Response to interactive comment from Anonymous Referee #2

Authors responses are shown in blue.

General comments

In this study, the authors test different particle filer setups for jointly assimilating a set of snowpack variables, such as snow depth, SWE and snow surface temperature. The study is a valuable contribution to previous studies, which have assessed the performance of the particle filter for the assimilation of only one snowpack variable in most cases. However, the study needs some improvements before final publication.

On behalf of all authors, we thank Anonymous Reviewer #2 for his/her detailed and relevant suggestions, which have allowed us to significantly improve our manuscript.

Four important issue are:

- The authors only use 100 particles when testing the performance of the filter. In some situations, such a low number of particles might give good filter performance. However, in the case of multivariate assimilation of several variables, more particles may be needed. Therefore, I would urge the authors to test the sensitivity of the filter performance by varying the number of particles. The authors should also present results showing the effective sample size after each update in order to test whether the number of particles is sufficient.

We would like to thank the reviewer for this remark of key importance. Actually, we did not test the system sensitivity to the ensemble size, even though in a multivariate DA application this critical issue need to be addressed. Therefore, as suggested, we have performed a further experiment (called nP_Exp, current Sect. 3.5) with the aim of assessing the effective ensemble size, the ensemble spread (current Figure 11) and the performance of the multivariate DA scheme (current Table 7) as the particles number increases: 100-, 200-, and 500-particles. This experiment considers a sample of one randomly-chosen winter season for each analysed experimental site. As shown and explained in the manuscript, the results generally do not show a significant system sensitivity to the ensemble size.

- I could not find any information about the uncertainty of the different measurement. The specification of the observation uncertainties is critically important for the filter behavior and should be reported.

We have considered the following observational uncertainties: 2°C for the surface temperature; 10 mm for the SWE; 0.15 for the surface albedo; 0.05 m for the snow depth; 50 kg/m^3 for the snow density. As required, we have reported this information in the current Table 4.

- Some of the figures contain too much information, foremost figure 4, 6 and 10. The large number of results shown in these figures makes it hard to see which filter setup performs best. This is further complicated by the different axes limits used in the figures (see for example the performance metric NER for CPD and SWE in figure 4, 6 and 10). Overall, I think the presentation of the results would improve by removing some of the performance metrics. The conclusions from this study may also be clearer if the authors could summarize their results in fewer graphs.

Thank you for this useful remark. We definitively agree that the comparison among the multivariate DA configuration was not clear and quite hard to assess. With the aim of ensuring a more concise and effective presentation of the results, we are proposing to replace the previous 4 statistical indices, namely Correlation coefficient, RMSE, Efficiency and Net Error Reduction, with only 2 evaluation metrics. The first one is the Kling-Gupta Efficiency (KGE) coefficient, a deterministic metrics allowing to jointly take account of the correlation coefficient, an estimate of the relative variability between simulated and observed quantities, and a measure of the overall bias. We have replaced Figures 4, 6, and 10 with the current Figure 7, which strictly compares the multi-year KGE values resulting from all the experiments. The second newly-introduced evaluation metrics is an ensemble-based probabilistic score, namely the Continuous Ranked Probability Skill

Score (CRPSS), whose values are listed in an overview table ensuring a quick comparison among the different DA configurations (current Table 6).

- First, the result sections contain more discussions about methods, rather than presentations of their results and quantitative comparisons between them.

This remark has allowed us to significantly improve the readability of our manuscript. We have revised the results section by properly separating the description of methodology from the results discussion. We have introduced a new section, namely Sect. 2.5.2, which presents all the assimilation experiments, also listed in the current Table 5. Furthermore, with the aim of make the manuscript more consistent, we are proposing a further section, namely Sect. 2.5.3, focused on the detailed description of the control open loop simulations (ex-Sect. 3.1.1).

In detail, we have moved:

- the beginning of Sect. 3.1.2, p. 11 l. 19-22 in Sect. 2.5.2 to introduce the first experiment [M_Exp], namely the DA simulations with the perturbation of meteorological forcing;

- the beginning of Sect. 3.2, p. 13 l. 8-11 in Sect. 2.5.2 where the second experiment [MP_Exp(1)] is described, namely the DA simulations with the perturbation of meteorological forcing and model parameters;

- Sect. 3.2.1, p. 13 l. 13-28, namely the preliminary analysis of model parameters, in Sect. 2.4.2;

- Sect. 3.2.1, p. 13 l. 29-30 in Sect. 2.5.2 with the aim of improving the consistency of our manuscript;

- Sect. 3.2.2, p. 14 l. 2-6 in Sect. 2.5.2, where the second experiment [MP_Exp(1)] is described, namely the DA simulations with the perturbation of meteorological forcing and model parameters;

- Sect. 3.3.1, p. 15 l. 23-25 in Sect. 3.4, namely the section focused on the fourth experiment [MPP_Exp];

- Sect. 3.3.1, from p. 15 l. 25 to p. 16 l. 7, in Sect. 2.5.2, where the fourth experiment [MPP_Exp] is described, namely the DA simulations with the additional snow density model.

Furthermore, we propose a new numbering of the third Section “Results and Discussion”:

- Section 3.1: **Multivariate DA simulations with perturbed meteorological input data**

- Section 3.2: **Multivariate DA simulations with perturbed model parameters**

- Section 3.3: **Sensitivity analysis of the multivariate DA scheme to the SWE measurement frequency**

- Section 3.4: **Multivariate DA simulations with proxy information of snow mass-related variables**

- Section 3.5: **Sensitivity analysis of the multivariate DA scheme to the ensemble size**

Second, the authors often states that one filter setup performs better than another setup. However, how large those improvements are is not presented in numbers between the setups. It is therefore very hard to judge whether the simulation results actually improved.

We have substantially revised the presentation of the experimental results in Section 3, with the aim of improving and making easier the quantitative comparison among the different multivariate DA configurations, through fewer and more comprehensive evaluation metrics (current Figure 7 and Table 6).

Specific comments

Abstract: I think the abstract lacks clear “take home messages”. What are the most important results and conclusions obtained in this study?

Thank you for this remark. Actually, the reference to the main results were completely lacking in the Abstract. Therefore, we have added a brief overview of the most important results with the aim of underlining the key conclusions.

Page 2, Lines 1-8: It is also possible to include observation uncertainties using the optimal interpolation scheme.

We thank you for pointing out this relevant mistake. We have revised this short paragraph.

Page 2, Lines 11-15: I think the Enkf was not invented "with the aim of overcoming the inaccuracy of the linearization procedure", but to avoid the need for linearization of the system equations, which in many cases is impossible or simply unfeasible.

Thank you for this remark. We have accordingly revised the manuscript.

Page 2, Lines 24-32: Please state the study goals in more detail using, for example, research questions or hypothesis. Perhaps remove the summary part stretching from line 27 to 32.

As suggested, we have introduced research questions, which definitively improve the readability of this paragraph. We propose to maintain the brief summary providing a quick overview of the manuscript.

Section 2.1: The Torgnon site description includes information about the measurement equipment, whereas the other site descriptions lack this information. I think it would be better to present the same amount of information for each of the field site. If including information about the measurement equipment for all field sites, perhaps better add a table to the paper with this information. Furthermore, the Torgnon site description includes some numbers about climatic conditions. Such information should be included for the two other field sites as well.

We agree that the experimental sites were not homogeneously described. We would prefer not to go into detail of the measurement equipment of each experimental site, since it is not the focus of this Section, aiming at a more general description. Therefore, we have removed all the information on the instrumental equipment at the Italian site. As required, we have also added information on the climatic conditions at the French and Swiss sites.

Page 7, Lines 26-28: I do not understand this part of the sentence: "a resampling procedure is frequently introduced to restore the sample variety through a Markov chain chaotic Monte Carlo". What is a "Markov chain chaotic Monte Carlo"?

The correct wording should be "Markov Chain Monte Carlo" (MCMC), a widely-used method to probe the posterior probability. We have accordingly updated the text.

Equation 8: The effective sample size should be calculated using the square root of the weights.

Thank you for this remark, there was actually a mistake. We have properly revised the Equation 8.

Section 2.3.2: What uncertainty was used for the different observations? This information is essential and must be included in the manuscript.

As required, we have included this information in Table 4, where the assimilated variables are listed.

Page 8, Lines 19-20: Why was not longwave radiation perturbed?

The longwave radiation is not perturbed since it is not a model input. Indeed, both the longwave radiation terms (i.e. incoming and outgoing components) are estimated through the Stephan-Boltzmann law. The outgoing term is calculated as a function of the surface temperature of snow or soil in snowy or snowless conditions, respectively. The incoming component is estimated as a function of the air temperature. While the emissivities of snow and soil are considered as constant model parameters, the air emissivity is time variant and it is evaluated according to both wind speed and air temperature.

Page 9, Line 14: I do not understand this sentence: "Therefore, tuning parameters are properly set to guarantee a significant variance of the parameters distribution."

We have used the approach proposed in Moradkhani et al. (2015), where the authors defined a small multiplicative coefficient (tuning parameter) to tune the variance of the random noise used to perturb the model parameters. Since in our study the parameters perturbation has been introduced with the main aim of enlarging the ensemble spread, we have properly defined these tuning coefficients in order to ensure a significant spread of the ensemble of the model parameters by still preserving the physical consistency. We have realised that this was a too specific and technical consideration. Therefore, we have revised the sentence.

Section 2.4.2: How are the parameter values perturbed? By additive or multiplicative noise?

After the resampling procedure, the ensemble of the model parameters is restored by perturbing the parameters through an additive noise. We have introduced this lacking information in the manuscript (Sect. 2.4.2).

Page 10, Lines 10-12: Perhaps remove: "The SWE is one of the most relevant snow-related quantities from a hydrological point of view, since its accuracy in estimate strongly impacts discharge simulations".

We have shortened and modified this sentence.

Equation 11 and 12: These two equations are probably not needed since the two metrics are very common.

As previously explained, these two equations have been removed since we are proposing new evaluation metrics.

Page 12, Lines 1-21: In this part of the manuscript, I think the authors are mainly discussing filter degeneracy, which is a well-known problem for these kind of applications, in particular when the dimensions of the observation space is high. Please shorten this general discussion by citing relevant literature (e.g. Ades et al., 2013), and provide results more specific to the actual study. Whether such a degeneracy occurs can be assessed by either calculating the efficient sample size, or qualitatively by plotting the time series of the particle spread after assimilation. It would probably be good to include one or both of those analyses to the result sections.

Ades, M. and P. J. Van Leeuwen, 2013: An exploration of the equivalent weights particle filter. Quartely Journal of Meteorology, 139, 820-840.

Thank you for this useful remark and the suggested reference. We have shortened this part discussing the filter degeneracy. We have added Figure 8, which compares M_Exp and MP_Exp by showing both the particles spread and the effective filter updating of SWE simulations at an assimilation time step. As regards the efficient sample size, we have restricted this analysis within the assessment of the system sensitivity to the ensemble size ([nP_Exp], current Sect. 3.5).

Page 14, Line 8: In many places throughout the manuscript the authors refers to "parameters resampling" or similar terminology. I do not understand this terminology, and I am not sure it is a correct since particles are being replicated or terminated in the resampling step, and parameter values are only affected indirectly. Please consider rephrasing.

We have used "parameters resampling" with the aim of stressing that the particles are here resampled not only in model state-space, rather in the state-parameter space. However, with the aim of improving the manuscript readability, in most of cases we have replaced "parameters resampling" with "parameters perturbation".

Page 14, Lines 8-10: Please add some quantitative measures on how large this improvement actually is.

We definitively agree that it was hard to compare the experiments. As previously explained, we have introduced a new ensemble-based evaluation score, namely the CRPSS, to quantitatively assess each DA configuration (Table 6).

Page 14, Lines 10-11: How are the model parameters better estimated? What are the best values of the parameters?

In this sentence we mean that since the particles resampling is here performed in the state-parameters space, it is possible to better take account of the parameters seasonality, rather than assuming constant parameters values.

Page 15, Lines 8-20, including Figure 8: The spread between the particles in the figures seems very small, indicating sample impoverishment. I suspect that the number of particles is not high enough for these kind of experiments, or that a more appropriate filter technique for high dimensional problems should be used requiring fewer particles. Please analyze whether sample impoverishment is occurring or not, and provide appropriate results from such an analysis in the manuscript.

The current Figure 10 (previously Figure 8) shows the time series of the mean of both the snow depth and SWE ensemble simulations. In this Figure the ensemble envelop is not shown. Actually the lack of this information can lead to misleading considerations. Therefore, we have better specified the caption.

Page 16, Lines 9-11: Provide quantitative results on how much the simulation improves.

As previously explained, we have newly introduced the CRPSS score, which allows to more properly assess and compare the different multivariate DA configurations (Table 6).

Page 16, Lines 9-20: What observation uncertainty was assigned to the "proxy information of snow mass-related variables"?

We have maintained the same observational uncertainties. Indeed, when increasing the uncertainty of the indirect estimates of the snow mass-related variables (i.e. SWE and snow density) the benefit of assimilating this proxy information is almost nullified due to unbalanced uncertainty values among the assimilated variables.

Page 16, Line 30: I think it should be "perturbations of parameters" instead of "parameter resampling" as mentioned above.

In most cases we have replaced "parameters resampling" with "parameters perturbation".

Conclusions: The conclusions mainly lists problems with the SIR-PF for the current application. In addition, I would like to know the answer to questions like these: What filter setup worked best for the current application? Does the filter work better for sites with low (CDP) or high (WFJ) snow amounts? What assimilation frequency worked best? Such information is currently missing in the conclusions.

Thank you for this useful remark. We have revised the Conclusions to address these missing considerations. We have better pointed out what filter setup works best and what the main limitations are at the analysed sites, according to their local features.

Technical corrections

Page 1, Line 10: Perhaps remove "multivariate Sequential Importance Resampling".

If the reviewer agrees, we would prefer to keep this sentence.

Page 4, Line 34: Probably wrong reference: Wever, 2015?

We are referring to “Wever, N., Schmid, L., Heilig, A., Eisen, O., Fierz, C., and Lehning, M.: Verification of the multi-layer SNOWPACK model with different water transport schemes. The Cryosphere, 9(6), 2271-2293, 2015”.

Page 10, Lines 15-21: All variables are not explained (e.g. Exp, Obs). It is pretty clear what they mean, but I think for completeness they should be described.

We have removed all these equations.

Page 12, Line 14: Change to: “Firstly, it is intended to properly identify the parameters affecting the model simulations most” or something better.

We have accordingly updated the text “Firstly, it is intended to properly identify the parameters mostly affecting the model simulations”.