seasonal patterns that vary geographically. Moderate correlation and anti-correlations are also noticed between LAI and soil moisture ~~in spring, summer and autumn, their absolute value tending,~~ 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, showing that assimilating LAI can have an influence on soil moisture. Finally an independent evaluation of both assimilation approaches

20 is conducted using satellite estimates of evapotranspiration (ET) and gross primary production (GPP) as well as measures of river discharges from gauging stations. The EnSRF shows a systematic albeit moderate improvement ~~for evapotranspiration and GPP and a highly~~ of root-mean square differences and correlations for ET and GPP products, but its main improvement is observed on river discharges with a high positive impact on ~~river discharges, while the SEKF exhibits~~ Nash-Sutcliffe efficiency scores. Compared to the EnSRF, the SEKF displays a more contrasting performance.

1 Introduction

Land surface variables (LSVs) are key components of the Earth ~~system taking part, for example, in the~~ s water, vegetation and carbon cycles. Understanding their behaviour and simulating their evolution is a challenging task that has significant implications on various topics, from vegetation monitoring to weather prediction and climate change (Bonan, 2008; Dirmeyer et al., 2015; Schellekens et al., 2017). Land surface models (LSMs) play an important role in ~~achieving these goals~~ improving our knowledge of land surface processes and their interactions with the other components of the climate system such as the atmosphere. Forced by atmospheric data and coupled with river-routing models, they aim to simulate LSVs such as soil moisture (SM), biomass and leaf area index (LAI). However, LSMs are prone to errors owing to inaccurate ~~initialisation, misspecified forcing and parameters~~ initialization, misspecified parameters, flawed forcing or inadequate model physics. Another way to monitor LSVs is to use observations either from in situ networks or satellites. While in situ networks ~~are generally sparse~~ generally provide sparse spatial coverage, remote sensing provides a global coverage of LSVs at spatial resolutions ranging from ~~25 km x 25 km to 300 m x 300 m~~ the kilometre scale to the metre scale but at a daily frequency at best (Lettenmaier et al., 2015; Balsamo et al., 2018). ~~Also satellites do not observe every LSV such as~~ 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).

Combining observations with LSMs can overcome flaws ~~of in~~ both approaches. This is the objective of Land Data Assimilation Systems (LDASs). Many of them focus on assimilating observations related to surface soil moisture (SSM), either using passive microwave brightness temperatures, microwave backscatter coefficients or soil moisture retrievals obtained from the ~~two previous~~ mentioned satellite observations, to estimate soil moisture profiles (Lahoz and De Lannoy, 2014; Reichle et al., 2014; De Lannoy et al., 2016; Maggioni et al., 2017, and references therein). One popular approach has been the Simplified Extended Kalman Filter (SEKF). Introduced at Meteo-France by Mahfouf et al. (2009), it was initially designed for assimilating screen level observations to correct soil moisture estimates in the context of numerical weather prediction and is now involved in the operational systems of ~~e.g. both~~ the European Centre for Medium range Weather Forecast (ECMWF, Drusch et al., 2009; de Rosnay et al., 2013) and the UK Met Office. The SEKF has also been ~~rapidly extended to the~~ 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). Even though the SEKF approach has provided good results, it suffers from several limitations. It ~~involves in particular~~ relies on a climatological background error covariance matrix assuming uncorrelated variables between grid points and involves the computation of a Jacobian matrix ~~obtained by 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 ~~their number~~ the size of the control vector. That is why SEKF has been in competition with more flexible approaches, such as the Ensemble Kalman Filter (EnKF) (Reichle et al., 2002; Fairbairn et

al., 2015; Blyverket et al., 2019, among others) and particle filters (see e.g. Pan et al., 2008; Plaza et al., 2012; Zhang et al., 2017; Berg et al., 2019) for estimating soil moisture profiles. Those various approaches have been extensively compared in ~~that~~
60 ~~context~~ the context of the sole assimilation of soil moisture retrievals (Reichle et al., 2002; Sabater et al., 2007; Fairbairn et al., 2015).

LDASs are, however, not restricted to soil moisture. Recently, monitoring vegetation dynamics through LDASs has gained attention. ~~Defined~~ LAI is a key land biophysical variable, it is defined as half the total area of green elements of the canopy per unit horizontal ground area, ~~the LAI is a key land biophysical variable~~. One way to monitor LAI is to assimilate observations
65 already used for surface soil moisture and ~~link them to LAI. Both brightness temperature indirectly linked to LAI, such as~~ brightness temperature for low microwave frequencies (see e.g. Vreugdenhil et al., 2016) and radar backscatter coefficient (Lievens et al., 2017; Shamambo et al., 2019, among others) ~~are influenced by vegetation dynamics~~. This is the approach followed by Sawada et al. (2015) and Sawada (2018) who assimilate brightness temperatures using a particle filter to jointly estimate soil moisture profiles and LAI in the Coupled Land Vegetation LDAS (CLVLDAS).

Another way to ~~monitor LAI through LDASs is to assimilate observations of LAI~~ constrain LAI is through the assimilation of
70 direct LAI observations in LDASs. Satellite derived LAI products ~~have benefited~~ benefit from recent advances in remote sensing (Fang et al., 2013; Baret et al., 2013; Xiao et al., 2013) and datasets are now available at the global scale and at high resolution. While ~~assimilating LAI has been done very often with crop models to estimate crop yields~~ other studies have assimilated LAI
75 in crop models and at a more local scale (see e.g. Pauwels et al., 2007; Ines et al., 2013; Jin et al., 2018), such assimilation has been, to our knowledge, seldom performed by LDASs. Jarlan et al. (2008) and Sabater et al. (2008) have succeeded ~~introducing~~
~~such~~ in introducing such an approach in LDASs. The latter study has notably shown that assimilating jointly observations of SSM and LAI can improve the quality of root-zone SM estimates for one location in ~~South-West~~ southwest France. This work has been carried out with the CO₂-responsive version of the Interactions between Soil, Biosphere and Atmosphere (ISBA) LSM (Calvet et al., 1998, 2004; Gibelin et al., 2006) developed by ~~Meteo-France research centre (Centre National de Recherches~~
80 ~~Météorologiques, CNRM)~~ CNRM. This version of ISBA allows for the simulation of vegetation dynamics including biomass and LAI. ~~Stemming from~~ Building on that work, Albergel et al. (2010), Rüdiger et al. (2010) and Barbu et al. (2011) ~~have~~
introduced a SEKF assimilating jointly SSM and LAI and tested the approach on ~~a site~~. ~~This~~ the SMOSREX site located in
southwest France. Their study has been extended to a series of locations over France (Dewaele et al., 2017) and to France (Barbu et al., 2014; Fairbairn et al., 2017) leading to the development of LDAS-Monde (Albergel et al., 2017). ~~Available~~
85 ~~through~~ The LDAS-Monde is available through the CNRM modelling platform SURFEX (SURFace EXternalisée, Masson et al., 2013) ~~, LDAS-Monde and it~~ has been successfully applied to various parts of the globe: Europe and the Mediterranean basin (Albergel et al., 2017, 2019; Leroux et al., 2018), contiguous United States (Albergel et al., 2018b) and Burkina Fasso (Tall et al., 2019).

Lately other LDASs have started assimilating LAI using an EnKF assimilation approach. For example ~~Fox et al. (2018) has~~
90 , Fox et al. (2018) assimilated LAI and biomass in order to reconstruct the vegetation and carbon cycles for a site in Mexico, ~~and Ling et al. (2019) has~~ while Ling et al. (2019) compared various approaches for the assimilation of LAI at global scale. In addition Kumar et al. (2019) ~~has~~ assimilated LAI with an EnKF in the North American Land Data Assimilation System phase

2 (NLDAS-2) and studied its impact not only on vegetation but also soil moisture, ~~water and carbon cycles~~, those LSVs being updated indirectly through the model using the updated LAI. ~~Nevertheless these studies, contrary to LDAS-Monde, have not considered soil moisture estimation, i.e. no soil moisture variables were in the control vector~~ These studies did not update both SM and LAI, as we will do in this study.

This paper aims to develop an EnKF approach for the joint assimilation of LAI and SSM in the LDAS-Monde. ~~While the SEKF has been routinely employed in LDAS-Monde, 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), ~~has been proposed by Fairbairn et al. (2015) for the assimilation of SSM only in order to estimate soil moisture profiles. This study has also-~~ 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 southwest France. While performing better on synthetic experiments, the EnSRF provides results that are equivalent to the SEKF for real cases. Related to that work, Blyverket et al. (2019) ~~uses~~ used another deterministic EnKF to assimilate ~~SMAP~~-satellite derived SSM from the Soil Moisture Active Passive satellite (SMAP) over contiguous United States with the ISBA LSM focusing on soil moisture in the near surface while not updating root-zone soil moisture directly through data assimilation.

~~Building upon the work of Fairbairn et al. (2015), the~~ The present paper aims to (1) adapt the EnSRF to the joint assimilation of LAI and SSM within the LDAS-Monde, (2) study the impact of assimilating LAI and SSM ~~to-on~~ LSVs using an ensemble approach, and (3) compare the EnSRF with the routinely used SEKF ~~on-and~~ its ability to provide improved LSV estimates. To achieve ~~such these~~ goals, LDAS-Monde with EnSRF and SEKF is applied on the Euro-Mediterranean region (~~longitude from 11.5°W to 62.5°E, latitude from 25.0°N to 75.5°N, see Fig. 1 for the extent of the domain~~) ~~for for~~ a 10 year experiment (from 2008 to 2017):

- using the vegetation interactive ISBA-A-gs LSM (Calvet et al., 1998, 2004; Gibelin et al., 2006) with the multi-layer soil diffusion scheme from Decharme et al. (2011),
- coupled daily with CNRM version of the Total Runoff Integrating Pathways river routing model (CTRIP, Decharme et al., 2019) to simulate hydrological variables such as river discharges,
- forced by the latest ERA-5 atmospheric reanalysis from ECMWF (~~Hersbach and Dee, 2016~~) (Hersbach and Dee, 2016; Hersbach et al., 2016)
- and assimilating satellite derived Soil Water Index (SWI, as a proxy for SSM) and LAI products from the Copernicus Global Land Service (CGLS).

The performance of both DA approaches is assessed with (i) satellite-driven model estimates of land evapotranspiration (ET) from the Global Land Evaporation Amsterdam Model (GLEAM, Miralles et al., 2011; Martens et al., 2017), (ii) upscaled ground-based observations of gross primary production (GPP) from the FLUXCOM project (Tramontana et al., 2016; Jung et al., 2017), and (iii) river discharges from the Global Runoff Data Centre (GRDC). The paper is organised as follows: Sect. 2 details the various components involved in LDAS-Monde including the data assimilation schemes. Sect. 3 describes the

experimental setup and the different datasets used in the experiment ~~such as atmospheric forcing, assimilated observations or for evaluation.~~ or assimilated observations. Sect. 3 also details the datasets used to assess the performance of the EnSRF and the SEKF. The impact of the EnSRF on LSVs is then assessed in Sect. 4, including the comparison with the SEKF. Finally, the paper discusses the issues encountered during the experiment and provides prospects for future work in Sect. 5, before concluding in Sect. 6.

2 LDAS-Monde

LDAS-Monde (Albergel et al., 2017) is the offline, global scale and sequential data assimilation system dedicated to land surfaces developed by the Meteo-France research centre CNRM (Albergel et al., 2017). Embedded within the open-access SURFEX surface modelling platform (Masson et al., 2013, <https://www.umr-cnrm.fr/surfex/>), it consists of the ISBA land surface model coupled with the CTRIP river routing system and data assimilation ~~routines that assimilates.~~ Those routines assimilate routinely satellite-based products of SSM and LAI to analyse and update soil moisture and LAI modelled by ISBA. The most recent SURFEX_v8.1 implementation is used ~~to conduct in~~ our experiments. We quickly recall the main components of LDAS-Monde and subsequently detail the novel EnSRF approach for the joint assimilation of SSM and LAI. More information can be found in Albergel et al. (2017), see also <https://www.umr-cnrm.fr/spip.php?article1022&lang=en>.

2.1 ISBA land surface model

The ISBA (Noilhan and Planton, 1989; Noilhan and Mahfouf, 1996) LSM aims to simulate the evolution of LSVs such as soil moisture, soil heat or biomass (Noilhan and Planton, 1989; Noilhan and Mahfouf, 1996). We use in this paper the ~~soil ISBA~~ multilayer diffusion scheme ~~version (Boone et al., 2000; Decharme et al., 2011) involving a discretization of 14 layers of soil over 12 m depth to solve which solves~~ the mixed form of Richards equations (Richards, 1931) for water and the one-dimensional Fourier law for heat ~~.The (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 14 soil layers (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 12.0 m depth, see also Fig 1. of Decharme et al., 2013) was chosen to minimise errors coming from solving numerically the numerical approximation of the diffusion equations.

Regarding vegetation dynamics and interactions between the water and carbon cycles, we use the ISBA-A-gs configuration (Calvet et al., 1998, 2004; Gibelin et al., 2006). This CO₂-responsive version represents the relationship between the leaf-level net photosynthesis rate (A) and stomatal aperture (gs). Dynamics of vegetation variables such as LAI or biomass are induced by photosynthesis in response to atmospheric variations. LAI growing phase from a prescribed threshold (1.0 m².m⁻² for coniferous trees, 0.3 m².m⁻² for every other type of vegetation) results from an enhanced photosynthesis and CO₂ uptake. On the contrary, a deficit of photosynthesis leads to higher mortality rates and a decreased LAI. Leaf biomass is determined from LAI (and vice-versa) through dividing LAI by the specific leaf area (one of the ISBA parameters depending on the vegetation

type). For arctic regions, hydraulic and thermal soil properties are modified in order to include a dependency on soil organic carbon content (Decharme et al., 2016).

From a practical point of view, ISBA is run in this paper at a regular 0.25° spatial resolution. Each ISBA grid cell is divided into 12 generic patches: 9 representing different types of vegetations (deciduous forests, coniferous forests, evergreen forests, C3 crops, C4 crops, C4 irrigated crops, grasslands, tropical herbaceous and wetlands), and three others depicting bare soils, bare rocks and permanent snow or ice surfaces. Each patch covers a varying percentage of grid cells. Denoted $\alpha_{[p]}$ for patch p of a given grid cell, this percentage is also known as the patch fraction. Vegetation and soil parameters for each patch and grid cell of ISBA are derived from the ECOCLIMAP II land cover database (Faroux et al., 2013) that is fully integrated in SURFEX.jj

2.2 CTRIP river routing model

The ISBA LSM is coupled with CTRIP to simulate hydrological variables at continental scale. Based originally on the work of Oki and Sud (1998), CTRIP aims to convert simulated runoff into simulated river discharges. The model is fully described in the following papers: Decharme et al. (2010), Decharme et al. (2012), Vergnes and Decharme (2012), Vergnes et al. (2014) and Decharme et al. (2019).

CTRIP is available at a 0.5° spatial resolution. ~~Coupling~~ The coupling between ISBA and CTRIP occurs on a daily basis through the OASIS3-MCT coupler (Voltaire et al., 2017). ISBA provides updated runoff, drainage, ~~groudwater~~ groundwater and floodplain recharges to CTRIP, while the river routing model returns the water table depth or rise, floodplain fraction and flood potential infiltration to the LSM.

2.3 Data assimilation

LDAS-Monde is a sequential data assimilation system ~~working-on-with~~ a 24h assimilation window. Each cycle is divided in two steps: forecast and analysis. Quantities produced during the forecast step (analysis step) are denoted with a superscript f (superscript a). The state of the studied system is described by $\mathbf{x}_{[p]}$ the control vector that contains every prognostic variable of the ISBA LSM for a patch p and a given grid point. In this paper, we consider LAI and soil moisture from layer 2 (1-4 cm depth, SM2) to 7 (60-80 cm depth, SM7) in the control vector, soil moisture in layer 1 being driven mostly by atmospheric forcings (Draper et al., 2011; Barbu et al., 2014). As in many LDASs, LDAS-Monde perform DA for each grid point independently (no spatial covariances are considered).

The forecast step consists of propagating the state of the system from a time t to ~~24h later~~ Since patches and grid cells $t + 24h$ using ISBA. Patches in each ISBA grid cell do not interact between each other ~~in ISBA~~ , denoted by $\mathcal{M}_{[p]}$ for . This implies that, for a patch p , the forecast ~~step can be written as:~~ of $\mathbf{x}_{[p]}$, denoted by $\mathbf{x}_{[p]}^f(t + 24h)$, 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:

$$\mathbf{x}_{[p]}^f(t + \underline{24h24h}) = \mathcal{M}_{[p]} \left(\mathbf{x}_{[p]}^a(t) \right) \quad (1)$$

The analysis step then corrects forecast estimates by assimilating observations of LAI and SSM.

2.3.1 Simplified Extended Kalman Filter

190 LDAS-Monde uses routinely a Simplified Extended Kalman Filter for the analysis step (Mahfouf et al., 2009). Observations (SSM and/or LAI) \mathbf{y}^o are available at a grid cell level. To compute the analysis, we first need to calculate model equivalents of observations at the same grid cell and interpolate them 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 to obtain the model equivalent \mathbf{y}^f of the ISBA grid cell using a weighted average:

$$\mathbf{y}^f = \sum_{k=1}^{12} \alpha_{[k]} \mathbf{H} \mathbf{x}_{[k]}^f \quad (2)$$

\mathbf{H} denotes the linear operator selecting model equivalent from each patch (modelled LAI for observed LAI, modelled soil moisture in layer 2 for SSM).

200 ~~The~~ Then, the SEKF analysis step ~~then~~ 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 by a fixed with a fixed prescribed error covariance matrix ~~Band~~ uses as observation operator. The observation operator is the product of the model state evolution over the 24h window from t to $t + 24\text{h}$ and the conversion of the model state into observation equivalents. We further suppose that the fixed \mathbf{B} matrix is diagonal. This implies that there is no covariances between patches. 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 ~~pit~~ gives, we have:

$$\mathbf{x}_{[p]}^a = \mathbf{x}_{[p]}^f + \alpha_{[p]} \mathbf{B} (\mathbf{H} \mathbf{M}_{[p]})^T \mathbf{C}_{\text{SEKF}}^{-1} (\mathbf{y}^o - \mathbf{y}^f) \quad (3)$$

and

$$\mathbf{C}_{\text{SEKF}} = \sum_{k=1}^{12} \alpha_{[k]}^2 (\mathbf{H} \mathbf{M}_{[k]}) \mathbf{B} (\mathbf{H} \mathbf{M}_{[k]})^T + \mathbf{R} \quad (4)$$

210 with \mathbf{R} the observation error covariance matrix and $\mathbf{M}_{[p]}$ the Jacobian matrix of $\mathcal{M}_{[p]}$. In practice, columns of $\mathbf{M}_{[p]}$ are calculated by finite differences using perturbed model runs. For each component x_j of the control vector and its perturbation δx_j , the j -th column of $\mathbf{M}_{[p]}$ can be written as:

$$[\mathbf{M}_{[p]}]_j = \frac{\partial x^f(t + 24h)}{\partial x_j} \approx \frac{\mathcal{M}_{[p]}(\mathbf{x}^a(t) + \delta x_j) - \mathbf{x}^f(t + 24h)}{\delta x_j} \quad (5)$$

Details on how to obtain eq. (3) and (4) can be found in supplementary material.

215 2.3.2 Ensemble Square Root Filter

We adapt the EnSRF from Whitaker and Hamill (2002) to the context of LDAS-Monde following the work of Fairbairn et al. (2015). The EnSRF is an EnKF-based approach in which the state of a system and associated uncertainties are described by

an ensemble of N_e control vectors $\{\mathbf{x}_{[p]}^{(i)}, i = 1, \dots, N_e\}$ for patch p of a given grid cell. The EnKF ~~as a Monte Carlo approach~~ approximates the classical Kalman Filter equations using the ensemble mean

$$220 \quad \bar{\mathbf{x}}_{[p]} = \sum_{i=1}^{N_e} \mathbf{x}_{[p]}^{(i)} \quad (6)$$

to describe the state of the system and the ensemble covariance matrix

$$\mathbf{P}_{[p]} = \frac{1}{N_e - 1} \mathbf{X}_{[p]} \mathbf{X}_{[p]}^T \quad (7)$$

~~with where~~ $\mathbf{X}_{[p]} = [\mathbf{x}_{[p]}^{(1)} - \bar{\mathbf{x}}_{[p]}, \dots, \mathbf{x}_{[p]}^{(N_e)} - \bar{\mathbf{x}}_{[p]}]$ is the ensemble perturbation matrix, to describe the uncertainties of the estimation.

225 ~~The forecast steps simple~~ In the forecast step, we propagate as in Eq. (1) each ensemble member from ~~a time to 24h later~~ time t to $t + 24h$ using the ISBA LSM. The analysis step then corrects the ensemble mean and the ensemble perturbation matrix by assimilating observations. To that end, we first calculate the model equivalent of the observations by aggregating the mean of the forecast ensemble over all the patches

$$\mathbf{y}^f = \sum_{k=1}^{12} \alpha_{[k]} \mathbf{H} \bar{\mathbf{x}}_{[k]}^f \quad (8)$$

230 ~~Then the EnSRF analysis creates an analysed ensemble whose~~ The analysis step then updates the ensemble whose analysed mean and covariance matrix matches exactly the analysis of the Kalman Filter when the observation operator is linear.

We choose to neglect ~~here the~~ ensemble covariances between patches in the analysis step of the EnSRF. This ~~assumption~~ assumption is in line with the ~~successful methodology of the SEKF and SEKF method and it~~ ensures a fair comparison between ~~both approaches. It also allows control variables from each patch to be updated independently. This places our approach~~ in a similar context as others EnKFs involved in LDASs where ensembles with no more than the two approaches. The approach outlined here is in line with other studies (Fairbairn et al., 2015; Carrera et al., 2015) showing that 1D-EnKFs can achieve promising results with around 20 members have achieved satisfactory results. For example, Fairbairn et al. (2015) and Carrera et al. (2015) have shown that sampling errors from the finite ensemble size were not significant for ensembles greater than 20 ensemble members.

240 Following this assumption, for a given patch ~~p~~ p , the analysed mean and perturbation matrix are given by the following equations:

$$\bar{\mathbf{x}}_{[p]}^a = \bar{\mathbf{x}}_{[p]}^f + \alpha_{[p]} \mathbf{P}_{[p]}^f \mathbf{H}^T \mathbf{C}_{\text{EnSRF}}^{-1} (\mathbf{y}^o - \mathbf{y}^f) \quad (9)$$

and

$$\mathbf{X}_{[p]}^a = (\mathbf{I} - \alpha_{[p]} \tilde{\mathbf{K}}_{[p]} \mathbf{H}) \mathbf{X}_{[p]}^f \quad (10)$$

245 with

$$\mathbf{C}_{\text{EnSRF}} = \sum_{k=1}^{12} \alpha_{[k]}^2 \mathbf{H} \mathbf{P}_{[k]}^f \mathbf{H}^T + \mathbf{R} \quad (11)$$

and

$$\tilde{\mathbf{K}}_{[p]} = \alpha_{[p]} \mathbf{P}_{[p]}^f \mathbf{H}^T (\mathbf{C}_{\text{EnSRF}}^T)^{-1/2} (\mathbf{C}_{\text{EnSRF}}^{1/2} + \mathbf{R}^{1/2})^{-1} \quad (12)$$

Such an approach, contrary to the SEKF, updates the state covariance matrix that will evolve in time. This ensures that
250 ~~statistics of the estimates keep information from past observations~~ information from the analysis is stored in the ensemble and is propagated forward in time.

3 Experimental setup and data sets

~~The effect of SEKF and EnSRF on LSVs is compared over the Euro-Mediterranean region (longitude from 11.5°W to 62.5°E, latitude from 25.0°N to 75.5°N) at a 0.25° spatial resolution during the decade 2008–2017.~~ We detail in the following
255 subsections the atmospheric forcing, the assimilated observations, the validation data sets employed in this ~~comparison paper~~ before detailing the experimental setup.

3.1 Atmospheric forcing

The ISBA LSM is forced with the ~~ERA-5 atmospheric reanalysis (Hersbach and Dee, 2016)~~ ERA5 atmospheric reanalysis (Hersbach and Dee, 2016; Hersbach et al., 2019) developed by ECMWF. ERA-5 reanalysis is available with an hourly fre-
260 quency at a ~~31-km spatial horizontal~~ 31 km horizontal spatial resolution. To be used, surface atmospheric variables such as air temperature, surface pressure, solid and liquid precipitations, incoming shortwave and longwave radiations ~~or~~ and wind speed are interpolated to ISBA-0.25° spatial resolution with a using bilinear interpolation. ~~Albergel et al. (2018a, b) have shown that forcing ISBA with ERA-5 compared to ECMWF previous ERA-Interim atmospheric reanalysis improves the quality~~ Replacing ECMWF’s atmospheric ERA-Interim reanalysis by ERA5 has been shown beneficial in the context of LSVs reanalyses with
265 LDASs (Albergel et al., 2018a, b).

3.2 Observations for assimilation

In this paper we assimilate observations from the SWI-001 and GEOV1 LAI data sets, both being distributed by the Copernicus Global Land Service. These satellite-derived products have ~~been already~~ already been successfully assimilated in the LDAS-Monde (e.g. Leroux et al., 2018; Albergel et al., 2019).

270 The SWI-001 product consists of Soil Water Indices (SWI) obtained through a recursive exponential filter (Albergel et al., 2008) using backscatter observations from the ASCAT C-band radar (Wagner et al., 1999; Bartalis et al., 2007). A one-day time scale is used in the recursive filter in order to ~~measure~~ estimate the wetness of the first centimetres of the soil. This product is available daily at a 0.1° spatial resolution. The raw SWI-001 averaged over the 2008-2017 period can be seen in Figure 1 (a).

~~To be assimilated~~ Prior to the assimilation, the SWI-001 product needs to be rescaled to the model climatology to avoid
275 introducing any bias in the LDAS system (Reichle and Koster, 2004; Drusch et al., 2005). We apply a linear rescaling to SWI-001 to match the observation mean and variance to the mean and variance of the modelled soil moisture in the second layer

of soil (1–4 cm). Introduced by Scipal et al. (2008), this rescaling ~~also known as~~ 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). Draper et al. (2009) and Barbu et al. (2014) have highlighted the importance of allowing seasonal variability in the
280 rescaling.

The GEOLAND2 Version 1 (GEOV1) LAI product is obtained through a neural network algorithm (Baret et al., 2013) transforming observations of reflectance from SPOT-VGT and PROBA-V satellites into LAI. This dataset is available every 10 days ~~at best at a one kilometre spatial resolution~~ with the finest spatial resolution being 1 km. The GEOV1 LAI averaged over the 2008–2017 period can be seen in Figure 1 (b).

285 Following Barbu et al. (2014), both observation datasets are interpolated on the model grid (0.25° spatial resolution) where and when at least half of observation grid points are available. As in previous LDAS-Monde studies, we use a 24h assimilation window and observations are assimilated at 9:00 UTC.

3.3 Validation data sets

We consider independent datasets of evapotranspiration (ET), gross primary production (GPP) and river discharges to assess
290 the validity of our approach and measure the influence of the EnSRF on the improvement of LSV reanalyses.

Satellite-derived estimates of ~~evapotranspiration~~ ET come from the GLEAM v3.3b product (Miralles et al., 2011; Martens et al., 2017). Daily estimates available for the period 1980 – 2018 at a 0.25° spatial resolution are fully driven by satellite observations and, as such, are independent from LDAS-Monde estimates. Figure 1 (c) displays GLEAM ~~evapotranspiration~~ ET averaged over the period 2008–2017 considered for validation in this paper.

295 Observations of ~~Gross Primary Production (GPP)~~ GPP are derived from the FLUXCOM project. This dataset is obtained by merging upscaled measurements from ~~eddy-eddy-covariance flux~~ towers and satellite observations using machine learning. More details can be found in Tramontana et al. (2016) and Jung et al. (2017). ~~They~~ The FLUXCOM data are available at a 0.5° spatial resolution on a monthly basis for the period 1982 – 2013. Figure 1 (d) shows FLUXCOM GPP averaged over the period 2008–2013 considered for validation in this paper.

300 River ~~discharges obtained with CTRIP~~ discharge output from the CTRIP river routing model are compared to daily stream-flow data obtained from the Global Runoff Data Centre (<https://www.bafg.de/GRDC>). Due to the low resolution of CTRIP (0.5° spatial resolution), we only consider data for sub-basins with rather large drainage areas (greater than 10 000 km²) with a long enough time series (4 complete years or more over 2008 – 2017).

3.4 Experimental setup

305 To assess the impact of EnSRF on LSV reanalyses and compare its ~~efficiency~~ skill with the routinely used SEFK, we have run LDAS-Monde over the Euro-Mediterranean region ~~for the period 2008–~~ (longitude from 11.5°W to 62.5°E, latitude from 25.0°N to 75.5°N) at a 0.25° spatial resolution during the decade 2008 – 2017 for three different configurations: one model run without assimilation (i.e. open loop), one using the SEKF and another one using the EnSRF with a 20-members ensemble. This size of the ensemble is consistent with Fairbairn et al. (2015) and Carrera et al. (2015). All three configurations start from

310 the same initial state obtained after spinning-up ISBA-CTrip twenty times over 2008. This provides an initial state for which the system has reached equilibrium.

For the SEKF configuration, the Jacobian matrix Eq. (5) is obtained by finite differences using perturbed model runs. Following Draper et al. (2009) and subsequent studies, perturbations are taken proportional to the dynamic range (difference between the volumetric field capacity w_{fc} and the wilting point w_{wilt}) for soil moisture variable. In practice ~~they are set to~~
315 ~~$10^{-4} \times (w_{fc} - w_{wilt})$~~ . ~~For LAI, perturbations,~~ perturbations for SM are set to ~~following Rüdiger et al. (2010)~~ $10^{-4} \times (w_{fc} - w_{wilt})$. Regarding the fixed background error covariance, we prescribe a mean volumetric standard deviation (SD) of $0.04 \text{ m}^3 \cdot \text{m}^{-3}$ for ~~soil moisture SM~~ in the second layer and $0.02 \text{ m}^3 \cdot \text{m}^{-3}$ for ~~soil moisture SM~~ in deeper layers, both are then scaled by the dynamic range ~~of SM~~. For LAI, perturbations are set to a fraction (0.001) of the modelled LAI following Rüdiger et al. (2010). LAI background error SD is set to 20% of the LAI value for modelled values above $2.0 \text{ m}^2 \cdot \text{m}^{-2}$ and
320 to a constant $0.4 \text{ m}^2 \cdot \text{m}^{-2}$ for modelled values below $2.0 \text{ m}^2 \cdot \text{m}^{-2}$. This SEKF configuration is the same as the ~~successful~~ one detailed in Albergel et al. (2017).

About the EnSRF configuration, the initial ensemble is obtained by perturbing the initial state using ~~Gaussian perturbations~~ perturbations sampled from a multivariate Gaussian distribution with a zero-mean and using ~~the prescribed B for the covariance matrix~~ covariance matrix used in the SEKF as the covariance matrix of that multivariate Gaussian distribution. Ensemble
325 Kalman Filters tend to underestimate variances and ensembles spreads. This brings about an artificially ~~shrunk small~~ spread leading ultimately to filter divergence if not counteracted. Hamill and Whitaker (2005) has shown that adding random perturbations to each ensemble member (additive inflation) at the start of each assimilation cycle can overcome this issue. It can also be used to represent model error. As in Fairbairn et al. (2015) we use time-correlated model errors using a first-order auto-regressive model. We prescribe an associated ~~white Gaussian~~ noise with zero mean and a SD of $\lambda(w_{fc} - w_{wilt})$ for ~~soil moisture (SM) SM~~, with $\lambda = 0.5$ for SM in layer 2 (1-4 cm depth), 0.2 for SM in layer 3 (4-10 cm depth), 0.05 for SM in layer 4 (10-20 cm depth) and 0.02 for SM in deeper layers. These values are in line with Fairbairn et al. (2015). For LAI, we prescribe a ~~white Gaussian~~ noise with zero mean and a SD of $0.5 \text{ m}^2 \cdot \text{m}^{-2}$. We also fix the time correlation to 1 day for SM in the second layer and 3 days for SM in deeper layers. This is similar to the work of Reichle et al. (2002) and Mahfouf (2007). For LAI, a rather small 1-day time correlation has to be used in order to avoid a collapse of the ensemble during the winter season due to
335 the LAI threshold in ISBA.

For both SEKF and EnSRF configurations, we follow previous LDAS-Monde studies and set ~~soil moisture SSM~~ observational errors to $0.05 \text{ m}^3 \cdot \text{m}^{-3}$ scaled to the dynamic range and LAI observational errors to 20% of the observed LAI values (see e.g Albergel et al., 2017; Leroux et al., 2018; Tall et al., 2019).

3.5 Evaluation strategy

340 As a ~~sanity~~ check, we first verify that EnSRF estimates of SSM and LAI are closer to observations than their ~~model-free run~~ open loop counterparts. We also compare the impact of EnSRF and SEKF on ~~these two LSVs~~ SM in layer 2 (1-4 cm depth, SM2) and LAI. This is achieved using scores such as biases, correlation coefficients (R), root mean square differences (RMSD) and normalised root mean square differences (nRMSD, RMSD divided by the averaged value of the studied variable).

The impact of assimilation on unobserved control variables (SM in deeper layers) is then assessed using daily analysis
 345 increment. Moreover, we study the evolution of the ensemble correlations between unobserved and observed variables in the
 EnSRF configuration. They drive (as Jacobian values in the SEKF configuration) the influence of observations on unobserved
 control variables. We focus on SM in layer 4 (10-20 cm depth, SM4) and layer 6 (40-60 cm depth, SM6) as SM in layer 3 (4-10
 cm depth) exhibits the same behaviour as SM4 and soil moisture in layer 5 (20-40 cm depth) and layer 7 (60-80 cm depth)
 have ~~a similar response to the same behaviour as~~ SM6 (not shown).

350 Potential improvements ~~on in~~ EnSRF and SEKF estimates ~~for evapotranspiration of ET~~ and GPP are measured using the
 same metrics as for SSM and LAI.

Finally the influence on river discharges for both DA approaches is measured by the Nash-Sutcliffe efficiency (NSE) score:

$$\text{NSE} = 1 - \frac{\sum_{t=1}^T (Q_t^s - Q_t^o)^2}{\sum_{t=1}^T (Q_t^o - \bar{Q}^o)^2} \quad (13)$$

with Q_t^s the simulated or analysed river discharge at time t , Q_t^o the observed river discharge at the same time and \bar{Q}^o the
 355 observed averaged river discharge. The NSE is a quantity between $-\infty$ and 1. A NSE value of 1 means that the model/analysis
 matches perfectly observations. A NSE value of 0 means that the model/analysis has the same ~~accuracy-NSE~~ as the observed
 averaged river discharge. Improvements or degradations caused by the SEKF or the EnSRF compared to the ~~model-run-open~~
~~loop~~ is measured with the Normalised Information Contribution Index (NIC):

$$\text{NIC}_{\text{NSE}} = 100 \times \frac{\text{NSE}_{\text{analysis}} - \text{NSE}_{\text{model}}}{1 - \text{NSE}_{\text{model}}} \quad (14)$$

360 4 Results

4.1 Impact of assimilation on LAI

Figure 2 displays ~~for the model-run~~the open loop, SEKF and EnSRF analyses and ~~for observations,~~observed LAI 10-days time
 series ~~of LAI~~ averaged over Europe and the Mediterranean basin and spanning the period 2008 – 2017. It shows that the model
 simulation underestimates LAI compared to observations during winter and summer. The growing phase of vegetation occurs
 365 at a slower pace with averaged LAI reaching its maximum early August instead of late June – early July for observations. The
 senescence phase subsequently takes place later in the autumn compared to observations. Both ~~data-assimilation-DA~~ systems
 correct efficiently model simulations for that latter phase ~~with EnSRF estimates getting closer to observations than SEKF ones~~.
 However, both SEKF and EnSRF fail to compensate the slower LAI dynamics of the model during spring. ~~Nevertheless, both~~
~~approaches reduce RMSD~~ This is in compliance with what Albergel et al. (2017) and Leroux et al. (2018) have observed over
 370 the Euro-Mediterranean region. 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. During the senescence, LAI dynamics is driven by the rate of mortality, thus making DA more efficient.

As expected, both DA approaches produce estimates that are closer to the assimilated LAI observations than their open loop counterpart. RMSDs are reduced from $0.880 \text{ m}^2 \cdot \text{m}^{-2}$ for the ~~model open loop~~ to $0.671 \text{ m}^2 \cdot \text{m}^{-2}$ for SEKF and $0.694 \text{ m}^2 \cdot \text{m}^{-2}$ for EnSRF and increase correlations with observations (Correlations with assimilated observations are increased from 0.593 for the model to 0.732 for SEKF and 0.723 for EnSRF). A full summary of statistics for LAI can be found in Table 1. We also note that the maximum LAI for EnSRF is smaller than the model or the SEKF maxima. This is due in part to a systematic negative bias introduced by EnSRF model perturbations leading to an averaged bias of -0.201 . The averaged bias for the open loop is rather small with $-0.020 \text{ m}^2 \cdot \text{m}^{-2}$ compared to a bias of -0.020 , but, it hides a negative bias during winter and summer that is compensated by a positive bias during autumn. DA approaches mostly correct the positive autumnal bias, thus, making the averaged bias more negative, $-0.116 \text{ m}^2 \cdot \text{m}^{-2}$ for the ~~model~~ and -0.116 SEKF and $-0.201 \text{ m}^2 \cdot \text{m}^{-2}$ for the SEKF-EnSRF. The bias is more negative for the EnSRF than for the SEKF for every season. This is due in part to a systematic negative bias introduced by the EnSRF model perturbations. This bias can sometimes lead to degraded performances. As pointed out by Fairbairn et al. (2015), model perturbations can introduce a bias into the system in LDASs.

Figure 3 shows nRMSD calculated over 2008 – 2017 for ~~model outputs the open loop~~ (a) and the difference between nRMSD for the ~~model open loop~~ and the estimates produced with SEKF (b) and EnSRF (c). On average nRMSD is reduced from 0.57 (~~model open loop~~) to 0.42 (EnSRF) and 0.40 (SEKF). Both assimilation approaches display the same geographical patterns reducing significantly nRMSD over most parts of the Euro-Mediterranean region (in blue in Figure 3). For example, roughly 20% of the domain has a nRMSD reduced by 0.25. We note that largest nRMSD reductions occur in places where nRMSD are large. The main differences between the two methods occur in ~~Scandinavia, around the arctic circle, Ireland and Western Great Britain~~ Ireland, western Great Britain, northwest Spain, the Alps, Scandinavia and arctic regions, where the SEKF shows greater positive impact than EnSRF, the latter even providing slightly degraded estimates compared to the ~~model run open loop~~ for 3% of the total domain (in red in Figure 3 (c)).

The geographical patterns identified in Figure 3 can be explained in part by the type of vegetation covering grid cells. We investigate the impact of DA for each of the four main vegetation types encountered in the Euro-Mediterranean region: deciduous forests, coniferous forests, C3 crops and grasslands. To that end, we consider only grid cells (g.c.) in which at least 50% of their surface is covered by one of these vegetation types. Figure 4 displays the spatial distribution of those grid cells: 1589 g.c. for deciduous forests (5.7% of the domain), 4223 g.c. for coniferous forests (15.2%), 1672 g.c. for C3 crops (6.0%) and 1725 g.c. for grasslands (6.2%).

We calculate the averaged seasonal RMSD for ~~model outputs the open loop~~, SEKF and EnSRF analyses for the entire domain (Figure 5 (a)) and for each dominant vegetation type (Figure 5, (b)-(e)). The biggest impact of assimilating LAI occurs in autumn for deciduous forests (Fig. 5 (e)). For example, RMSD is reduced from $2.69 \text{ m}^2 \cdot \text{m}^{-2}$ for the ~~model open loop~~ to $1.72 \text{ m}^2 \cdot \text{m}^{-2}$ for the SEKF and $1.45 \text{ m}^2 \cdot \text{m}^{-2}$ for the EnSRF. For C3 crops (Fig. 5 (c)) both assimilation approaches reduce RMSD in a similar manner, the largest decrease happening between August and October. The SEKF and the EnSRF offer contrasting performances in the case of grasslands (Fig. 5 (d)) as RMSDs are decreased by $0.18 \text{ m}^2 \cdot \text{m}^{-2}$ from ~~model the open loop~~ to SEKF estimates but by $0.09 \text{ m}^2 \cdot \text{m}^{-2}$ for EnSRF estimates. Strongest Largest RMSD reductions occur in for both cases in April and September. This explains the reduced performance of the EnSRF compared to the SEKF over the grasslands-dominated

Ireland, western Great Britain and arctic regions. For coniferous trees (Fig. 5 (b)), ~~neither SEKF nor EnSRF has an~~ SEKF has a small positive impact on RMSDs ~~and the EnSRF a slightly negative impact~~. This explains the rather poor performance of the EnSRF over Scandinavia. This also explains what happens in northwest Spain and in the Alps. While not being dominated by one type of vegetation, coniferous trees and grasslands, the two types for which the EnSRF performs poorly, represent more than 70 % of the vegetation in those places.

The scale of reduction in RMSD for EnSRF analyses is directly connected to estimated variances and standard deviations from the ensemble. The bigger the ensemble variances are, the larger are the weight of observations in the DA system. 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. Ensemble standard deviations are clearly larger in summer than in winter peaking in July for c3 crops at $0.22 \text{ m}^2.\text{m}^{-2}$, in August for grasslands at $0.14 \text{ m}^2.\text{m}^{-2}$ and in September for coniferous forests at $0.07 \text{ m}^2.\text{m}^{-2}$. The maximum standard deviation is observed for deciduous forests and reaches $0.35 \text{ m}^2.\text{m}^{-2}$ also in September.

Standard deviations in the EnSRF relies heavily on the model perturbations. In the case of LAI, model perturbations applied to LAI in every vegetation patch are sampled from the same distribution. However, the behaviour of ensemble standard deviations varies greatly seasonally and for each type of vegetation. Standard deviations for coniferous trees are so low it leads to almost no impact of DA. Such behaviour can be explained by two caveats: first, ISBA modelled LAI evolves over a prescribed threshold ($1 \text{ m}^2.\text{m}^{-2}$ for coniferous forests, $0.3 \text{ m}^2.\text{m}^{-2}$ for other vegetation patches). Model perturbations can lead to LAI values below this threshold. To avoid model issues, estimated LAI is reset to that threshold when this is the case. It can lead to an artificially reduced ensemble standard deviation when modelled LAI is close to that threshold as in winter. Secondly, since LAI dynamics are smooth, reduced ensemble standard deviations due to the winter season still have an impact in spring through the ISBA LSM. An approach for model errors tailored for each vegetation patch could overcome the observed caveats.

4.2 Impact of assimilation on SSM

This section studies the impact of assimilating jointly LAI and SSM on estimated SSM. We firstly recall that observed SSM is derived from the SWI-001 satellite product and is matched to the model climatology of soil moisture in the second layer of soil (1-4 cm depth) using a seasonal ~~CDF-matching~~ linear rescaling. This means that assimilating observed SSM mostly corrects the short-term variability of estimated SSM and does not modify its climatological seasonal cycle. Results from either SEKF or EnSRF experiments are in line with this statement. For example, the bias between observed and estimated SSM remains, on average over 2008-2017, below $0.002 \text{ m}^3.\text{m}^{-3}$ all over the domain (see also Table 1 all the averaged scores with observed SSM).

Figure 7 displays RMSD calculated over 2008 – 2017 for ~~model-outputs~~ the open loop (a) and the difference between RMSD for the ~~model-open loop~~ and the estimates produced with SEKF (b) and EnSRF (c). On average, RMSD is reduced from $0.035 \text{ m}^3.\text{m}^{-3}$ (~~model-open loop~~) to $0.032 \text{ m}^3.\text{m}^{-3}$ (SEKF) and $0.027 \text{ m}^3.\text{m}^{-3}$ (EnSRF). ~~Model-RMSD~~ RMSD for the open loop tends to be generally larger in wetter places than in drier places with the exception of ~~South-East~~ southeast Spain and parts of ~~Northern~~ northern Africa where RMSDs can be larger than $0.050 \text{ m}^3.\text{m}^{-3}$. Both assimilation approaches reduce significantly

RMSD in many places over the domain (in blue in Figure 7 (b-c)). The main reduction occurs for both approaches in the southern part of the Euro-Mediterranean region where grid cells consists of bare soil and bare rocks. In those places, vegetation is sparse, and SSM is the main source of information in assimilated observations making its impact more straightforward. We also notice that the EnSRF tends to systematically produce estimates that are closer to observations than SEKF estimates. This is due to the model perturbations for the EnSRF and the prescribed background error covariance matrix in the SEKF. The prescribed model error for the EnSRF leads to ensembles with a bigger standard deviation than the one prescribed in the SEKF for SSM. This leads to a bigger weight to SSM observations in the EnSRF than in the SEKF, thus, making EnSRF estimates closer to SSM observations than SEKF estimates.

Assimilation also improves correlations with observed SSM from 0.544 for ~~model-outputs-the open loop~~ in average to 0.652 for the SEKF and 0.760 for the EnSRF. Figure 8 illustrates correlations for ~~model-outputs-the open loop~~ (a) and difference between correlations for the ~~model-open loop~~ and SEKF (b) and EnSRF (c) outputs. From correlation results, similar conclusions are drawn as from RMSDs. In particular the main improvement occurs in ~~Northern-northern~~ Africa for both approaches. Finally we observe negative correlations between ~~model-outputs-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. Further studies of such aspects are beyond the scope of this paper.

4.3 Correlations between observed and unobserved control variables

Examining Jacobians in the SEKF has provided interesting insights on the sensitivity of SSM and LAI on soil moisture in deeper layers (see e.g. Albergel et al., 2017, covering the Euro-Mediterranean region between 2000 and 2012). In the EnSRF, the ~~role-of-Jacobian-is devolved-to~~ Jacobian is replaced by correlations sampled from the ensemble covariance matrix. Figure 9 shows maps of correlations between soil moisture in layer 2 (1-4 cm depth, SM2 used as a proxy for SSM) and SM in layer 4 (10-20 cm depth, SM4) and layer 6 (40-60 cm depth, SM6) and correlations between LAI and SM2, SM4 and SM6. Correlations are averaged by season (December-January-February, March-April-May, June-July-August and September-October-November) over the whole period 2008 – 2017.

The first two rows of Figure 9 show the seasonal evolution of correlations between SM2 and SM4 and SM6. SM4 is highly correlated to SM2 (in blue), R being above 0.5 for most places of the domain for each season ~~and correlations with SM2~~. SM6 is also highly correlated to SM2 but to a lesser ~~extend extent~~ meaning that correlations with SSM decrease in absolute value when we reach deeper soil layers. We also notice seasonal tendencies. For example, correlations with SM2 tend to be larger in ~~Western-Europe-during-Spring~~ western Europe during spring while they reach their maximum during ~~Summer-summer~~ in Scandinavia. Negative correlations with SM2 (between -0.35 and -0.20) tend to appear during ~~Winter-winter~~ over Russia. It means that in those areas in winter, there is less liquid water in the surface when there is more liquid water in deeper layers. This is linked to snow and freezing as we only compare liquid soil moisture from the different layers of soil. We further notice

that SM2 and SM6 are uncorrelated in ~~Summer~~ summer over Spain and ~~Northern Africa~~. ~~Finally we remark that in northern Africa. This decorrelation between surface and root-zone soil moisture occurs during very dry conditions such as occurred in Spain and northern Africa during summer. The same phenomenon appears in~~ very arid places such as in Sahara SM2 is not correlated to soil moisture in deeper layers (either SM4 and SM6) for each season. This implies that assimilating SSM in those
480 areas will not modify soil moisture in deeper layers as we will show in the next section.

The last three rows of Figure 9 show the seasonal evolution of correlations between LAI and soil moisture in layers 2, 4 and 6. Soil moisture tends to be less correlated on average to LAI than to SSM nevertheless the values reached are relatively large (between -0.5 and 0.5). It means that assimilating LAI has an impact on estimated soil moisture. In detail, correlations between LAI and SM6 are larger in absolute value than with SM4 and with SM2 meaning that LAI is more correlated to
485 root-zone soil moisture than with SSM. We also observe seasonal geographical patterns. Positive correlations tend to appear in ~~Summer in Northern~~ summer in northern Europe where deciduous and coniferous forests are dominant meaning more water in the soil leads to a greater LAI. On the contrary in ~~Spring and Summer~~ spring and summer, negative correlations appear around the Mediterranean basin. This means a higher LAI leads to a reduced soil moisture due to plant transpiration in part. Barbu et al. (2011) has already highlighted this kind of behaviour for Jacobians for grassland places in southwest France.

490 Overall conclusions drawn from correlations are in accordance with those derived from the analysis of SEKF Jacobians drawn in Albergel et al. (2017) over the Euro-Mediterranean region and Tall et al. (2019) over Burkina-Fasso. Nevertheless, we note that correlation can be influenced by the way we apply model error. Another type of model error, perturbing for example atmospheric forcing, may have led to different characteristics of the covariances between the ISBA variables.

4.4 Impact of assimilation on soil moisture in deeper layers

495 Figure 10 displays soil moisture for layers 4 and 6 averaged over 2008 – 2017 from the ~~model~~ open loop (left) and the averaged difference with SEKF estimates (central panels) and EnSRF estimates (right). We observe that the SEKF ~~has the same and the EnSRF have overall similar~~ averaged SM4 as the ~~model~~. ~~Nevertheless we discern seasonal tendencies. Figure 11 shows analysis increments for SM4 and SM6 for SEKF (top row) and EnSRF (bottom row) for May, July and September. We see that increments on SM4 tend to be negative in May and September in most parts of the domain and positive in~~
500 ~~July in Northern Europe for SEKF. EnSRF estimates for SM4 tends to be similar to SEKF estimates.~~ open loop. The main difference occurs in ~~Northern~~ northern Africa and in the Arabian peninsula where the soil is estimated wetter than in SEKF with a difference reaching $0.02 \text{ m}^3 \cdot \text{m}^{-3}$. This disparity over arid ~~zones~~ regions is due solely to a wet bias introduced by model error ~~as assimilating SSM in those places has no influence due to negligible correlations between SSM and SM4. In those places, the EnSRF cannot correct this bias using observations of SSM or LAI. In other places, the EnSRF can correct the bias~~
505 ~~potentially introduced by the model perturbations to unobserved control variables through the help of correlations.~~ We also identify greater EnSRF SM4 estimates over places such as Poland and Spain but the difference ~~, being with the open loop is~~ always below $0.01 \text{ m}^3 \cdot \text{m}^{-3}$, ~~comes from more positive increments during the summer period (as shown for July in Figure 11). Except in arid areas, SM4 estimates and analysis increments for SEKF and EnSRF tend to be similar, thus, making our SM4 estimates less dependant on the data assimilation method.~~

510 Regarding SM6 estimates, both SEKF and EnSRF produce a drier soil layer than the model for most of the domain as shown 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~~ Also, EnSRF SM6 is wetter for regions where bare soil dominates in northern Africa than SM6 obtained with the SEKF or the open loop. Again this is due solely to the wet bias introduced by model soil moisture perturbations as SM6 and SM2 are uncorrelated in those places. Then, we can observe for SM6 an abrupt change in the arctic region for both SEKF and EnSRF compared to the open loop. This difference is due to modified hydraulic and thermal soil properties in ISBA for arctic regions. This modification has been implemented by (Decharme et al., 2016) in order to include a dependency on soil organic carbon content.

Figure 11 shows analysis increments in SM4 for SEKF (top row) and EnSRF (bottom row) for May, July and September. We see that increments in SM4 tend to be negative in May and September in most parts of the domain and positive in July, particularly in northern Europe for SEKF. The SM4 analyses increments for the SEKF and EnSRF tend to be similar, except for arid regions. This makes the SM4 estimates less dependant on the data assimilation method.

About analysis increments for SM6, SEKF increments are close to zero. ~~For EnSRF for every season (not shown). This implies that the drier estimates are solely due to the joint effect of the ISBA LSM and the updated LAI and soil moisture near the surface. For the EnSRF, this joint effect also occurs but, cycling is also responsible to this drying but~~ analysis increments are not negligible ($-0.01 \text{ m}^3 \cdot \text{m}^{-3}$ for biggest values) ~~and~~. The EnSRF SM6 analysis increments compensate the wet bias from model error in SM6 (not shown). ~~As for SM4, and lead to similar SM6 EnSRF estimates are larger than SEKF and model estimates in Northern Africa, but this time only for places where bare soil dominates as for places where bare rocks dominates the soil dries due to cycling. Again assimilation does not modify directly estimates as correlations are null. estimates as the SEKF in most places as shown previously.~~

530 Overall SEKF and EnSRF provide similar estimates for soil moisture in deeper layers for most places but not necessarily through the same mechanisms.

4.5 Evaluation using Evapotranspiration and Gross Primary Production

We now evaluate the performance of our data assimilation systems using independent satellite-based datasets of ~~evapotranspiration (ET) and gross primary production (GPP)~~ ET and GPP.

535 The ~~model open loop~~ tends to underestimate ET leading to an averaged negative bias of $-0.328 \text{ kg} \cdot \text{m}^{-2} \cdot \text{day}^{-1}$ reaching $-0.8 \text{ kg} \cdot \text{m}^{-2} \cdot \text{day}^{-1}$ in June and July. Both SEKF and EnSRF reduce this bias to $-0.114 \text{ kg} \cdot \text{m}^{-2} \cdot \text{day}^{-1}$ and $-0.059 \text{ kg} \cdot \text{m}^{-2} \cdot \text{day}^{-1}$, respectively. More statistics on ET can be found in Table 1. Figure 12 displays correlations between the GLEAM dataset and ~~model open loop~~ estimates (a) and the difference between correlations for the ~~model open loop~~ and the estimates produced with SEKF (b) and EnSRF (c). Overall the correlation is increased on average from 0.789 to 0.803 (SEKF) and 0.823 (EnSRF). EnSRF provides estimates that are more correlated with this independent dataset for almost ~~everywhere~~ all grid cells, it improves correlation (between 0.05 and 0.1) especially over Spain, ~~Northern-northern~~ Africa or around the Caspian Sea where correlations between the ~~model open loop~~ and GLEAM were poorer than for the rest of the domain, showing its positive impact

on ET. Similar conclusions can be drawn from geographical patterns observed for RMSD and nRMSD (not shown, see Table 1 for averaged results).

545 Figure 13 depicts correlation between GPP from the FLUXCOM dataset and ~~model-open loop~~ estimates (a) and the difference between correlations for the ~~model-open loop~~ and the estimates produced with SEKF (b) and EnSRF (c). As for ET, the EnSRF provides GPP estimates that are more correlated to the FLUXCOM dataset than ~~model-open loop~~ and SEKF estimates for almost everywhere, on average 0.817 compared to 0.784 for the model and 0.786 for the SEKF. The ~~best-biggest~~ improvements are noticeable ~~on~~ around the Caspian Sea (above 0.05) where correlations between the model and FLUXCOM GPP were
550 poorer than for the rest of the domain. Also contrary to the SEKF, degradations are confined to only few places in Iraq, Iran and close to the Arctic circle. Again similar conclusions can be drawn from geographical patterns observed for RMSD and nRMSD (not shown, see Table 1 for averaged results).

Overall the EnSRF exhibits moderate improvements for GPP and ET compared to SEKF, ~~thus validating our approach.~~

4.6 Evaluation using river discharges

555 We limit our evaluation to 92 stations over Europe with a model NSE above -1. The NIC of EnSRF compared to the ~~model-open loop~~ is displayed for those stations in Figure 14. Most stations are located in France and Germany. Blue circles denote a positive impact (above 3%) of EnSRF on estimated river discharges, red circles a negative one (below -3%) and grey diamonds a neutral impact (between -3 % and 3%). A positive NIC is observed for 61 stations and a negative NIC for only 11 stations. ~~The rest of the stations (20) showed a neutral impact.~~ Largest NIC are noticed for German stations. Such a positive influence
560 for EnSRF contrasts with the rather neutral effect of SEKF on river discharges. In compliance with previous studies (Albergel et al., 2017; Fairbairn et al., 2017), we observe a significantly positive NIC of SEKF for only 15 stations and a negative NIC for 3 stations (not shown).

5 Discussion

4.1 ~~Is the EnSRF able to provide improved estimates of LSVs?~~

565 ~~Section 4 shows overall the ability of the EnSRF to provide improved LSVs reanalyses when LAI and SSM are assimilated jointly.~~

~~For LAI, EnSRF estimates are on average as close as SEKF estimates to observations. We notice a stronger impact of both data assimilation approaches during the senescence phase than during the growing phase. This is in compliance with what Albergel et al. (2017) and Leroux et al. (2018) have observed over the Euro-Mediterranean region. During the growing phase, the system can artificially add LAI and biomass, but if atmospheric conditions are not favourable, the modelled biomass cannot
570 maintain its growing rate. During the senescence, LAI dynamics is driven by the rate of mortality, thus making DA more efficient. We further notice that DA has an impact that varies with the type of vegetation. Impact tends to be larger when the vegetation is dominated by deciduous forests. On the contrary, neither SEKF nor EnSRF has an impact on LAI for coniferous~~

575 forests. But the model performs well on places where more than 50% of plants are coniferous trees and LAI dynamics is weak in those places. Finally we observe that the EnSRF introduces a negative bias in LAI estimates compared to model or SEKF counterparts. This bias is caused by the ensemble perturbations coming from the model error. As pointed out by Fairbairn et al. (2015), model error can introduce a bias into the system in LDASs, thus showing its influence.

In the case of LAI, model errors applied to LAI in every vegetation patch are sampled from the same distribution. However, the behaviour of ensemble standard deviations varies greatly seasonally and for each type of vegetation. Standard deviations for coniferous trees are so low it leads to almost no impact of DA. Such behaviour can be explained by two caveats: first, ISBA modelled LAI evolves over a prescribed threshold ($1 \text{ m}^2 \cdot \text{m}^{-2}$ for coniferous forests, $0.3 \text{ m}^2 \cdot \text{m}^{-2}$ for other vegetation patches). When perturbed using model error, estimated LAI can be below this threshold. To avoid model issues, estimated LAI is reset to that threshold when it is the case leading to an artificially reduced ensemble standard deviation when modelled LAI is close to that threshold as in winter. Secondly, since LAI dynamics are smooth, reduced ensemble standard deviations due to the winter season still have an impact in spring through cycling. Overall we observe that ensemble standard deviations for LAI highly depend on the type of vegetation. An approach for model errors tailored for each vegetation patch could overcome the observed caveats.

590 Regarding SSM, EnSRF estimates tend to be closer to observations than SEKF estimates. This behaviour is systematic and is due to the prescribed model error for the EnSRF and the prescribed background error covariance matrix in the SEKF. We put more uncertainty in the additive model error as we are unsure of our approach compared to the background error covariance matrix that has been routinely used in the SEKF. We also observe that in arid places such as the Sahara, observed and modelled SSM are negatively correlated even after CDF match. This shows that the short-term variability of observations is completely different from what we model in ISBA. It raises the question of the quality of ISBA and/or SSM observations (after CDF matching) in arid places. Stoffelen et al. (2017) has shown that observed SSM derived from scatterometers can have a poor quality in arid places. Studying further such aspects is however beyond the scope of the paper.

About soil moisture in deeper layers, both approaches tend to provide similar estimates excepted in arid places where the EnSRF provides a wetter soil than the SEKF. This difference is solely due to a wet bias introduced again by the model error. When we perturb soil moisture, we ensure that soil moisture remains positive. In arid places where soil moisture is really low, perturbations fatally add water. As soil moisture in deeper layers is not correlated to SSM or LAI in those places, DA cannot correct this bias using observations of SSM or LAI. In other places, the EnSRF can correct the bias potentially introduced by model error to unobserved control variables through the help of correlations. They give insightful information on how the system works. Conclusions drawn from correlations are in accordance with those derived from the analysis of SEKF Jacobians drawn in Albergel et al. (2017) over the Euro-Mediterranean region and Tall et al. (2019) over Burkina Fasso. We further remark that in dry places in Summer, negative correlations between LAI and soil moisture appear. This means a higher LAI leads to a reduced soil moisture due to plant transpiration in part. Barbu et al. (2011) has already highlighted this kind of behaviour for Jacobians for grassland places in South-West France. Correlations explain the links between variables in ISBA LSM. However, they are influenced by the way we apply model error. Another model error, perturbing for example atmospheric forcing, may have led to different correlations.

Section 4 shows that EnSRF provides also improved estimates for evapotranspiration, gross primary production and river discharges. In the case of river discharges, their rather systematic improvement may ~~The rather systematic improvement of EnSRF estimates compared to the open loop may~~ be due in part to the assimilation of SSM and LAI. It may also be due in part to a bias added by the EnSRF ensemble formulation (as observed for other LSVs) that compensates an existing bias due to the coupling between ISBA and CTRIP. Further investigations have to be conducted to explore this question (~~out of scope~~). Moreover, a negative NIC is observed for most Spanish stations, where anthropogenic effects (irrigation, importance of dams, ...) ~~dominate hydrological cycles~~ can potentially modify soil moisture, streamflow and river discharges (Milano et al., 2013). Since CTRIP does not consider anthropogenic effects, this can explain poor performances of the LDAS-CTRIP system.

5 Discussion

5.1 How to deal with model errors in LDAS-Monde EnSRF?

As seen in the previous section, the quality of EnSRF estimates highly depends on the specified model error. We have seen that our system would benefit from a more tailored approach. One way that has been followed in the LDAS community is to use perturbed atmospheric forcings to generate ~~an ensemble of more physical states~~ more physical model perturbations and to obtain an ensemble whose covariances are more physically based. This can be done by either perturbing precipitations only (e.g. Fairbairn et al., 2015; Munier et al., 2015), operating a more complex system of perturbations that includes correlations between precipitation, short wave and long wave radiations (see among others Reichle et al., 2007; Liu et al., 2011; Kumar et al., 2014). Another possibility is to perturb land parameters such as the soil texture (Blyverket et al., 2019) or vegetation parameters. The main drawback of such approaches is that they tend to overcome underestimated ensemble variances by putting too much uncertainty on atmospheric forcings or model parameters that might be far better known than assumed. They can also induce a bias in model estimates (as shown by Fairbairn et al., 2015).

~~Model~~ The model error in Ensemble Kalman Filters aims to compensate insufficiencies of the model and forcings but is difficult to prescribe as it aims to compensate something we do not know. One way to curb this issue is to estimate model error. Dee (2005) describes a range of approaches to account for model biases in data assimilation systems. The last decade has also seen the development of techniques to estimate model error covariance matrices (see Tandeo et al., 2018, for a review of existing approaches). Approaches based on ~~Desroziers diagnostics~~ diagnostics developed in Desroziers et al. (2005) (Todling, 2015; Bowler, 2017) or on statistics of consecutive innovations (Berry et al., 2013; Harlim et al., 2014) seem affordable for LDASs from a computational point of view.

All these approaches help to estimate model deficiencies but do not necessarily provide the reasons of those caveats. For land surface models, they can come not only from possibly inadequate atmospheric or soil and vegetation parameters but also from inadequate model physics (missing processes, ... etc.). Finding the reasons of those is a complex task. An interesting step would be to assess the influence of atmospheric uncertainties on LSMs by using ensemble atmospheric forcings such as the 10-members atmospheric reanalysis included in ERA5 (available at a coarser spatial and temporal resolution though) or the

51 members of ECMWF ensemble medium-range forecasts. Such ~~idea has ideas have~~ been explored over Spain in the case of multi-models and multi-forcings ensembles by Ehsan Bhulyan et al. (2019).

5.2 The question of ~~cross-covariances~~1D or 3D filtering

Both SEKF and EnSRF in this paper do not consider covariances between patches and between grid cells. However, those
645 covariances are likely to exist. For example, each patch of a given grid cell is forced with the same atmospheric forcing, errors in the forcing would result ~~to in~~ correlated errors for the state of each patch. The same thing could be said for the state of two neighbouring grid cells since errors in atmospheric reanalyses are spatially correlated (Hersbach et al., 2019). Including those covariances could be beneficial to LSV reanalyses.

By construction, the SEKF cannot include these covariances by itself. Indeed the SEKF relies on the ISBA land surface
650 model to calculate covariances between variables by building the Jacobian matrix of the model. Since each patch of each grid cell of the model ~~run independently, it cannot create~~ do not interact with each other, the Jacobian between two variables of different patches is zero. The same occur for variables between different grid cells. Therefore, if we want to include covariances between patches or between grid cells, they have to be prescribed in the fixed background error covariance matrix.

On the contrary, Ensemble Kalman Filters can include this information automatically as estimated covariances are built from
655 the ensemble, thus making EnKFs more flexible than the SEKF. In our case, that would lead to a single state vector containing the LAI and SM in the various layers of soil of each patch and multiply by around 12 the size of this state. Fairbairn et al. (2015) and Carrera et al. (2015) have shown that LDASs can use a small ensemble to provide good LSVs estimates without experiencing the traditional undersampling issues or spurious ensemble covariances. However, ~~if we take into account including~~ covariances between patches or between grid cells, ~~this would be a different story. Nevertheless those two caveats would make~~
660 undersampling and spurious covariances more likely to occur due to the increased size of the state vector. Nevertheless these two potential issues can be overcome. Inflation aims to compensate undersampling by artificially inflate the ensemble spread. Approaches have been built to estimate inflation (under the form of a multiplicative coefficient). Anderson (2009) has proposed to add inflation as a parameter in the control vector leading to inflation being updated at each EnKF analysis. Bauser et al. (2018) has successfully applied this approach to a soil hydrology problem. Other approaches based on consistency diagnostics
665 developed by Desroziers et al. (2005) (Li et al., 2009; Miyoshi, 2011) or reformulated EnKFs (Bocquet, 2011; Bocquet and Sakov, 2012) have gained popularity.

Long-range spatial spurious covariances can be filtered out using localisation procedures either by artificially reducing distant spurious correlation (Hamill et al., 2001; Houtekamer and Mitchell, 2001) or by assimilating observations locally (Ott et al., 2004), LDAS-Monde could be seen as an extreme application of the second approach, because of the 1D nature of the
670 ISBA LSM. Localisation procedures are very efficient and are routinely used for a wide range of application applications.

Unfortunately, the problem of potentially spurious covariances between patches remains as we would need to fix a criterion to determine which covariance has to be reduced. Recently Farchi and Bocquet (2019) has proposed a localisation procedure based on augmented ensembles. Such formulation allows a covariance localisation not based on spatial ~~criteria and characteristics~~ and it could be used to include covariances between patches in LDAS-Monde EnSRF.

675 6 Conclusions

In this paper, we have adapted the Ensemble Square Root Filter used by Fairbairn et al. (2015) to the context of the joint assimilation of surface soil moisture and leaf area index within LDAS-Monde. The validity of our approach has then be assessed over the Euro-Mediterranean region for the period 2008 – 2017 and compared to a Simplified Extended Kalman Filter, that is routinely used in LDAS-Monde. Results ~~shows~~ show that the EnSRF provides estimates of LAI of a similar quality to the
680 SEKF. Estimated EnSRF surface soil moistures tend to get closer to observations than their SEKF counterparts. We also have examined the impact of EnSRF on controlled soil moisture for deeper soil layers. For soil moisture in near-surface layers (4-20 cm depth), analysis increments are similar for both approaches but EnSRF estimates tend to be wetter especially for arid places due to a bias introduced by the model error perturbations. For deeper layers (20-80 cm depth), SEKF and EnSRF estimates of soil moisture are similar but are obtained through different mechanisms. While drier soil moisture in SEKF is obtained through
685 the model by transferring information from updated soil moisture in (near-)surface, the EnSRF produces ~~those soil moisture~~ estimates partly because of the data assimilation routine itself, acting like a bias correction procedure for ~~those layers soil layers~~ either near the surface or in the root zone to compensate for the wet model bias via the correlations between soil moisture in deeper layers and surface soil moisture and LAI. Finally, validation of our approach has been carried out using datasets of ~~evapotranspiration, gross primary production~~ ET, GPP and river discharges, showing a moderate positive impact for ~~the two~~
690 ~~previous~~ ET and GPP but a marked positive one for ~~the latter. While involving a crude model error, this~~ river discharges. This paper shows the potential of the EnSRF within LDAS-Monde and constitutes a good basis for further developments.

One limitation of assimilating LAI is that LAI products are only available every 10 days (for CGLS products) ~~making their estimates being correctly updated every ten days (assimilating)~~. This only allows for an update of LAI every 10-days, as the assimilation of surface soil moisture ~~has a~~ is found to have negligible impact on ~~LAI~~ the LAI analyses. LDAS-Monde would
695 benefit from having observations linked to vegetation available every day. ~~This is the case of radar backscatter coefficients as shown by~~ Lievens et al. (2017) and Shamambo et al. (2019) ~~that are already used in our system through the assimilated ASCAT-derived soil water indices~~ have shown that ASCAT radar backscatter coefficients can be linked to surface soil moisture and LAI (or vegetation optical depth) through a water cloud model. The development ~~of an observation operator and the calibration of the water cloud model~~ linking surface soil moisture and LAI to ~~those coefficients~~ radar backscatter coefficient is
700 currently under development at CNRM. ~~Once fully tested, it should, hopefully, provide~~ Assimilating ASCAT radar backscatter coefficients would replace the assimilation of ASCAT-derived soil water indices. It would open the possibility to have access to daily indirect observations of LAI and improve LDAS-Monde daily updates of LAI and soil moisture.

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

first link). The developments presented in this study stemmed on SURFEX version 8.1. LDAS-Monde technical documentation and contact point are freely available at: <https://opensource.umr-cnrm.fr/projects/openldasmonde/files>

710 *Author contributions.* Conceptualization, BB, CA; Investigation, BB; Methodology, BB; Writing—original draft, BB; Writing—review and editing, All

Competing interests. The authors declare no conflict of interest.

Acknowledgements. Results were generated using the Copernicus Climate Change Service Information, 2017. The Authors would like to thanks the Copernicus Global Land Service for providing the satellite derived Leaf Area Index and Surface Soil Moisture.

715 **References**

- Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Pignatelli, 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.
- Albergel, C., Calvet, J.-C., Mahfouf, J.-F., Rüdiger, C., Barbu, A. L., Lafont, S., Roujean, J.-L., Walker, J. P., Crapeau, M. and Wigneron, J.-P.: Monitoring of water and carbon fluxes using a land data assimilation system: a case study for southwestern France, *Hydrol. Earth Syst. Sci.*, 14, 1109-1124, 10.5194/hess-14-1109-2010, 2010.
- Albergel, C., Munier S., Leroux, D. J., Dewaele, H., Fairbairn, D., Barbu, A. L., Gelati, E., Dorigo, W., Faroux, S., Meurey, C., Le Moigne, P., Decharme, B., Mahfouf, J.-F. and Calvet, J.-C.: Sequential assimilation of satellite-derived vegetation and soil moisture products using SURFEX_v8.0: LDAS-Monde assessment over the Euro-Mediterranean area, *Geosci. Model Dev.*, 10, 3889-3912, 10.5194/gmd-10-3889-2017, 2017.
- Albergel, C., Dutra, E., Munier, S., Calvet, J.-C., Sabater, J. M., de Rosnay, P. and Balsamo, G.: ERA-5 and ERA-Interim driven ISBA land surface model simulations: Which one performs better? *Hydrol. Earth Syst. Sci.*, 22, 3515-3532, 10.5194/hess-22-3515-2018, 2018.
- Albergel, C., Munier, S., Bocher, A., Bonan, B., Zheng, Y., Draper, C., Leroux, D. J. and Calvet, J.-C.: LDAS-Monde Sequential Assimilation of Satellite Derived Observations Applied to the Contiguous US: An ERA5 Driven Reanalysis of the Land Surface Variables. *Remote Sens.*, 10, 1627, 10.3390/rs10101627, 2018.
- Albergel, C., Dutra, E., Bonan, B., Zheng, Y., Munier, S., Balsamo, G., de Rosnay, P., Sabater, J. M. and Calvet, J.-C.: Monitoring and Forecasting the Impact of the 2018 Summer Heatwave on Vegetation. *Remote Sens.*, 11, 520, 10.3390/rs11050520, 2019.
- Anderson, J. L.: An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A*, 59, 210-224, 10.1111/j.1600-0870.2006.00216.x, 2009.
- Balsamo, G., Agusti-Panareda, A., Albergel, C., Arduini, G., Beljaars, A., Bidlot, J., Bousserez, N., Bousssetta, S., Brown, A., Buizza, R., Buontempo, C., Chevallier, F., Choulga, M., Cloke, H., Cronin, M. F., Dahoui, M., De Rosnay, P., Dirmeyer, P. A., Drusch, M., Dutra, E., Ek, M. B., Gentile, P., Hewitt, H., Keeley, S. P. E., Kerr, Y., Kumar, S., Lupu, C., Mahfouf, J.-F., McNorton, J., Mecklenburg, S., Mogensén, K., Muñoz-Sabater, J., Orth, R., Rabier, R., Reichle, R., Ruston, B., Pappenberger, F., Sandu, I., Seneviratne, S. I., Tietsche, S., Trigo, I. F., Uijlenhoet, R., Wedi, N., Woolway, R. I. and Zeng, X: Satellite and in situ observations for advancing global Earth surface modelling: A review. *Remote Sens.*, 10, 2038, 10.3390/rs10122038, 2018.
- Barbu, A. L., Calvet, J.-C., Mahfouf, J.-F., Albergel, C. and Lafont, S.: Assimilation of Soil Wetness Index and Leaf Area Index into the ISBA-A-gs land surface model: grassland case study, *Biogeosciences*, 8, 1971-1986, 10.5194/bg-8-1971-2011, 2011.
- Barbu, A. L., Calvet, J.-C., Mahfouf, J.-F., and Lafont, S.: Integrating ASCAT surface soil moisture and GEOV1 leaf area index into the SURFEX modelling platform: a land data assimilation application over France, *Hydrol. Earth Syst. Sci.*, 18, 173-192, 10.5194/hess-18-173-2014, 2014.
- Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmared, H., Pacholczyk, P. and Smetse, B.: GEOV1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products, Part 1: Principles of development and production, *Remote Sens. Environ.*, 137, 299-309, 10.1016/j.rse.2012.12.027, 2013.
- Bartalis, Z., Wagner, W., Naeimi, V., Hasenauer, S., Scipal, K., Bonekamp, H., Figa, J. and Anderson, C.: Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT). *Geophys. Res. Lett.*, 34, L20401, 10.1029/2007GL031088, 2007.

- Bauser, H. H., Berg, D., Klein, O., and Roth, K.: Inflation method for ensemble Kalman filter in soil hydrology, *Hydrol. Earth Syst. Sci.*, 22, 4921-4934, 10.5194/hess-22-4921-2018, 2018.
- Berg, D., Bauser, H. H. and Roth, K.: Covariance resampling for particle filter - state and parameter estimation for soil hydrology. *Hydrol. Earth Syst. Sci.*, 23, 1163-1178, 10.5194/hess-23-1163-2019, 2019.
- 755 Berry, T. and Sauer, T.: Adaptive ensemble Kalman filtering of non-linear systems. *Tellus A*, 65, 20331, 10.3402/tellusa.v65i0.20331, 2013.
- Bocquet, M.: Ensemble Kalman filtering without the intrinsic need for inflation, *Nonlin. Processes Geophys.*, 18, 735-750, 10.5194/npg-18-735-2011, 2011.
- Bocquet, M. and Sakov, P.: Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems, *Nonlin. Processes Geophys.*, 19, 383-399, 10.5194/npg-19-383-2012, 2012.
- 760 Bonan, G. B.: Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests, *Science*, 320, 5882, 1444-1449, 10.1126/science.1155121, 2008.
- Boone, A., Masson, V., Meyers, T. and Noilhan, J.: The influence of the inclusion of soil freezing on simulations by a soil-vegetation-atmosphere transfer scheme, *J. Appl. Meteorol.*, 39, 1544-1569, 10.1175/1520-0450(2000)039<1544:TIO>2.0.CO;2, 2000.
- Bowler, N. E.: On the diagnosis of model error statistics using weak-constraint data assimilation. *Q. J. Roy. Meteor. Soc.*, 143, 1916-1928, 10.1002/qj.3051, 2017.
- 765 Blyverket, J., Hamer, P. D., Bertino, L., Albergel, C., Fairbairn, D. and Lahoz, W. A.: An Evaluation of the EnKF vs. EnOI and the Assimilation of SMAP, SMOS and ESA CCI Soil Moisture Data over the Contiguous US. *Remote Sens.*, 11, 478, 10.3390/rs11050478, 2019.
- Calvet, J.-C., Noilhan, J., Roujean, J.-L., Bessemoulin, P., Cabelguenne, M., Olioso, A. and Wigneron, J.-P.: An interactive vegetation SVAT model tested against data from six contrasting sites, *Agr. Forest Meteorol.*, 92, 73-95, 10.1016/S0168-1923(98)00091-4, 1998.
- 770 Calvet, J.-C., Rivalland, V., Picon-Cochard, C. and Guehl, J.-M.: Modelling forest transpiration and CO₂ fluxes—response to soil moisture stress, *Agr. Forest Meteorol.*, 124, 143-156, 10.1016/j.agrformet.2004.01.007, 2004.
- Carrera, M. L., Bélair, S. and Bilodeau, B.: The Canadian Land Data Assimilation System (CaLDAS): Description and Synthetic Evaluation Study, *J. Hydrometeorol.*, 16, 1293-1314, 10.1175/JHM-D-14-0089.1, 2015.
- 775 Corazza, M., Kalnay, E., Patil, D. J., Yang, S.-C., Morss, R., Cai, M., Szunyogh, I., Hunt, B. R., and Yorke, J. A.: Use of the breeding technique to estimate the structure of the analysis "errors of the day", *Nonlin. Processes Geophys.*, 10, 233-243, 10.5194/npg-10-233-2003, 2003.
- Decharme, B., Alkama, R., Douville, H., Becker, M. and Cazenave, A.: Global Evaluation of the ISBA-TRIP Continental Hydrological System. Part II: Uncertainties in River Routing Simulation Related to Flow Velocity and Groundwater Storage, *J. Hydrometeorol.*, 11, 601-617, 10.1175/2010JHM1212.1, 2010.
- 780 Decharme, B., Boone, A., Delire, C. and Noilhan, J.: Local evaluation of the Interaction between Soil Biosphere Atmosphere soil multilayer diffusion scheme using four pedotransfer functions, *J. Geophys. Res.*, 116, D20126, 10.1029/2011JD016002, 2011.
- Decharme, B., Alkama, R., Papa, F., Faroux, S., Douville, H. and Prigent, C.: Global off-line evaluation of the ISBA-TRIP flood model, *Clim. Dynam.*, 38, 1389-1412, 10.1007/s00382-011-1054-9, 2012.
- Decharme, B., Martin, E. and Faroux, S.: Reconciling soil thermal and hydrological lower boundary conditions in land surface models, *J. Geophys. Res. Atmos.*, 118, 7819-7834, 10.1002/jgrd.50631, 2013.
- 785 [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, 853–877, 10.5194/tc-10-853-2016, 2016.](#)

- Decharme, B., Delire, C., Minvielle, M., Colin, J., Vergnes, J.-P., Alias, A., Saint-Martin, D., Séférian, R., Sénési, S. and Voldoire, A.:
 790 Recent changes in the ISBA-CTRIP Land Surface System; for use in the CNRM-CM6 climate model and in global off-line hydrological
 applications, *J. Adv. Model Earth Sy.*, 11, 1207-1252, <https://doi.org/10.1029/2018MS001545>, 2019.
- Dee, D. P.: Bias and data assimilation, *Q. J. Roy. Meteor. Soc.*, 131, 3323-3343, 10.1256/qj.05.137, 2005.
- De Lannoy, G. J. M., de Rosnay, P. and Reichle, R. H.: Soil Moisture Data Assimilation. In Duan, Q., Pappenberger, F., Thielen, J., Wood,
 A., Cloke, H. and Schaake J. (eds) *Handbook of Hydrometeorological Ensemble Forecasting*, pp 701-743, Springer, Berlin, Heidelberg,
 795 Germany, 10.1007/978-3-642-39925-1_32, 2016.
- de Rosnay, P., Drusch, M., Vasiljevic, D., Balsamo, G., Albergel, C. and Isaksen, L.: A simplified Extended Kalman Filter for the global
 operational soil moisture analysis at ECMWF. *Q. J. Roy. Meteor. Soc.*, 139, 1199-1213, 10.1002/qj.2023, 2013.
- Desroziers, G., Berre, L., Chapnik, B. and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space.
Q. J. Roy. Meteor. Soc., 131, 3385-3396, 10.1256/qj.05.108, 2005.
- 800 Dewaele, H., Munier, S., Albergel, C., Planque, C., Laanaia, N., Carrer, D. and Calvet, J.-C.: Parameter optimisation for a better representation
 of drought by LSMs: inverse modelling vs. sequential data assimilation, *Hydrol. Earth Syst. Sci.*, 21, 4861-4878, 10.5194/hess-21-4861-
 2017, 2017.
- Dirmeyer, P. A., Peters-Lidard, C. and Balsamo, G.: Land-atmosphere interactions and the water cycle, in: *Seamless prediction of the Earth
 system: from minutes to months*, edited by: Brunet, G., Jones, S. and Ruti, P. M., WMO-No. 1156, World Meteorological Organization,
 805 Geneva, Switzerland, 145-154, 2015.
- Draper, C. S., Mahfouf, J.-F. and Walker, J. P.: An EKF assimilation of AMSR-E soil moisture into the ISBA land surface scheme. *J. Geophys.
 Res.*, 114, D020104, 10.1029/2008JD011650, 2009.
- Draper, C., Mahfouf, J.-F., Calvet, J.-C., Martin, E., and Wagner, W.: Assimilation of ASCAT near-surface soil moisture into the SIM
 hydrological model over France, *Hydrol. Earth Syst. Sci.*, 15, 3829-3841, 10.5194/hess-15-3829-2011, 2011.
- 810 Drusch, M., Wood, E. F. and Gao, H.: Observations operators for the direct assimilation of TRMM microwave imager retrieved soil moisture,
Geophys. Res. Lett., 32, L15403, 10.1029/2005GL023623, 2005.
- Drusch, M., Scipal, K., de Rosnay, P., Balsamo, G., Andersson, E., Bougeault, P. and Viterbo, P.: Towards a Kalman Filter-based soil moisture
 analysis system for the operational ECMWF Integrated Forecast System. *Geophys. Res. Lett.*, 36, L10401, 10.1029/2009GL037716, 2009.
- Ehsan Bhulyan, M. A., Nikopoulos, E. I., Anagnostou, E. N., Polcher, J., Albergel, C., Dutra, E., Fink, G., Martinez-de la Torre, A. and
 815 Munier, S.: Assessment of precipitation error propagation in multi-model global water resource reanalysis. *Hydrol. Earth Syst. Sci.*, 23,
 1973-1994, 10.5194/hess-23-1973-2019, 2019.
- Fairbairn, D., Barbu, A. L., Mahfouf, J.-F., Calvet, J.-C. and Gelati, E.: Comparing the ensemble and extended Kalman filters for in situ soil
 moisture assimilation with contrasting conditions, *Hydrol. Earth Syst. Sci.*, 19, 4811-4830, 10.5194/hess-19-4811-2015, 2015.
- Fairbairn, D., Barbu, A. L., Napoly, A., Albergel, C., Mahfouf, J.-F. and Calvet, J.-C.: The effect of satellite-derived surface soil moisture
 820 and leaf area index land data assimilation on streamflow simulations over France, *Hydrol. Earth Syst. Sci.*, 21, 2015-2033, 10.5194/hess-
 21-2015-2017, 2017.
- Fang, H., Jiang, C., Li, W., Wei, S., Baret, F., Chen, J. M., Garcia-Haro, J., Liang, S., Liu, R., Myneni, R.B., Pinty, B., Xiao, Z. and Zhu,
 Z.: Characterization and intercomparison of global moderate resolution leaf area index (LAI) products: Analysis of climatologies and
 theoretical uncertainties. *J. Geophys. Res. Biogeosci.*, 118, 529-548, 10.1002/jgrg.20051, 2013.
- 825 Farchi, A. and Bocquet, B.: On the efficiency of covariance localisation of the ensemble Kalman filter using augmented ensembles. *Front.
 Appl. Math. Stat.*, 5, 3, 10.3389/fams.2019.00003, 2019.

- Faroux, S., Kaptué Tchuenté, A. T., Roujean, J.-L., Masson, V., Martin, E. and Le Moigne, P.: ECOCLIMAP-II/Europe: a twofold database of ecosystems and surface parameters at 1 km resolution based on satellite information for use in land surface, meteorological and climate models, *Geosci. Model Dev.*, 6, 563-582, 10.5194/gmd-6-563-2013, 2013.
- 830 Fox, A. M., Hoar, T. J., Anderson, J. L., Arellano, A. F., Smith, W. K., Litvak, M. E., MacBean, N., Schimel, D. S. and Moore, D. J. P.: Evaluation of a Data Assimilation System for Land Surface Models Using CLM4.5. *J. Adv. Model Earth Sy.*, 10, 2471-2494, 10.1002/2018MS001362, 2018.
- Gibelin, A.-L., Calvet, J.-C., Roujean, J.-L., Jarlan, L. and Los, S. O.: Ability of the land surface model ISBA-A-gs to simulate leaf area index at the global scale: Comparison with satellites products, *J. Geophys. Res.*, 111, D18102, 10.1029/2005JD006691, 2006.
- 835 Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, *Mon. Weather Rev.*, 129, 2776-2790, 10.1175/1520-0493(2001)129<2776%3ADDFOBE>2.0.CO%3B2, 2001.
- Hamill, T. M. and Whitaker, J. S.: Accounting for the error due to unresolved scales in ensemble data assimilation: a comparison of different approaches, *Mon. Weather Rev.*, 133, 3132-3147, 10.1175/MWR3020.1, 2005.
- Harlim, J., Mahdi, A. and Majda, A. J.: An ensemble Kalman filter for statistical estimation of physics constrained nonlinear regression
840 models. *J. Comput. Phys.*, 257, 782-812, 10.1016/j.jcp.2013.10.025, 2014.
- Hersbach, H. and Dee, D.: "ERA-5 reanalysis is in production", ECMWF newsletter, number 147, Spring 2016, p. 7, 2016.
- [Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Holm, E., Janiskova, M., Keeley, S., Laloyaux, P., Lopez, P., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. and Thépaut, J.-N.: The ERA5 Global Reanalysis, *Q. J. Roy. Meteor. Soc.*, submitted, 2019.](#)
- 845 [The ERA5 Global Reanalysis, *Q. J. Roy. Meteor. Soc.*, submitted, 2019.](#)
- Houtekamer, P. L. and Mitchell, H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation, *Mon. Weather Rev.*, 129, 123-137, 10.1175/1520-0493(2001)129<0123%3AASEKFF>2.0.CO%3B2, 2001.
- Ines, A. V., Das, N. N., Hansen, J. W. and Njoku, E.G.: Assimilation of remotely sensed soil moisture and vegetation with a crop simulation
850 model for maize yield prediction. *Remote Sens. Environ.* 138, 149-164, 10.1016/j.rse.2013.07.018, 2013.
- Jarlan, L., Balsamo, G., Lafont, S., Beljaars, A., Calvet, J.-C. and Mougín, E.: Analysis of leaf area index in the ECMWF land surface model and impact on latent heat and carbon fluxes: Application to West Africa. *J. Geophys. Res.*, 113, D24117, 10.1029/2007JD009370, 2008.
- Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G. and Wang, J.: A review of data assimilation of remote sensing and crop models, *Eur. J. Agron.*, 92, 141-152, 10.1016/j.eja.2017.11.002, 2018.
- 855 Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P., Gans, F., Ichii, K., Jain, A. K., Kato, E., Papale, D., Poulter, B., Raduly, B., Rödenbeck, C., Tramontana, G., Viovy, N., Wang, Y. P., Weber, U., Zaehle, S. and Zeng, N.: Compensatory water effects link yearly global land CO₂ sink changes to temperature, *Nature*, 541, 516-520, 10.1038/nature20780, 2017.
- Kumar, S. V., Peters-Lidard, C. D., Mocko, D., Reichle, R. H., Liu, Y., Arsenault, K. R., Xia, Y., Ek, M., Riggs, G., Livneh, B. and Cosh, M:
860 Assimilation of Remotely Sensed Soil Moisture and Snow Depth Retrievals for Drought Estimation. *J. Hydrometeorol.*, 15, 2446-2469, 10.1175/JHM-D-13-0132.1, 2014.
- Kumar, S. V., Mocko, D. M., Wang, S., Peters-Lidard, C. D. and Borak, J.: Assimilation of remotely sensed Leaf Area Index into the Noah-MP land surface model: Impacts on water and carbon fluxes and states over the Continental U.S.. *J. Hydrometeorol.*, 10.1175/JHM-D-18-0237.1, 2019.

- 865 Lahoz, W. A. and De Lannoy, G. J. M.: Closing the Gaps in Our Knowledge of the Hydrological Cycle over Land: Conceptual Problems. *Surv. Geophys.*, 35, 623-660, 10.1007/s10712-013-9221-7, 2014.
- Leroux, D. J., Calvet, J.-C., Munier, S. and Albergel, C.: Using satellite-derived vegetation products to evaluate LDAS-Monde over the Euro-Mediterranean Area, *Remote Sens.*, 10, 1199, 10.3390/rs10081199, 2018.
- Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G.J., Pan, M. and Wood, E.F.: Inroads of remote sensing into hydrologic science during
870 the WRR era. *Water Resour. Res.*, 51, 7309-7342, 10.1002/2015WR017616, 2015.
- Li, H., Kalnay, E., and Miyoshi, T.: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter, *Q. J. Roy. Meteor. Soc.*, 135, 523-533, 10.1002/qj.371, 2009.
- Lievens, H., Martens, B., Verhoest, N. E. C., Hahn, S., Reichle, R. H. and Miralles, D. G.: Assimilation of global radar backscatter and radiometer brightness temperature observations to improve soil moisture and land evaporation estimates. *Remote Sens. Environ.*, 189,
875 194-210, 10.1016/j.rse.2016.11.022, 2017.
- Ling, X.-L., Fu, C.-B., Yang, Z.-L., and Guo, W.-D.: Comparison of different sequential assimilation algorithms for satellite-derived leaf area index using the Data Assimilation Research Testbed (version Lanai), *Geosci. Model Dev.*, 12, 3119-3133, 10.5194/gmd-12-3119-2019, 2019.
- Liu, Q., Reichle, R. H., Bindlish, R., Cosh, M. H., Crow, W. T., de Jeu, R., De Lannoy, G. J. M., Huffman, G. J. and Jackson, T. J.: The
880 Contributions of Precipitation and Soil Moisture Observations to the Skill of Soil Moisture Estimates in a Land Data Assimilation System. *J. Hydrometeorol.*, 12, 750-765, 10.1175/JHM-D-10-05000.1, 2011.
- Livingston, D. M., Dance, S. L. and Nichols, N. K.: Unbiased Ensemble Square Root Filters. *Physica D*, 237, 1021-1028, 10.1016/j.physd.2008.01.005, 2008.
- Mahfouf, J.-F.: L'analyse dans le sol à Météo-France. Partie 1: Evaluation et perspectives à l'échelle locale, *Meteo-France technical report*,
885 Toulouse, France, 2007.
- Mahfouf, J.-F., Bergaoui, K., Draper, C., Bouyssel, C., Taillefer, F. and Taseva, L.: A comparison of two offline soil analysis schemes for assimilation of screen-level observations, *J. Geophys. Res.*, 114, D08105, 10.1029/2008JD011077, 2009.
- Maggioni, V. and Houser, P. R.: Soil Moisture Data Assimilation. In Park, S. and Lu, X. (eds) *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. III)*, pp 195-217, Springer, Cham, Switzerland, 10.1007/978-3-319-43415-5_9, 2017.
- 890 Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A. and Verhoest, N. E. C.: GLEAM v3: Satellite-based land evaporation and root-zone soil moisture, *Geosci. Model Dev.*, 10, 1903-1925, 10.5194/gmd-10-1903-2017, 2017.
- Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gibelin, A.-L., Giordani, H., Habets, F., Jidane, M.,
895 Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J.-C., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V. and Voldoire, A.: The SURFEXv7.2 land and ocean surface platform for coupled and offline simulation of earth surface variables and fluxes, *Geosci. Model Dev.*, 6, 929-960, 10.5194/gmd-6-929-2013, 2013.
- [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.](#)
- 900 Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A. and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations. *Hydrol. Earth Syst. Sci.*, 15, 453-469, 10.5194/hess-15-453-2011, 2011.

- Miyoshi, T.: The Gaussian Approach to Adaptive Covariance Inflation and Its Implementation with the Local Ensemble Transform Kalman Filter, *Mon. Weather Rev.*, 139, 1519-1535, 10.1175/2010MWR3570.1, 2011.
- 905 Munier, S., Polebistki, A., Brown, C., Belaud, G. and Lettenmaier, D. P.: SWOT data assimilation for operational reservoir management on the upper Niger River Basin, *Water Resour. Res.*, 51, 554-575, 10.1002/2014WR016157, 2015.
- Noilhan, J. and Planton, S.: A simple parameterization of land surface processes for meteorological models. *Mon. Weather Rev.*, 117, 536-549, 10.1175/1520-0493(1989)117<0536%3AASPOLS>2.0.CO%3B2, 1989.
- Noilhan, J. and Mahfouf, J.-F.: The ISBA land surface parameterisation scheme, *Global Planetary Change*, 13, 145-159, 10.1016/0921-910 8181(95)00043-7, 1996.
- Oki, T. and Sud, Y. C.: Design of Total Runoff Integrating Pathways (TRIP), a global river channel network, *Earth Interact.*, 2, 1-36, 10.1175/1087-3562(1998)002<0001:DOTRIP>2.3.CO;2, 1998.
- Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J., and Yorke, A.: A local ensemble Kalman filter for atmospheric data assimilation, *Tellus A*, 56, 415-428, 10.1111/j.1600-0870.2004.00076.x, 2004.
- 915 Pan, M., Wood, E. F., Wojcik, R. and McCabe, M. F.: Estimation of regional terrestrial water cycle using multi-sensor remote sensing observations and data assimilation. *Remote Sens. Environ.*, 112, 1282-1294, 10.1016/j.rse.2007.02.039, 2008.
- Pauwels, V. R. N., Verhoest, N. E. C., De Lannoy, G. J. M., Guissard, V., Lucau, C. and Defourny, P.: Optimization of a coupled hydrology-crop growth model through the assimilation of observed soil moisture and leaf area index values using an ensemble Kalman filter. *Water Resour. Res.*, 43, W04421, 10.1029/2006WR004942, 2007.
- 920 Plaza, D. A., De Keyser, R., De Lannoy, G. J. M., Giustarini, L., Matgen, P., and Pauwels, V. R. N.: The importance of parameter resampling for soil moisture data assimilation into hydrologic models using the particle filter, *Hydrol. Earth Syst. Sci.*, 16, 375-390, 10.5194/hess-16-375-2012, 2012.
- Reichle, R. H., Walker, J. P., Koster, R. D. and Houser, P. R.: Extended versus Ensemble Kalman Filtering for land data assimilation. *J. Hydrometeorol.*, 3, 728-740, 10.1175/1525-7541(2002)003<0728%3AEVEKFF>2.0.CO%3B2, 2002.
- 925 Reichle, R. H. and Koster, D.: Bias reduction in short records of satellite soil moisture, *Geophys. Res. Lett.*, 31, L19501, 10.1029/2004GL020938, 2004.
- Reichle, R. H., Koster, R. D., Liu, P., Mahanama, S. P. P., Njoku, E. G. and Owe, M.: Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *J. Geophys. Res. Atmos.*, 112, D09108, 10.1029/2006JD008033, 2007.
- 930 Reichle, R. H., De Lannoy, G. J. M., Forman, B. F., Draper, C. S. and Liu, Q.: Connecting Satellite Observations with Water Cycle Variables Through Land Data Assimilation: Examples Using the NASA GEOS-5 LDAS. *Surv. Geophys.*, 35, 577-606, 10.1007/s10712-013-9220-8, 2014.
- Richards, L. A.: Capillary conduction of liquids through porous mediums, *Physics*, 1, 318-333, 10.1063/1.1745010, 1931.
- Rüdiger, C., Albergel, C., Mahfouf, J.-F., Calvet, J.-C., and Walker, J. P.: Evaluation of Jacobians for Leaf Area Index data assimilation with 935 an extended Kalman filter, *J. Geophys. Res.*, 115, D09111, 10.1029/2009JD012912, 2010.
- Sabater, J. M., Jarlan, L., Calvet, J.-C., and Boyssel, F.: From near-surface to root-zone soil moisture using different assimilation techniques, *J. Hydrometeorol.*, 8, 194-206, 10.1175/JHM571.1, 2007.
- Sabater, J. M., Rüdiger, C., Calvet, J.-C., Fritz, N., Jarlan, L. and Kerr Y.: Joint assimilation of surface soil moisture and LAI observations into a land surface model, *Agr. Forest Meteorol.*, 148, 1362-1373, 10.1016/j.agrformet.2008.04.003, 2008.

- 940 Sakov, P. and Oke, P. R.: A deterministic formulation of the Ensemble Kalman Filter: an alternative to ensemble square root filters. *Tellus A*, 60, 361-371, 10.1111/j.1600-0870.2007.00299.x, 2008.
- Sawada, Y., Koike, T. and Walker, J. P.: A land data assimilation system for simultaneous simulation of soil moisture and vegetation dynamics, *J. Geophys. Res. Atmos.*, 120, 5910-5930, 10.1002/2014JD022895, 2015.
- Sawada, Y.: Quantifying Drought Propagation from Soil Moisture to Vegetation Dynamics Using a Newly Developed Ecohydrological Land
945 Reanalysis. *Remote Sens.*, 10, 1197, 10.3390/rs10081197, 2018.
- Schellekens, J., Dutra, E., Martínez-de la Torre, A., Balsamo, G., van Dijk, A., Sperna Weiland, F., Minvielle, M., Calvet, J.-C., Decharme, B., Eisner, S., Fink, G., Flörke, M., Peßenteiner, S., van Beek, R., Polcher, J., Beck, H., Orth, R., Calton, B., Burke, S., Dorigo, W., and Weedon, G. P.: A global water resources ensemble of hydrological models: the earth2Observe Tier-1 dataset, *Earth Syst. Sci. Data*, 9, 389-413, 10.5194/essd-9-389-2017, 2017.
- 950 [Schmugge, T. J.: Remote Sensing of Soil Moisture: Recent Advances, IEEE T. Geosci. Remote, GE21, 145-146, 10.1109/TGRS.1983.350563, 1983.](#)
- Scipal, K., Drusch, M. and Wagner, W.: Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system, *Adv. Water Resour.*, 31, 1101-1112, 10.1016/j.advwatres.2008.04.013, 2008.
- Shamambo, D. C., Bonan, B., Calvet, J.-C., Albergel, C. and Hahn, S.: Interpretation of radar scatterometer observations over land: a case
955 study over southwestern France, *Remote Sens.*, submitted, 2019.
- Stoffelen, A., Aaboe, S., Calvet, J.-C., Cotton, J., De Chiara, G., Figua-Saldana, J., Mouche, A. A., Portabella, M., Scipal, K. and Wagner, W.: Scientific developments and the EPS-SG scatterometer, *IEEE J. Sel. Top. Appl.*, 10, 2086-2097, 10.1109/JSTARS.2017.2696424, 2017.
- Tall, M., Albergel, C., Bonan, B., Zheng, Y., Guichard, F., Dramé, M. S., Gaye, A. T., Sintondji, L. O., Hountondji, F. C. C., Nikiema, P. M. and Calvet, J.-C.: Towards a Long-Term Reanalysis of Land Surface Variables over Western Africa: LDAS-Monde Applied over Burkina
960 Faso from 2001 to 2018. *Remote Sens.*, 11, 735, 10.3390/rs11060735, 2019.
- Tandeo, P., Ailliot, P., Bocquet, M., Carrassi, A., Miyoshi, T., Pulido, M. and Zhen, Y.: Joint Estimation of Model and Observation Error Covariance Matrices in Data Assimilation: a Review. *Mon. Weather Rev.*, submitted, available at: <https://arxiv.org/abs/1807.11221v2>, 2018.
- Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M. and Whitaker, J. S.: Ensemble Square Root Filters. *Mon. Weather Rev.*, 131, 1485-1490, 10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2, 2003.
965
- Todling, R.: A lag-1 smoother approach to system-error estimation: sequential method. *Q. J. Roy. Meteor. Soc.*, 141, 1502-1513, 10.1002/qj.2460, 2015.
- Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S. and Papale, D.: Predicting carbon dioxide and energy fluxes across global FLUXNET
970 sites with regression algorithms, *Biogeosciences*, 13, 4291-4313, 10.5194/bg-13-4291-2016, 2016.
- Vergnes, J.-P. and Decharme, B.: A simple groundwater scheme in the TRIP river routing model: global off-line evaluation against GRACE terrestrial water storage estimates and observed river discharges, *Hydrol. Earth. Syst. Sci.*, 16, 3889-3908, 10.5194/hess-16-3889-2012, 2012.
- Vergnes, J.-P., Decharme, B. and Habets, F.: Introduction of groundwater capillary rises using subgrid spatial variability of topography into
975 the ISBA land surface model, *J. Geophys. Res. Atmos.*, 119, 11065-11086, 10.1002/2014JD021573, 2014.
- Voltaire, A., Decharme, B., Pianezze, J., Lebeauin Brossier, C., Sevault, F., Seyfried, L., Garnier, V., Bielli, S., Valcke, S., Alias, A., Accensi, M., Ardhuin, F., Bouin, M.-N., Ducrocq, V., Faroux, S., Giordani, H., Léger, F., Marsaleix, P., Rainaud, R., Redelsperger, J.-L.,

- Richard, E. and Riette, S.: SURFEX v8.0 interface with OASIS3-MCT to couple atmosphere with hydrology, ocean, waves and sea-ice models, from coastal to global scales, *Geosci. Model Dev.*, 10, 4207-4227, 10.5194/gmd-10-4207-2017, 2017.
- 980 Vreugdenhil, M., Dorigo, W. A., Wagner, W., de Jeu, R. A. M., Hahn, S. and van Marle, M. J. E.: Analyzing the Vegetation Parameterization in the TU-Wien ASCAT Soil Moisture Retrieval. *IEEE T. Geosci. Remote*, 54, 6, 3513-31, 10.1109/TGRS.2016.2519842, 2016.
- Wagner, W., Lemoine, G. and Rott, H: A method for estimating soil moisture from ERS scatterometer and soil data. *Remote Sens. Environ*, 70, 191-207, 10.1016/S0034-4257(99)00036-X, 1999.
- Whitaker, J. S. and Hamill, T. M.: Ensemble data assimilation without perturbed observations, *Mon. Weather Rev.*, 130, 1913-1924, 985 10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2, 2002.
- Xiao, Z., Liang, S., Wang, J., Chen, P., Yin, X., Zhang, L. and Song, J.: Use of general regression neural networks for generating the GLASS leaf area index product from time-series MODIS surface reflectance. *IEEE Trans. Geosci. Remote*, 10.1109/TGRS.2013.2237780, 2013.
- Zhang, H., Hendricks Franssen, H.-J., Han, X., Vrugt, J. A., and Vereecken, H.: State and parameter estimation of two land surface models using the ensemble Kalman filter and the particle filter, *Hydrol. Earth Syst. Sci.*, 21, 4927-4958, 10.5194/hess-21-4927-2017, 2017.

Table 1. Statistics (Root Mean Square Difference (RMSD), normalized RMSD (nRMSD), correlation (R), and bias) between LDAS-Monde estimates (~~Model-run~~open loop, SEKF and EnSRF) and observations for CGLS SSM, CGLS LAI, GLEAM Evapotranspiration (~~E~~ET) and FLUXCOM GPP averaged over the Euro-Mediterranean region for the period 2008–2017 (for SSM, LAI and E) or 2008–2013 (for GPP).

Variable	Exp.	RMSD	nRMSD	R	Bias
LAI	Model-run <u>open loop</u>	0.880 m ² .m ⁻²	0.568	0.593	-0.020 m ² .m ⁻²
	SEKF	0.671 m ² .m ⁻²	0.401	0.732	-0.116 m ² .m ⁻²
	EnSRF	0.694 m ² .m ⁻²	0.419	0.723	-0.201 m ² .m ⁻²
SSM	Model-run <u>open loop</u>	0.035 m ³ .m ⁻³	0.161	0.544	0.002 m ³ .m ⁻³
	SEKF	0.032 m ³ .m ⁻³	0.138	0.652	0.001 m ³ .m ⁻³
	EnSRF	0.027 m ³ .m ⁻³	0.117	0.760	0.001 m ³ .m ⁻³
Evapotranspiration <u>ET</u>	Model-run <u>open loop</u>	0.833 kg.m ⁻² .day ⁻¹	0.712	0.789	-0.328 kg.m ⁻² .day ⁻¹
	SEKF	0.778 kg.m ⁻² .day ⁻¹	0.689	0.803	-0.114 kg.m ⁻² .day ⁻¹
	EnSRF	0.745 kg.m ⁻² .day ⁻¹	0.678	0.823	-0.059 kg.m ⁻² .day ⁻¹
GPP	Model-run <u>open loop</u>	1.369 g(C).m ⁻² .day ⁻¹	0.913	0.784	-0.412 g(C).m ⁻² .day ⁻¹
	SEKF	1.393 g(C).m ⁻² .day ⁻¹	0.962	0.786	-0.146 g(C).m ⁻² .day ⁻¹
	EnSRF	1.344 g(C).m ⁻² .day ⁻¹	0.908	0.817	-0.105 g(C).m ⁻² .day ⁻¹

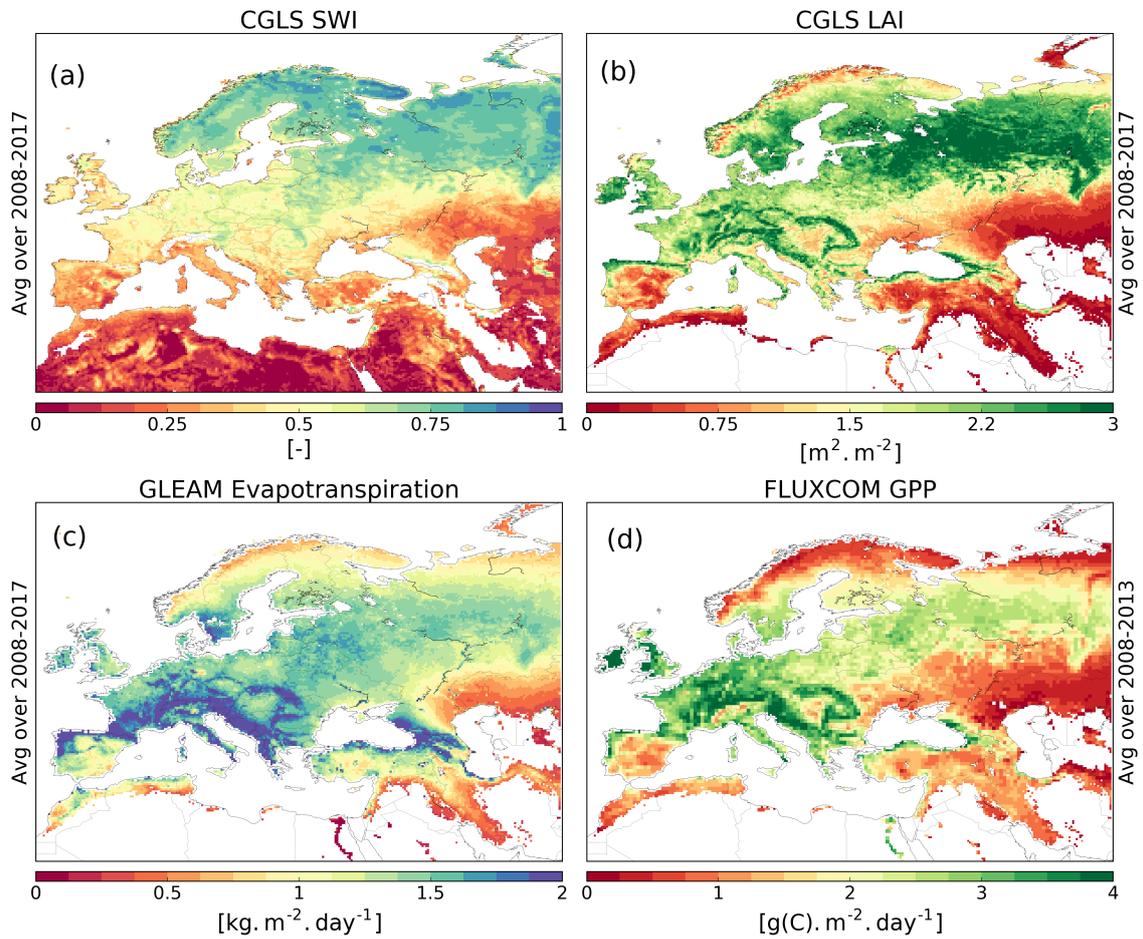


Figure 1. Satellite-derived products of (a) original Soil Water Index (SWI), (b) Leaf Area Index (LAI), (c) evapotranspiration (ET) and (d) Gross Primary Production (GPP). They are averaged over 2008-2017 for (a), (b) and (c) and over 2008-2013 for (d).

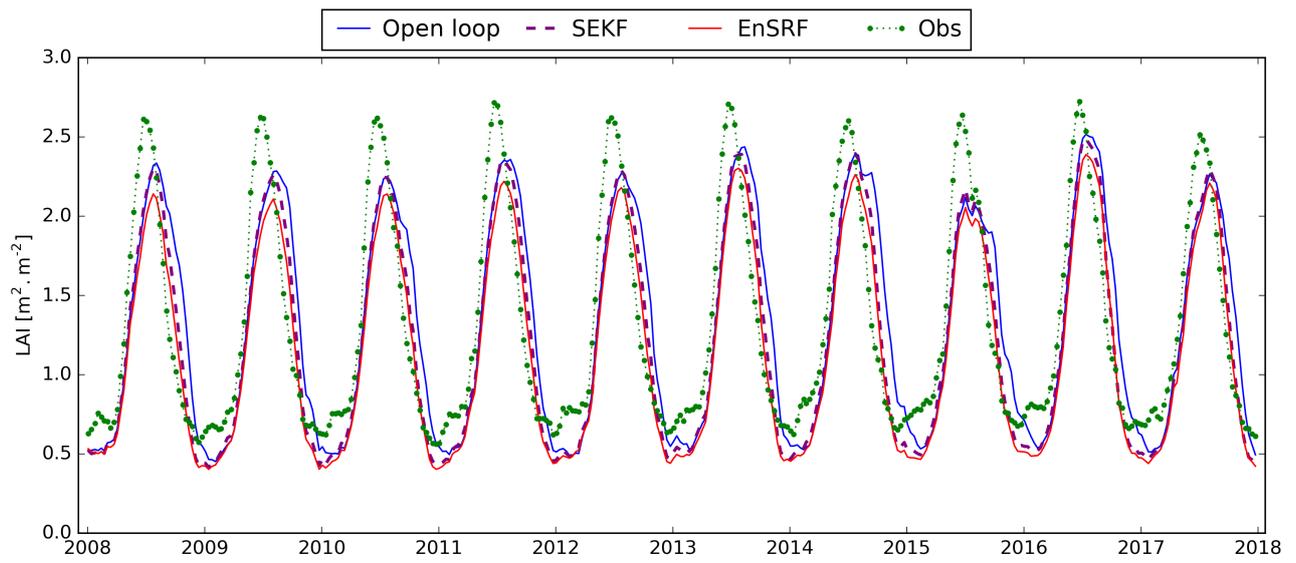


Figure 2. 10-days time series of LAI averaged over the whole domain from the ~~model~~-open loop (blue line), the observations (green dots and dotted line) and analyses obtained with the SEKF (dashed purple line) and the EnSRF (red line) for the period 2008-2017.

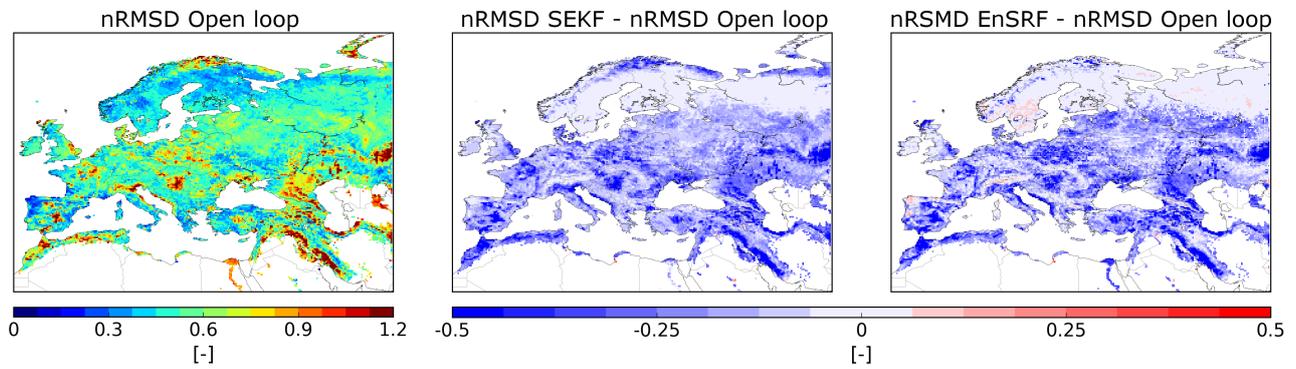


Figure 3. (a) Normalized RMSD (nRMSD) between observed LAI and its model open loop equivalent for the period 2008-2017 and nRMSD difference between nRMSD for assimilation experiments (SEKF in (b) and EnSRF in (c) vs nRMSD Model) and the open loop.

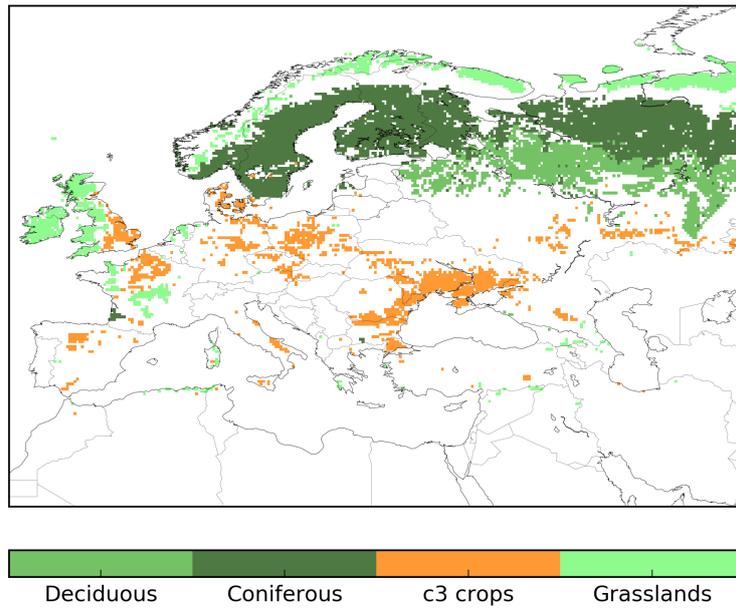


Figure 4. Grid cells of the domain where a vegetation type (or patch) is predominant (patch fraction above 50%). Coniferous trees are dominant for around 15% of the domain that has plants (dark green), deciduous broadleaved trees (green), c3 crops (orange) and grasslands (light green) are in majority for 6% of the domain each.

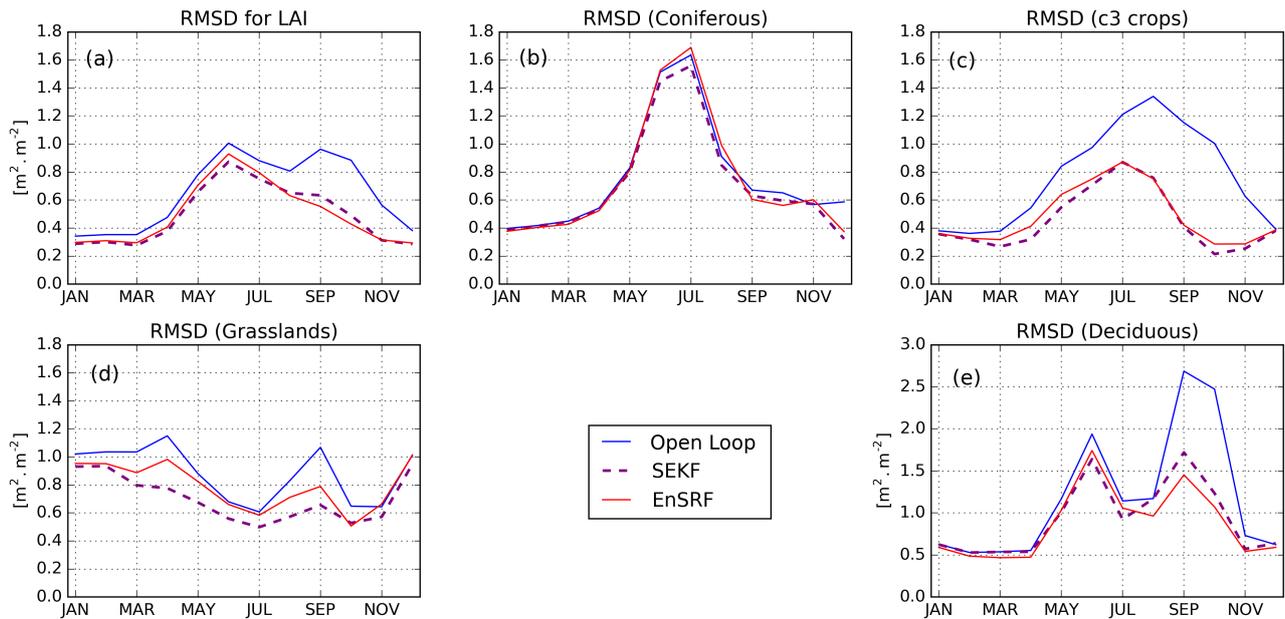


Figure 5. Seasonal RMSD between LAI from observations and the model-open loop (blue line), the SEKF analysis (dashed purple line) and the EnSRF analysis (red line) averaged over: (a) the whole domain, and grid cells where (b) coniferous trees, (c) c3 crops, (d) grasslands, (e) deciduous broadleaved trees represent more than 50% of plants for the period 2008-2017.

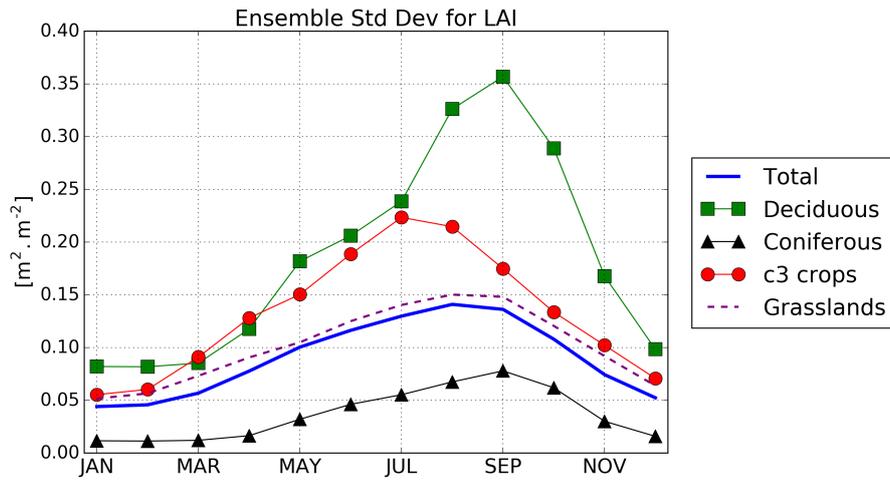


Figure 6. Seasonal standard deviation of the ensemble from the EnSRF averaged over: the whole domain (thick blue line), and grid cells where deciduous broadleaved trees (green squares), coniferous trees (black triangles), c3 crops (red circles) and grasslands (dashed purple line) represent the majority of plants for the period 2008-2017.

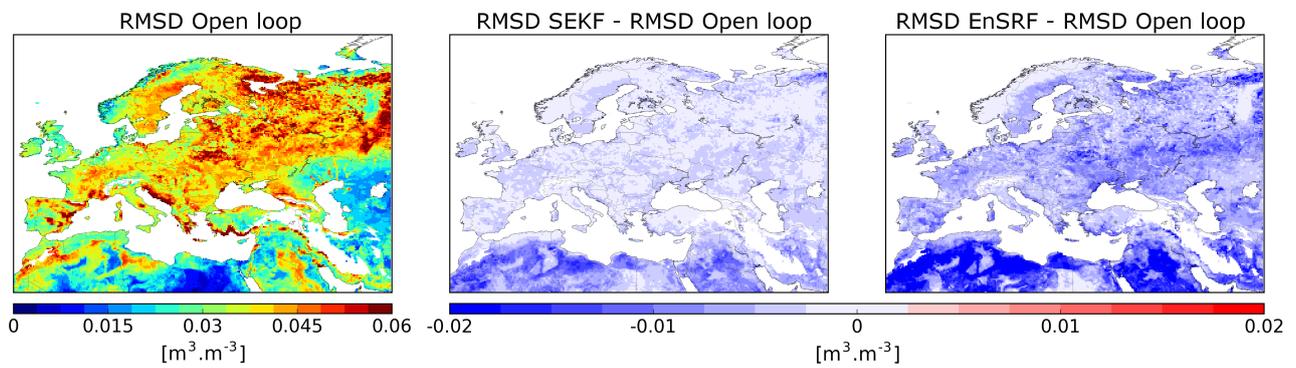


Figure 7. (a) Root mean square difference (RMSD) between observed (rescaled) SSM and its model-open loop equivalent for the period 2008-2017 and RMSD difference between RMSD-for-assimilation experiments (SEKF in (b) and EnSRF in (c)vs-RMSD-Model) and the open loop.

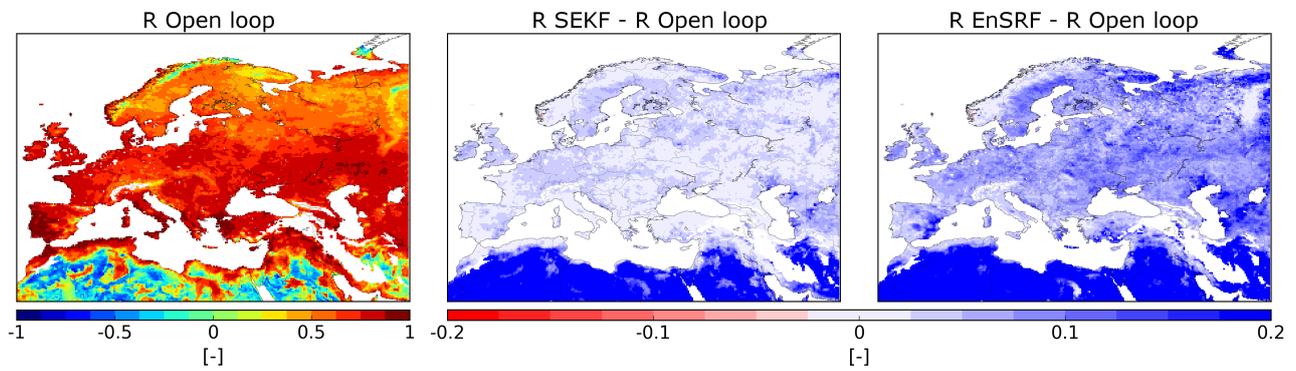


Figure 8. (a) Correlation (R) between observed (rescaled) SSM and its model-open loop equivalent for the period 2008-2017 and R difference between R-for-assimilation experiments (SEKF in (b) and EnSRF in (c)vs R-Model) and the open loop..

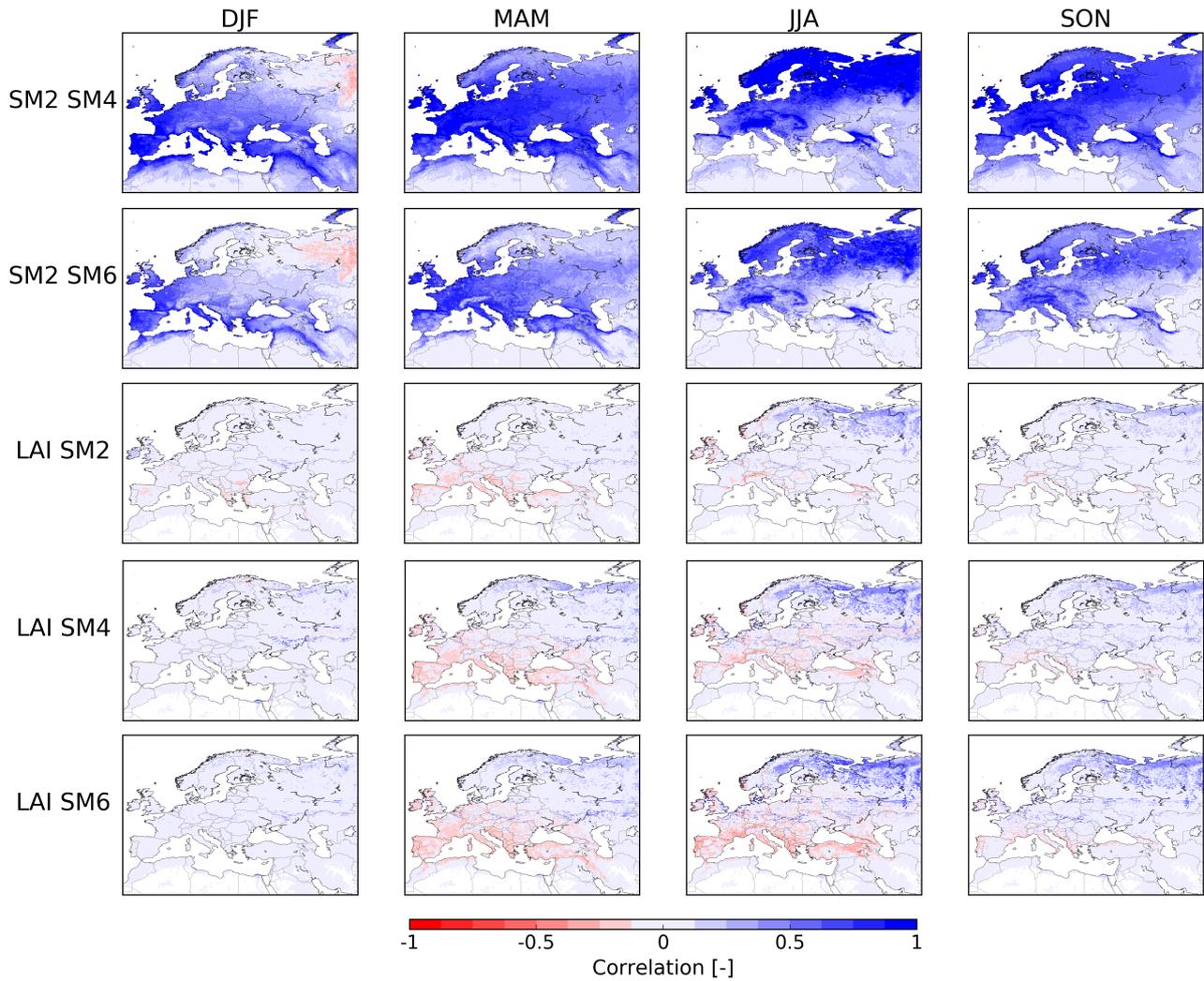


Figure 9. Correlation between the model variables sampled from ensembles and averaged seasonally (DJF=December-January-February, MAM=March-April-May, JJA=June-July-August and SON=September-October-November). From top to bottom, correlation between soil moisture in the second layer (1-4cm, SM2) and the fourth layer (10-20cm, SM4), between SM2 and soil moisture in the sixth layer (40-60cm, SM6), between LAI and SM2, LAI and SM4 and LAI and SM6. Areas in blue exhibit positive correlations, areas in red exhibit anti-correlations.

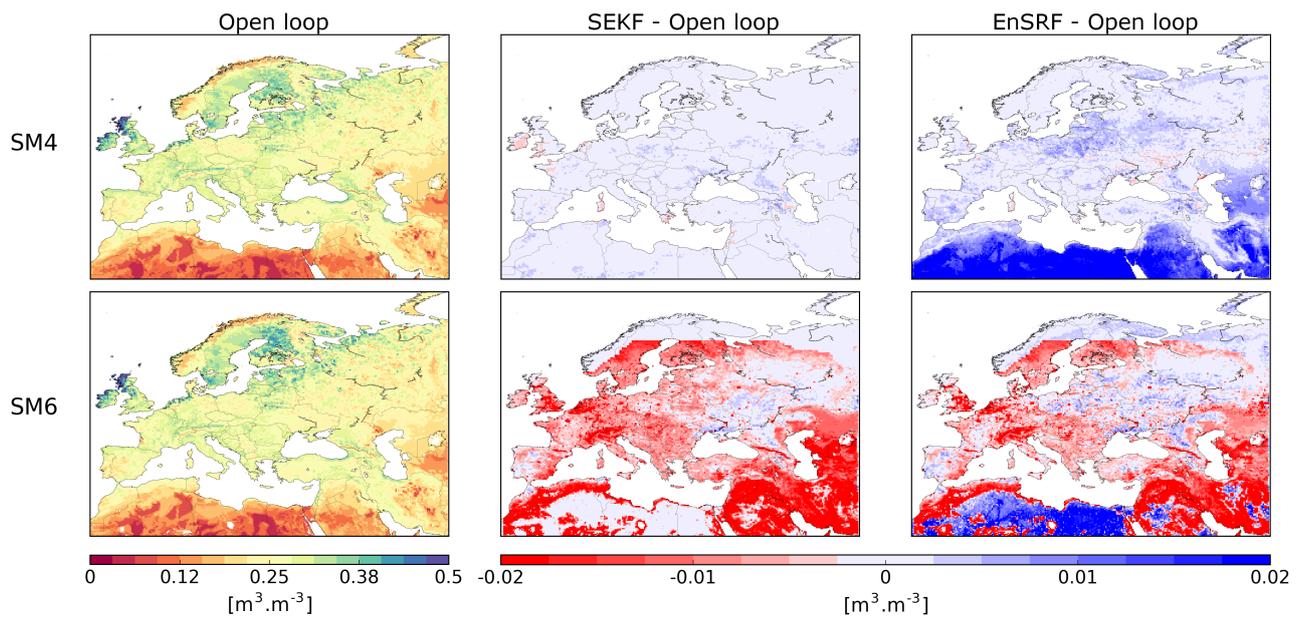


Figure 10. From left to right, averaged soil moisture [for the open loop](#) (fourth layer, 10-20 cm, SM4 and sixth layer, 40-60 cm, SM6) over 2008-2017, averaged analysis impact for SEKF (central) and EnSRF (right).

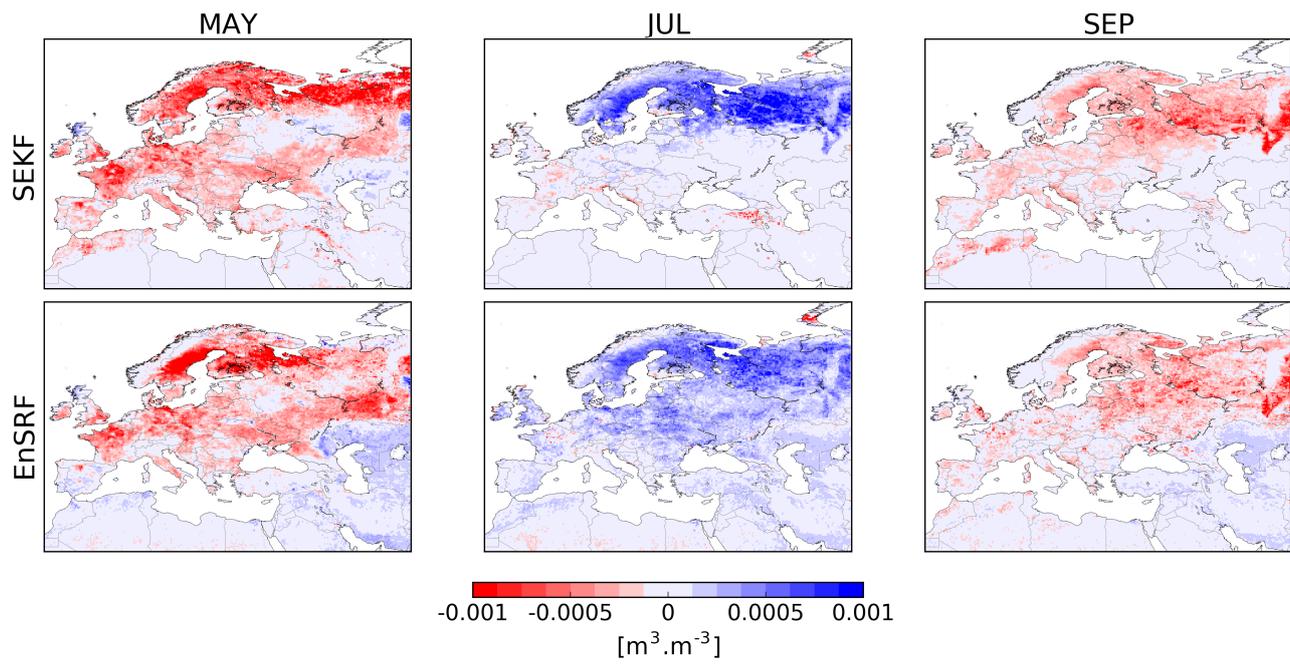


Figure 11. Averaged analysis increments for soil moisture in fourth layer (10-20cm, SM4) for SEKF and EnSRF for all months of May (left), July (central) and September (right).

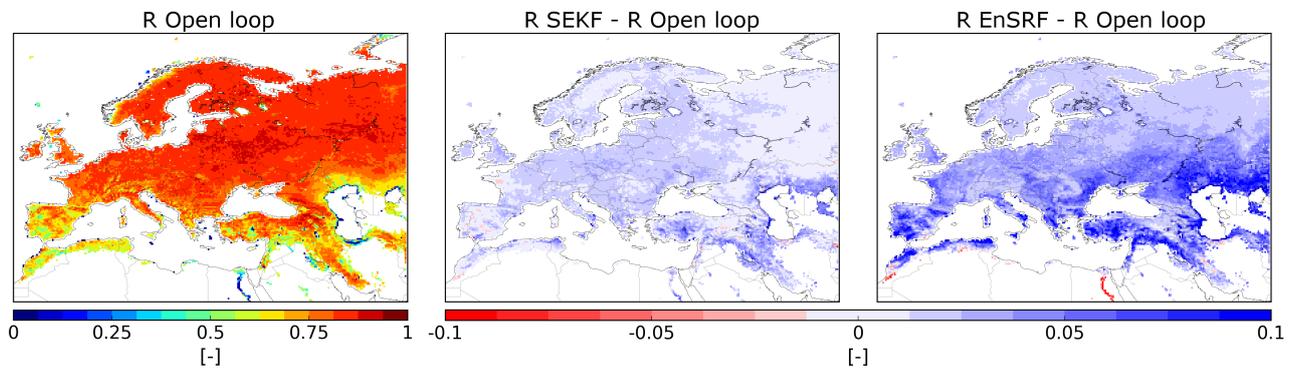


Figure 12. (a) Correlation (R) between observed GLEAM Evapotranspiration-ET and its model-open loop equivalent for the period 2008-2017 and R difference between R for assimilation experiments (SEKF in (b) and EnSRF in (c) vs R-Model) and the open loop.

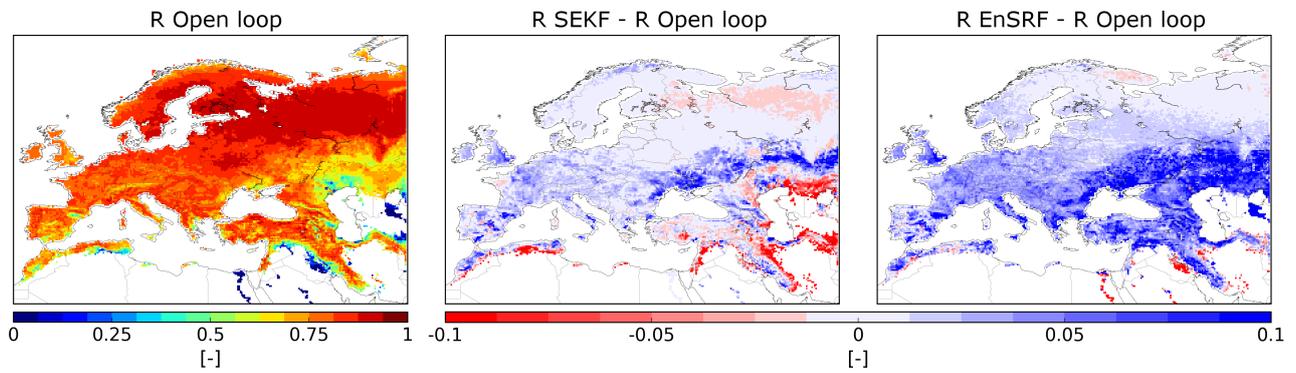


Figure 13. (a) Correlation (R) between observed FLUXCOM gross primary production and its model-open loop equivalent for the period 2008-2013 and R difference between R -for-assimilation experiments (SEKF in (b) and EnSRF in (c) vs R -Model) and the open loop.

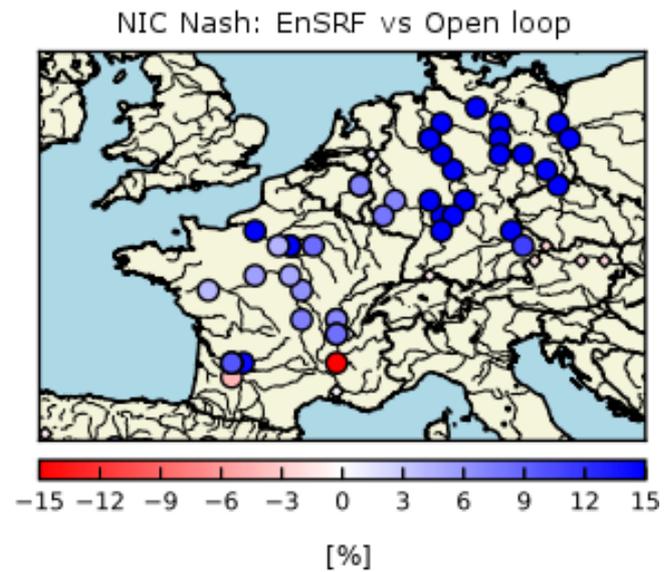
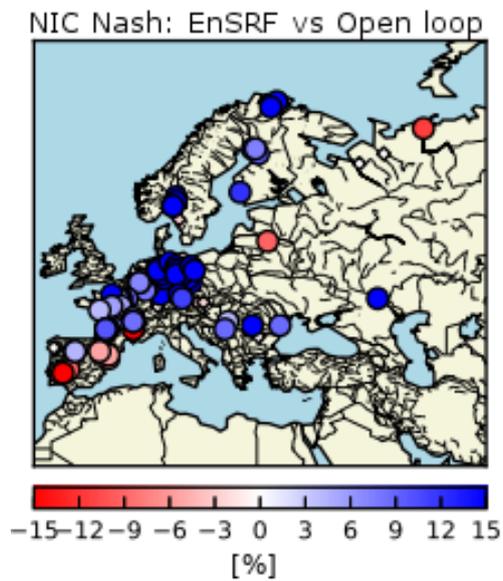


Figure 14. Normalised Information Contribution Index (NIC) assessing the improvement of Nash-Sutcliffe efficiency indices for EnSRF river discharge estimates compared to model-open-loop counterparts. Blue circles assess a positive impact of DA, red circles a negative impact and small diamonds a neutral impact.

SUPPLEMENTARY MATERIAL TO

“AN ENSEMBLE SQUARE ROOT FILTER FOR THE JOINT ASSIMILATION OF SURFACE SOIL MOISTURE AND LEAF AREA INDEX WITHIN LDAS-MONDE: APPLICATION OVER THE EURO-MEDITERRANEAN REGION”

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*

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

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

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

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

\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

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

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

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

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:

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

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

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 $\mathbf{x}_{[p]}$, $p = 1, \dots, 12$, 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:

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

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:

$$\mathbf{H}_k \mathbf{x} = \sum_{j=1}^{12} \alpha_{[j]} \tilde{\mathbf{H}}_k \mathbf{x}_{[j]} \quad (9)$$

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

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

$$\mathbf{H}_k = \left(\alpha_{[1]} \tilde{\mathbf{H}}_k \quad \alpha_{[2]} \tilde{\mathbf{H}}_k \quad \dots \quad \alpha_{[12]} \tilde{\mathbf{H}}_k \right) \quad (10)$$

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:

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

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

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

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:

$$\mathbf{B} = \begin{pmatrix} \tilde{\mathbf{B}} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{B}} & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \tilde{\mathbf{B}} \end{pmatrix} \quad (13)$$

with $\tilde{\mathbf{B}}$ the prescribed covariance matrix for control variables within a patch. In practice $\tilde{\mathbf{B}}$ 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:

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

and

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

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

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

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 :

$$\mathbf{x}_{[p],k}^a = \mathbf{x}_{[p],k}^f + \mathbf{K}_{[p],k} \left(\mathbf{y}_k^o - \sum_{j=1}^{12} \alpha_{[j]} \tilde{\mathbf{H}}_k \mathbf{x}_{[j],k}^f \right) \quad (17)$$

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

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 $\tilde{\mathbf{H}}_k \mathbf{M}_{[p],k-1}$ 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.