Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/hess-2020-672-RC1, 2021

Comment on hess-2020-672

Anonymous Referee #1

Referee comment on "Ensemble-based data assimilation of atmospheric boundary layer observations improves the soil moisture analysis in idealized limited-area experiments" by Tobias Sebastian Finn et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-672-RC1, 2021

Review

on the manuscript "Ensemble-based data assimilation of atmospheric boundary layer observations improves the soil moisture analysis" by Tobias Sebastian Finn, Gernot Geppert and Felix Ament

The manuscript "Ensemble-based data assimilation of atmospheric boundary layer observations improves the soil moisture analysis" is devoted to the very important and interesting topic of the comparison between two methods in the land-surface data assimilation (DA) in NWP, belonging to two families: of Extended Kalman Filters and of Ensemble Kalman Filters (namely, Simplified Extended Kalman Filter and Localized Ensemble Transform Kalman Filter). Currently in NWP, Extended Kalman Filters are widely used. However the family of Ensemble Kalman Filters arouses strong interest, due to the development of ensemble systems and using ensemble methods in the upper air DA. Also, there is a hope that ensemble methods will allow to develop the coupled DA, where forecast errors both in the land surface and atmospheric models will be corrected by observations in the surface layer (2-metre observations). Currently in NWP we have so-called weekly coupled systems, because assimilation of the 2-metre observations affects only the land surface model variables (soil moisture and temperature). In the manuscript, the attempt is made to to test feasibility of the ensemble DA methods for the land surface DA.

Unfortunately the manuscript is written so that it is impossible to understand it. The structure of the manuscript is illogical and even confusing. For example, a description of Simplified Extended Kalman Filter is given in Section 2.2.1, which is a subsection of Section 2 "Fully-coupled ensemble data assimilation framework". Simplified Extended Kalman Filter is neither fully-coupled, nor ensemble. Very general descriptions of DA methods and using of ensembles in DA are given (Section 2.2 and beginning of Section 2.2.2), which is not needed. At the same time, for the important description of SEKF and

LETKF (how they are applied? Which variables are in the state vector? What are observational operators and how they are linearized? Which variables are perturbed for the ensemble? How coupling in DA is provided?), only references are provided. Specific parameters of experiments, descriptions of different parts of the DA system, experiment setup, technical details, etc. - everything is mixed and scattered along the text. As a result, it takes too many efforts to read and to understand at least something.

Also, there are many places which are confusing or wrong. Authors are either too inaccurate, or even do not understand what they are doing. Examples are:

1) Lines 18-27. The authors mix physical coupling in the model and coupled DA. No effect of the 2m observations on the soil moisture happens when there is no physical coupling (!) between the soil and atmosphere. No physical coupling in the model, and also no in reality. For example, when there is a strong advection, or in very cloudy and windy conditions. Coupled DA will not help to correct the soil moisture from the 2m observations in these situations. But what coupled DA would do (what we expect from it) in these situations, it would correct atmospheric model lowest layer variables, for example cloudiness or wind.

2) Lines 37-39: "The soil moisture analysis is moreover often estimated in its own daily assimilation cycle in addition to assimilation cycles for the atmosphere on shorter, hourly-like, time-scales. To combine these assimilation cycles into one single cycle, EnKFs are one candidate because of their ensemble-based flow-dependency." Impossible to understand. What is estimated? Why flow-dependency will help to combine cycles? Why flow-dependency is important to soil? In soil, there is no flow.

3) Lines 41-43: "We additionally test with this EnKF setup the hypothesis of hourly updating the soil moisture based on a flow-dependent coupling between land surface and atmosphere." Impossible to understand. What is the hypothesis?

4) Lines 50-51: "Together with TerrSysMP, we perform idealized twin experiments, using the same system configuration for our nature run and our data assimilation experiments." What is "nature run"? I guess, it is without DA. But this is only a guess …

5) Lines 52-53: "With this distilled setup, we are able to isolate the effect of perturbations within the soil moisture on the 2-metre-temperature without having model errors." Absolutely impossible to understand.

6) Lines 74-75: "We further restrict our grid points to single plant functional types (PFTs) to simplify the setup." How it is possible to restrict grid points to plant types???

7) Lines 109-110: " This background forecast is updated at 00:00 UTC based on gridpoint observations at 12:00 UTC and Eq. (3)." Why you use this half day shift?

8) Lines 128-130: Impossible to understand. How Eq.3 can be used to estimate ensemble weights? There is nothing about weights in it. How the observation operator can be linearized around the ensemble mean? Linearization is calculating of a derivative. How the ensemble mean helps to calculate a derivative?

9) Lines 134-135: "These experiments are thus model-error-free ..." - no, they are initial state error free. Model errors remain.

10) Line 153: "Our data assimilation framework ..." - what is data assimilation framework? The code?

11) Line 155: " … the background and first guess are read-in as output files from the models" - what is the difference between the background and first guess in your case?

12) Line 158: "We define a nature run (NATURE) as our truth in this study and to get our 2-metre-temperature observations." Impossible to understand. You define a run without DA as a truth? How it can be?

13) Lines 168-169: "Our only perturbations in the atmosphere and soil are a result of initial soil moisture and soil temperature perturbations." Impossible to understand. You have no perturbations in the atmosphere?

14) Lines 169-170: "A single run with a similar model configuration and a spin-up of 6 years builds the foundation for our initial conditions in the atmosphere and soil." A model configuration similar to what? Why do you need so long spin-up?

15) Lines 171-173: "As horizontal correlation function, we use a truncated Gaussian kernel with a standard deviation of 14 grid points (14 km) and a truncation radius of 42 grid points. The same type of truncated Gaussian correlation is used in vertical dimensions with a standard deviation of 0.5 m and a truncation after 1 m." How a dimension of a standard deviation may be "grid points" or "metres"??? The dimension of a standard deviation is the same as the dimension of the random variable, for which the standard deviation is defined. What is your random variable so that its standard deviation is defined in grid points???

I stop here. Too much work is needed, for at least approximate understanding.

My recommendation for the authors is to rewrite the manuscript totally, to elaborate the text and to re-submit it. Usually for these kind of manuscripts, the plan is the following:

1. Introduction. Why this study happens? What is the purpose?

2. Observations. Which observations are used for assimilation? Which observations are used for verification (or validation)?

3. Methods. Description of the model and DA methods (including observation operators), and how they are applied.

4. Experiment setup. Description of the domain and specific parameters of the DA system (or, other important parameters).

5. Results. Description of results, including criteria and methods, how to estimate them. What is better, what is worth? How verification (or validation) is performed?

6. Discussion. What is the applicability of the results? What are limitations of the study?

7. Conclusions. General conclusions and suggestions for future studies.