Response to interactive comment from Referee #1 (Matthieu Lafaysse):

Authors responses are shown in blue.

General remarks

- I think that this manuscript is a very significant contribution in the field of data assimilation for snowpack modelling. The originality of this paper comes from the multivariate assimilation in the context of the particle filter algorithm. Another added value is the multi-sites application whereas recent applications of the particle filter in snowpack modelling were only focused on one specific site. The multivariate assimilation exhibits some promising advantages but also some discrepancies and challenges which have to be accounted for in the development of such systems. The paper gives a very interesting overview of these positive and negative effects, and their links with the model structure and with the frequency of available observations.

On behalf of all authors, we thank Matthieu Lafaysse for his detailed and relevant suggestions, which have allowed us to significantly improve our manuscript.

- The introduction gives a very good overview of the position of this work among state of-the-art methods.

- I have the feeling that the structure of the paper could be a bit improved before publication in two ways:

1/ First, the results section is a bit too long because it includes some details of the methodology itself which should be described in section 2. This is especially the case because the section already includes both results description and discussion. In particular, the beginning of sections 3.1.1, 3.1.2, 3.2.1, 3.2.2 and 3.3.1 introduce many methodological elements which could be detailed in section 2. I would suggest a paragraph 2.6 describing all the assimilation experiments and their objective. Thus, the presentation of results can become more concise.

We thank you for this useful remark, which allows us to markedly enhance the readability of our manuscript. We have revised the results section by properly separating the description of methodology from the results discussion. As you suggested, we have introduced a new section, namely Sect. 2.5.2, which presents all the assimilation experiments, listed in the current Table 5. Furthermore, with the aim of making the manuscript more consistent, we propose a further section, namely Sect. 2.5.3, focused on the detailed description of the open loop simulations, that are the reference ensemble simulations, as suggested in your major issues.

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**

2/ Then, the authors should better emphasize the lessons of their work for current and future developments of data assimilation in snowpack modelling systems, in a more general point of view than their particular study. This could be done either by introducing a dedicated discussion section either by adding complementary informations and perspectives in the conclusion. For example, the challenge of spatialization at larger scales should be mentioned because it will be a major issue for hydrological modelling. Then, can the authors give general recommendations for the implementation of data assimilation algorithms further than their particular case? From their results, do they recommend to always include parameter perturbations? Do they recommend to include parameter perturbations this way or to test other methods? Do they recommend to apply restrictions in terms of availability of observations to decide to assimilate a given variable? Do they recommend a minimal model structure to decide to assimilate some specific variables?

We are grateful to the reviewer this remark. Actually we discussed the results focusing on our analysed case studies since it is a first attempt to implement a multivariate PF scheme in the framework of snow modelling. However, as you suggested, we have introduced more general considerations and recommendations, according to our experience. We have stressed the importance of jointly perturbing both the meteorological data and model parameters, especially when dealing with spatialized systems. Furthermore, we have highlighted the potential of using indirect estimates of model state variables with the aim of limiting the system sensitivity to the measurement frequency.

Major issues

- My main concern is the fact that the skill of data assimilation is assessed by the comparison of deterministic scores between ensemble simulations including data assimilation and the deterministic reference simulation which is forced by in-situ meteorological measurement. However, in the real world, the quality of the meteorological forcing will be much lower that the quality of the forcing at the three stations of Col de Porte, Weissfluhjoch and Torgnon. Therefore, it makes sense to use perturbations which are not really representative of the uncertainty of these meteorological dataset but more typical of common meteorological errors. Although it is not clearly said in the paper (section 2.4.1), this is what is done here because the error statistics of Charrois et al, 2016 and Magnusson et al, 2017 come from a comparison between a meteorological analysis and in-situ observations. These errors do not represent the observation error, they represent the meteorological analysis error. As a consequence, data assimilation is expected to reduce the meteorological error introduced in the forcing. But it is very demanding to expect from data assimilation to come back to results of the same quality as simulations forced by in-situ measurements when perturbations higher than the observation uncertainty are introduced. There are several options to solve this issue: Option 1) using lower perturbations consistent with the meteorological forcing. The main limitation will be a low spatial transferability of the results as very few stations provide this quality of meteorological data. Option 2) using a meteorological forcing of lower quality more consistent with the perturbations. This option would require to run again all simulations. Option 3) changing the evaluation metrics to provide a comparison of skill between 2 ensembles, the first one with the perturbations but without assimilation and the second one with assimilation. This option does not imply to change the simulation runs, it only requires to compute new evaluation metrics. Therefore, I would recommend this option for this work. The easiest way will be to keep the same metrics but to apply them to the ensemble without assimilation instead of the reference run without perturbation. Thus, the blue points in Fig. 4, 6, 10 will be replaced by a boxplot which can be compared with the red boxplot (ensemble with assimilation). Note that it would also be possible to use ensemble metrics instead of deterministic metrics. For example you could compute the Continuous Ranked Probability Score (CRPS) of the ensembles with and without assimilation.

This remark is definitively of key importance and we would like to thank the reviewer for pointing out this issue. As suggested, we have chosen the Option n°3, namely we have considered the probabilistic open loop run as the control one. Therefore, we have compared the ensemble simulations resulting from each experiment with the ensemble open loop simulations. With the aim of improving the comparison among the experiments, we are also 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).

Furthermore, with the aim of improving the clarity of our manuscript, we have also modified Sect. 2.4.1, p. 9, l. 5-7:

"Even though this approach ensures to take account of the actual meteorological errors affecting the quality of the model predictions, the main limitation of this procedure is the lack of correlations among the perturbed forcing variables, which does not ensure their physical consistency (Charrois et al., 2016)."

- The second major issue is the fact that the scores are presented in a very high number of subplots (Fig. 4, 6, 10) which are very small. The comparison of the different experiments is difficult with these figures due to the lack of more synthetic metrics allowing a quicker comparison of the experiments. It is probably interesting to see the interannual variability of the scores for one example but I do not think that this is necessary for all scores, sites, and experiments. It is impossible to analyze in details all the scores provided in these 3 figures. Page 16, line 11, clearly the authors do not need all the metrics of Figure 10 for such a general conclusion! I think the authors should try to present multi-year scores in a synthetic table allowing a quick and representative overview of the model skill for the different experiments.

As previously explained, we have introduced more synthetic evaluation metrics, namely the KGE and CRPSS scores, to allow a quicker comparison among the experiments results. Actually, we agree that the interannual variability of the scores is not of significant relevance, since the main goal is to assess the overall performance of each multivariate DA configuration. Therefore, in place of Figures 4, 6, and 10, we are proposing the current Figure 7, which shows multi-year KGE scores. Moreover, we have listed the resulting CRPSS indices in the current Table 6.

Other remarks

Page 1 line 24: It would be useful to also mention that snow models are based on uncertain parameterizations and parameters (Essery et al, 2013; Lafaysse et al, 2017). Thus, it would become more natural to introduce further the perturbations of model parameters.

In this sentence we list some of the main real-world phenomena and causes which make it difficult to model the snowpack dynamics. Here we are not including modelling issues. Therefore, we are proposing to introduce the parameters uncertainty directly at the beginning of the current Section 2.4.2, namely the Section focused on the parameters perturbation.

Page 1 line 28: It is not obvious that there is a link between the complexity and the skill of the data assimilation algorithm.

We agree with this remark. We have revised this sentence.

Page 1 line 29: "they allow to process" → they allow taking benefit from

We have accordingly revised this sentence.

Page 2 line 4: snow models (plural)

We have accordingly revised the text.

Page 2 lines 5-8: Optimal interpolation also allows accounting for observation uncertainty.

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

Page 2 line 15: EnKF can also be based on ensembles obtained from other methods than the Monte-Carlo approach.

We have accordingly revised this sentence.

Page 2 lines 21-30: The authors could also add that in the context of more complex models, EnKF is also complicated by the need to compute averages of the snowpack profiles. This can be a challenge for the models based on a lagrangian discretization with a variable number of snowpack layers.

Thank you for this remark. We have added this further consideration.

Page 2 line 33: "the full prior density" → coming from the ensemble

We have revised this sentence.

Page 3 lines 18-20: Please add "at the local scale" because these conclusions might not be true in spatialized simulations.

We have accordingly revised this sentence.

Page 4, lines 9-10: I do not agree that instrumental biases are representative of observation uncertainties. Even on these well-maintained sites, environmental errors are the prevailing source of uncertainty. Therefore, the instrumental accuracy provided by manufacturers does not provide a good assessment of observation error. For example, the radiation sensors are generally more affected by environmental issues (hoar or snow on the sensor) than for instrumental accuracy. Similarly, precipitation measurement is mainly affected by undercatch in case of wind.

We agree that this statement was improper. Therefore, we have removed this sentence.

Page 4, line 13: "all the requirements" → to force and evaluate a snow model

We have accordingly revised the sentence.

Page 4, lines 23-28; page 5, lines 2-7: I think that it is not necessary to provide so many details about the available observations at Col de Porte and Weissfluhjoch. The observations which are not used in this paper (temperature profiles, ground temperatures, liquid water content, runoff, etc.) do not need to be described.

Thank you for this suggestion. We have removed extra information on observations that are not used in our study.

Page 5, line 30: Can you detail what represent the 2 distinct layers? I assume that there is a surface layer? Does it have a fixed depth?

Yes, there is a surface layer. The thickness of the snow layers can vary and no limit is set for any of them. The snow distribution between the two layers is ruled by the empirical parameterization, which allows maintaining the surface layer thinner than the underlying one. This approach is intended to allow to consider the top layer temperature as an acceptable approximation of the skin temperature, whose measures can thus more efficiently assimilated. We refer to Piazzi et al. (2018, accepted) for the detailed description of the snow modelling scheme.

Section 2.2 Can you explain how the energy balance is computed without the availability of a longwave radiation forcing?

The longwave radiation is not a model input. 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 6 line 27: model input vector → meteorological input vector

We have accordingly revised the sentence.

Page 6 line 32: why do you prefer here the word "noise" to "error"? I think it would be more accurate to talk about observation error.

We agree. We have replaced "observational noise" with "observation error".

Page 7 line 4: missing space after t-1

We have added the lacking space.

Page 7 line 12: This statement could be more general. Indeed, as mentioned before, the Monte Carlo sampling is not the only method to build an ensemble.

Thank you, we have revised this sentence.

Page 7 lines 17-18: I followed the formalism until here but I do not fully understand the sentence "Particles are drawn from a known proposal distribution according to the ´ Sequential Importance Sampling approach". Can you clarify this part so that it can ˙ be understood without reading the references associated with the SIS approach?

Actually a further binding sentence was lacking. Therefore, we have revised the text by extending the explanation: "**It is noteworthy to consider that the direct sampling of particles from the posterior density is generally difficult, since its distribution is often non-Gaussian. Therefore, particles […]**".

Page 8 line 6: I think that the reference to Fig. 2 in the text does not take all the benefit of this figure to clarify the methodology. I would suggest to refer separately to the different subplots in the text to be more illustrative. Can you also comment the reasons which explain the slight differencies between Fig 2b (weights) and 2d (number of resampled particles)?

We think that the main reason of these slight differences can result from a combined effect depending on both the drawn sample and the shape of the empirical cumulative distribution. We have added the references to each single subplot.

Page 8 equation 10: Can you explain by words the practical implication of this equation?

This equation describes how the particles weights are updated at each assimilation time step, namely by evaluating the value of the likelihood function, which is assumed to be a multi-dimensional Gaussian distribution. The likelihood value of each particle depends on how it is placed with respect to all the available observations.

Page 8 line 23: “additive stochastic noise” Can you detail the process applied for precipitation? I assume it is probably not possible to apply directly an additive noise in that case? Is there a different treatment between occurrency and intensity?

Following the approach proposed by Magnusson et al. (2017), for the precipitation, as well as for wind speed, we assumed an additive stochastic noise having a lognormal distribution, and this is now specified in the text. We perturbed only the precipitation intensity.

Page 9 line 1-2: I agree with this remark. However, following my first major remark, the perturbations used in this study are not representative of the error of the in-situ measurements at Col de Porte.

Thank you for this comment. According to your main remark, in Sect. 2.4.1 we have better specified that the perturbations are not representative of the error of ground-based measurements, rather of the common meteorological analysis error.

Page 9 lines 4-6: The perturbation of model parameters is introduced through a very "mechanical" point of view for the data assimilation algorithm. I think it would be useful to remind that errors exist in the snow model itself and that it is natural that perturbations of the meteorological inputs are not sufficient to cover all uncertainty.

We agree with this remark. Actually we introduced and discussed the perturbation of model parameters only with regard to its potential within the DA framework. Firstly, we are proposing to modify the title of the Section 2.4.2, “**Perturbation of model parameters**”. Secondly, we have added an introduction sentence at the beginning of this Section.

From page 9 line 29 to page 10 line 3: Are the authors aware that weekly measurement of bulk density are available at Col de Porte at 3 different places in the plot? These data should be preferred for data assimilation than a computation from SWE and snow depth. Indeed, very unrealistic values are obtained with such a computation because the spatial variability in the plot is responsible for a different accumulation between both sensors.

Thank you for this interesting remark. Actually we knew that weekly measurements of bulk density are available at the Col de Porte. We have chosen to use indirectly derived estimates of snow density since they can be evaluated with a daily frequency, which is an interesting benefit to test the system sensitivity to difference in the measurements frequency, with respect to the other sites. We have taken account of the higher uncertainty of the density estimates with respect to the other variables. Before using these snow density estimates, we have qualitatively assessed the consistency of their values, neglecting those deemed unreliable.

Page 10 Line 14 Can you provide the time step and the hour used for the surface temperature?

The surface temperature is updated every 15 minutes, according to the model integration time step. The observations are assimilated every 24 hours, at 11 a.m.

Page 10 Line 16 "snowless periods are neglected" How do you define the snow free periods? Is it only based on observations? This choice can lead to eliminate some data for which some particles do have snow and to include data for which some particles do not have snow. This is a usual issue in the evaluations of a snowpack model so please be accurrate on that point.

We have estimated the melt-out date of each winter season according to the observations. For each station, the snowless periods start after the last melt-out date evaluated over the whole dataset, with the aim of being conservative as much as possible.

Page 10 line 28: "spurious trends" → unexpected biases

We have accordingly revised the text (current Sect. 2.5.3).

Page 10 lines 25-28: The goal of section 3.1.1 should also be to check if the perturbations are able to realistically depict the uncertainty of snow simulations.

We have added this further consideration (current Sect. 2.5.3).

Page 11 line 5: It would be interesting to notice that despite unbiased perturbations, the control run is not identical to the ensemble mean.

We have added this further consideration (current Sect. 2.5.3).

Page 11 line 6: A more comprehensive assessment of the fact the control run is included in the ensemble spread would be to use Talagrand rank histograms over the whole period to ckeck that the control run has a random position in the ensemble. Note that it would be even more informative to check if the observation is included in the ensemble spread with a random position. This would be very useful to strengthen the discussion between lines 1-21 of page 12.

As you suggested, we have analysed the Talagrand rank histograms to assess whether both observations and the deterministic control simulations are included within the ensemble spread. We have introduced the current Figure 6 showing the Talagrand histogram of the SWE ensemble open loop simulations throughout the overall CDP dataset.

Page 12 line 7 "their measures" → the measure of the corresponding variable

We have accordingly revised this sentence.

Page 13 line 15 "uselessly" Can you be more explicit? Larger computation requirements without a significant improvement of the spread and further of the filter efficiency.

We have better specified this sentence.

Page 13 lines 18-20: It is unclear if more parameters were also tested in a preliminary sensitivity analysis and which specific metric was used to select the parameters to disturb. Note that the parameters could also be chosen a priori based on previously published sensitivity analyses of other snowpack models.

The preliminary sensitivity analysis involved also other model parameters. We have selected only those a significant impact on the simulations. According to the approach described in Piazzi et al. (2018, accepted), we use the KGE coefficient as evaluation metric. We have added the reference in the text.

Page 13 lines 22-23: The albedo parameters could also have an impact of snow mass during the melting season.

Thank you, actually we omitted this information. We have added this further consideration.

Page 13 line 30: Can you provide a better description of Fig 5 in the text? The fact that the spread of viscosity is increasing in the melting period should be noticed. Does it suggest that melting issues in the model are compensated by this parameter?

Yes, the gradual increase of the ensemble spread (current Figure 3) can definitively suggest an offsetting effect throughout the melting period. Thank you for this interesting remark. We have added these considerations in the text (current Sect. 2.4.2).

Page 15 line 28: Can you give more details about the new density function and how it differs from the original relationship between SWE and snow depth in your model?

According to the approach proposed by Jonas et al. (2009), the snow density is estimated through an empirical parameterization relying on the reported 4 main factors. The authors defined through a linear regression the two coefficients [b, a] best fitting an extended observational dataset of snow depths and snow densities. Therefore, the snow density (ρ_{estim}) is evaluated as [$\rho_{estim} = a \cdot SD_{obs} + b$], where SD_{obs} is the observed snow depth. The resulting SWE estimate SWE estimation (SWE_{estim}) is retrieved as [$SWE_{estim} = SD_{obs} \cdot \rho_{estim}$].