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 [$SWE_{estim} = SD_{obs} \cdot \rho_{estim}$].

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”.

A Particle Filter scheme for multivariate data assimilation into a point-scale snowpack model in Alpine environment

Gaia Piazzi[1], Guillaume Thirel[2], Lorenzo Campo[1], Simone Gabellani[1]

[1]CIMA Research Foundation, Savona, 17100, Italy

[2] ~~Hydrosystems and Bioprocesses~~ Catchment Hydrology Research Group, HYCAR Research Unit ~~(HBAN)~~, Irstea, Antony, 92160, France

Correspondence to: Gaia Piazzi (gaia.piazzi@cimafoundation.org)

Abstract. The accuracy of hydrological predictions in snow-dominated regions deeply depends on the quality of the snowpack simulations, whose dynamics strongly affects the local hydrological regime, especially during the melting period. With the aim of reducing the modelling uncertainty, data assimilation techniques are increasingly being implemented for operational purposes. This study aims at investigating the performance of a multivariate Sequential Importance Resampling - Particle Filter scheme designed to jointly assimilate several ground-based snow observations. The system, which relies on a multilayer energy-balance snow model, has been tested at three Alpine sites: Col de Porte (France), Torgnon (Italy), and Weissfluhjoch (Switzerland). The implementation of a multivariate data assimilation scheme faces several challenging issues, which are here addressed and extensively discussed: (1) the effectiveness of the perturbation of the meteorological forcing data in preventing the sample impoverishment; (2) the impact of the parameters perturbation~~resampling~~ on the filter updating of the snowpack state; ~~(3)~~ the system sensitivity to (3) the frequency of the assimilated observations and (4) the ensemble size.

The perturbation of the meteorological forcing data turns out to be generally not sufficient to prevent the sample impoverishment of the particles sample, which is highly limited when jointly perturbating key model parameters. The parameters perturbation proves, however, to sharpen the system sensitivity to the frequency of the assimilated observations, which can be successfully relaxed by introducing indirectly estimated information on snow mass-related variables. The ensemble size is found not to greatly impact the filter performance in this point-scale application.

1. Introduction

Snow-dominated areas play a distinctive role in water supply in terms of soil moisture, runoff, and groundwater recharge (Vivoroli et al., 2007; Dettinger, 2014). The knowledge of the spatio-temporal distribution of snow cover is therefore of critical importance to several applications (Viviroli et al., 2011; Fayad et al., 2017). When dealing with hydrological predictions in mountain regions, the modelling of snow dynamics is a challenging issue due to complex interactions among site-dependent factors, namely the meteorological forcing (Bormann et al., 2013; Luce et al., 2014), local topography (Molotch and Meromy, 2014; Revuelto et al., 2014), the presence of vegetation and the wind-induced phenomena (Gascoin et al., 2013; Zheng et al., 2016; Quéno et al., 2016).

Recently, an increasing interest focuses on investigating the potentials of Data Assimilation (DA) schemes in consistently improving the model simulations by assimilating ground-based measurements or remotely sensed snow-related observations (Bergeron et al., 2016; Dziubanski and Franz, 2016; Griessinger et al., 2016; Huang et al., 2017).

~~S~~everal DA methodologies have been developed, each one characterized by different performances mainly according to its degree of complexity. The sequential DA techniques are widely used for real-time applications, since they allow taking benefit from ~~to process~~ the observational data as it becomes available and sequentially update the model state. The most basic approach~~es~~ relies~~y~~ on the direct insertion (Liston et al., 1999; Rodell and Houser, 2004; Malik et al., 2012), which promotes the simple replacement of model predictions with observations, whenever available. ~~and the optimal interpolation schemes (Brasnett,1999; Liston and Hiemstra, 2008).~~ Even though this~~ese~~ conceptually simple DA scheme~~s~~ is~~are~~ an attractive method~~s~~, ~~their~~ its implementation within complex, multi-layered snow models is not straightforward, mainly because of possible model shocks resulting from physical inconsistencies among state variables (Magnusson et al., 2017). More advanced are the optimal interpolation schemes (Brasnett,1999; Liston and Hiemstra, 2008), the Cressman scheme (Cressman, 1959; Drusch et al., 2004; Dee et al., 2011; Balsamo et al. 2015) and the nudging method (Stauffer and Seaman, 1990; Boni et al. 2010), allowing to take into account the observational uncertainty, which is a priori defined. At a higher level are the Kalman filters, which are among the most commonly used sequential DA techniques (Kalman, 1960). The standard version of the Kalman Filter (KF) (Gelb, 1974), which relies on the system linearity assumption, was upgraded to the Extended Kalman Filter (EKF) (Miller et al., 1994) allowing to deal with nonlinear dynamic models through a linearized statistical approach (Sun et al., 2004; Dong et al., 2007). With the aim of overcoming the need for a statistical linearization, which can be unfeasible when dealing with strongly nonlinear models~~the inaccuracy of the linearization procedure affecting the filter performance due to possible strong model nonlinearities (Moradkhani, 2008)~~, the Ensemble Kalman Filter (EnKF) has been developed (Evensen, 1994). Unlike the KF and EKF schemes, this method does not require a model linearization since the error estimates are evaluated from an ensemble of possible model realizations, commonly generated ~~using~~ through the Monte Carlo approach (Evensen, 2003). In the recent past, an increasing number of studies on snow hydrology have contributed to confirm the EnKF as a well-performing technique enabling to enhance the accuracy of hydrological simulations by consistently updating model predictions through the assimilation of snow-related observations (Andreadis and Lettenmaier, 2005; Durand and Margulis, 2006; Clark et al., 2006; Slater and Clark, 2006; Su et al., 2008; Durand and Margulis, 2008; Su et al., 2010; De Lannoy et al., 2012; Magnusson et al., 2014; Griessinger et al., 2016; Huang et al., 2017).

Even though the EnKF scheme provides a flexible framework to explicitly handle both observational and modelling uncertainties (Salamon and Feyen, 2009), some constraining assumptions hinder filter performance (Chen, 2003). Firstly, in Kalman filtering the analysis step relies on the second-order moments (Moradkhani et al., 2005). However, because the state variables in stochastic-dynamic systems are modelled as random variables, the involved probability distributions are not supposed to follow a Gaussian distribution (Weerts and El Serafy, 2006). Thus, in strongly nonlinear systems the first two moments are not likely to be sufficient to properly approximate the posterior probability distributions, whose estimates require the tracking of higher-order moments (Moradkhani et al., 2005). Secondly, the EnKF is limited to the linear updating procedure with significant simplification affecting the filter performance. Recently, Piazzi et al. (2018~~7~~) investigated the main limitations in implementing a multivariate EnKF scheme to assimilate ground-based and remotely-sensed snow data in the framework of snow modelling. Furthermore, since the EnKF involves state-averaging operations, the implementation of this DA technique into highly-detailed complex snowpack models (e.g. with varying number of snow layers) is even more challenging, or even unfeasible.

In order to overcome these limiting issues, filter methods for non-Gaussian, nonlinear dynamical models have been developed. These sequential Monte Carlo techniques, also known as Particle Filter (PF) schemes (Gordon et al., 1993), have the main advantage of relaxing the need for restrictive assumptions on the form of the probability distributions, since the full prior density derived from the ensemble is used within the updating procedure (Arulampalam et al., 2002). Thanks to their suitability to better succeed in handling systems nonlinearities, PF schemes are currently garnering a growing attention for snow modelling applications. Leisenring and Moradkhani (2011) compared the performances of common sequential EnKF-based DA methods and PF variants at assimilating synthetic SWE measurements to improve its seasonal predictions and to estimate some sensitive parameters in a small-scale snowpack model. The results suggested that all the

DA techniques succeeded in enhancing the SWE dynamics. Even though PF-based techniques generally revealed a higher accuracy, the resulting bias was comparable with the Kalman filters one. Dechant and Moradkhani (2011) evaluated the PF performance in assimilating remotely~~-~~sensed microwave radiance data to update the states of a snow model. The results showed that the DA scheme allowed to improve simulations of SWE as well as discharge forecasts. Thirel et al. (2013) investigated the implementation of the PF technique to assimilate MODIS SCA data into a physical distributed hydrological model, in order to enhance snowmelt-related stream flow predictions. Both synthetic and real experiments showed clear improvements of model discharge simulations, especially for intermediate values of observation error. Margulis et al. (2015) tested a newly-proposed PF approach to improve SWE estimates when assimilating historical Landsat-derived observations of fractional snow-covered area into a land surface model. This technique has been recently applied by Cortés et al. (2016). Charrois et al. (2016) investigated the performances of the Sequential Importance Resampling PF (SIR-PF) scheme in assimilating MODIS-like synthetic data of optical reflectance into a detailed multilayer snowpack model. The study assessed the impact of the assimilation, which well succeeded in reducing RMSE values on both snow depth and SWE with a resulting reduction of the uncertainty on the snow melt-out date. An even larger bias reduction was achieved by updating the model assimilating synthetic snow depth observations, except for thin snowpack. However, they proved that the joint~~ed~~ assimilation of remotely sensed reflectance and measurements of snow depth can be the best combination to provide a significant improvement of the model simulations at the local scale. Magnusson et al. (2017) found that the assimilation of daily snow depth measurements within a multi-layer energy-balance snow model through the PF scheme resulted in an improvement of the simulations of SWE and snowpack runoff over the whole analysis period. However, model daily runoff dynamics did not substantially benefit from the snow depths assimilation, except during the melt-out period.

In view of the promising performances of PF-based schemes in snow-related univariate DA applications, this study aims at contributing to this research field by ~~addressing~~ investigating the potential of this technique in performing multivariate DA. Main issues of key importance are therefore addressed: (1) How does the PF scheme succeed in consistently updating the snowpack system state by jointly assimilating several in-situ snow-related point data? (2) What are the most constraining limitations in implementing a multivariate PF-based scheme in the framework of snow modelling and how to overcome them? (3) What is the impact of the uncertainties of meteorological data and model parameters on the filter effectiveness? (4) How much does the filter performance depend on the observations availability? What is an effective approach to limit the system sensitivity to the difference in the measurement frequency of the assimilated variables?~~The goal focuses on investigating how the PF scheme succeeds in consistently updating the system state by jointly assimilating several in-situ snow-related point data in a snow dynamic model.~~

Section 2 firstly describes the ~~analyzed~~analysed case studies and the modelling system consisting of a multilayer energy-balance model and the DA scheme, whose main features are discussed. After sketching the experimental design, Section 3 presents and assesses the main results of the experiments on different configurations of the multivariate DA scheme.~~explains the development of the PF-based DA scheme.~~ The main issues hindering the filter efficiency are thoroughly discussed by ~~analyzing~~analysing the impact of the meteorological perturbation, the uncertainty of model parameters and the use of an additional snow density model to reduce the system sensitivity to the in-situ measurement frequency. Lastly, conclusions are outlined in Section 4.

2. Materials and methods

2.1 Case studies

With the aim of testing the snow modelling system at a point-scale, the selection of the case studies has been restricted among pilot experimental sites, where automated weather stations supply meteorological and snow-related measurements of high quality and completeness. This choice has allowed to reliably investigate the filter performance regardless of possible inconsistent measures, since generally the in-situ observations are extensively verified through a quality control and data gaps filling (Morin et al., 2012; Essery et al., 2013; Lafaysse et al., 2017). ~~Since the strategic requirements conditioning the placement of the test sites entail slightly impacting local features affecting the spatial snow distribution (e.g. wind-induced phenomena, slope, exposure) (Wever et al., 2015), the instrumental biases have been deemed as properly representative of the observational uncertainties. Moreover, w~~With the purpose of investigating the snow model sensitivity to various meteorological conditions, measurement sites located at different elevations have been chosen. Moreover, the selection has been limited over the domain of interest, namely the Alpine region. Among the Alpine measurement sites, three snow experimental sites, meeting all the requirements to force and evaluate a snow model, have been selected: Col de Porte (France), Weissfluhjoch (Switzerland) and Torgnon (Italy).

Col de Porte site

The Col de Porte observatory (CDP) is located near Grenoble, in the Chartreuse massif in the French Alps (45°30' N, 5°77' E) at an elevation of 1325 m a.s.l.. This pilot site is placed in a grassy meadow surrounded by a coniferous forest on the eastern side. Snow cover is usually present from December to April, on average (Lafaysse et al., 2017). Nevertheless, during the winter season surface melt and rainfall events can frequently occur at the relatively low altitude of the experimental site. The mean annual precipitation is about 1110 mm of rain and 570 mm of solid precipitation, and the air temperature falls below 0°C generally only during December and March. In-situ meteorological data, at the hourly resolution, include measurements of 2-m air temperature and relative humidity, 10-m wind speed, incoming short- and longwave radiations and precipitation rates. Precipitation phase is manually assessed using all possible ancillary information (Lafaysse et al., 2017). Snow-related observations are provided both at daily and hourly resolution. ~~Weekly manual snow pits enable to collect internal snowpack information (e.g. snow temperature, density profiles, liquid water content). Furthermore, hourly records of temperature and height of vertically-free settling disks are available.~~ Along with weekly manual SWE measurements, since the season 2001-2002 SWE is automatically measured on a daily basis by a ground-based cosmic rays counter. Hourly snow albedo data are estimated through the radiation sensors, as the ratio between incoming and reflected shortwave radiation (Morin et al., 2012). Moreover, measurements of snow surface ~~and soil~~ temperature~~s~~ are hourly available. ~~Snowmelt runoff is hourly estimated through two lysimeters, as well.~~

Weissfluhjoch site

The Weissfluhjoch site (WFJ) (46.82°N, 9.80°E) is located at an altitude of 2540 m a.s.l. in the Swiss Alps, near Davos, Switzerland (WSL Institute for Snow and Avalanche Research SLF, 2015b). This snow experimental site is placed in an almost flat area of a south-easterly oriented slope. At WFJ the snow season generally starts in October/November and lasts until June/July (Wever et al., 2015). The average air temperature exceeds 0°C generally only between May and October and the mean annual precipitation is about 1180 mm of solid precipitation and 556 mm of rain. An automated weather station provides a comprehensive multiyear dataset including measurements of air temperature and relative humidity, wind speed and direction, incoming and outgoing short- and longwave radiation, snow/ice surface temperature, ~~temperature of soil at its interface with the snowpack,~~ snow depth and precipitation (Schmucki et al., 2014). ~~Snow temperatures are measured at 50, 100 and 150 cm above the ground surface. From September 2013 onwards, soil~~

~~temperatures are measured at 50-, 30- and 10-cm depth. Additionally, snowpack runoff data are supplied by a snow lysimeter.~~ The rain gauge ~~and snow lysimeter~~ measures at an interval of 10 min, whereas most other measurements are done at 30-min intervals. ~~Lastly, an upward-looking ground-penetrating radar provides 30-min data on liquid water percolation allowing to monitor the progress of the meltwater front (Schmid et al., 2014).~~ Every two weeks, generally in early and in the middle of each month depending on weather conditions, a manual full-depth snow profile is performed in order to provide measurements of snow temperature and snow density (WSL Institute for Snow and Avalanche Research SLF, 2015a). Snow density and SWE are also manually measured through snow cores.

Torgnon site

The Torgnon site (TGN) (Tellinod, 45°50' N, 7°34' E) is located in Aosta Valley, a mountain region in north-western Italian Alps. The experimental site is subalpine grassland, at an elevation of 2160 m a.s.l. (Filippa et al., 2015). The area slopes lightly (4°) and it is characterized by a typical intra-alpine semi-continental climate, with an average annual temperature of around 3°C and a mean annual precipitation of 880 mm (Galvagno et al., 2013). On average, the snow season lasts from the end of October to late May, when the test site is covered by a thick snow cover (90-120 cm). Since 2008, an automatic weather station provides 30-min averaged records of different meteorological parameters, including air and surface temperatures ~~(HMP45, Vaisala, SI-111 and therm107, Campbell Scientific)~~, incoming and outgoing short- and longwave radiations, and surface albedo ~~(CNR4, Kipp&Zonen)~~, precipitation ~~(OTT Pluvio2, Weighing Rain Gauge)~~, soil water content ~~(CS-616, Campbell Scientific)~~, snow depth ~~(SR50A-L, Campbell Scientific)~~ and wind speed and direction ~~(WINDSONIC1-L, Campbell Scientific)~~. Furthermore, SWE in-situ measurements are available with a sampling resolution of 6 hours for the snow seasons 2013-2014 and 2015-2016 ~~(CS725, Campbell & Scientific)~~. The sensor provides comprehensive SWE measures over a sizeable area of 50-100 m^2, with a resulting lower impact of several local factors (e.g. snow drifting, vegetation). Bi-weekly manual measures of snow density (snow pits) are available during the winter season (3-4 measurements per month, on average).

According to the in-situ meteorological observations, the selected experimental sites are characterized through two of the most ruling climate forcing, namely air temperature and snowfall rate (Lòpez-Moreno and Nogués-Bravo, 2005), together with the snow depth trend (Figure 1).

2.2 Snow model

The snow model relies on a multilayer scheme consisting of two layers of both soil and snowpack. The model provides an estimate of several snow-related variables describing the snowpack state by simulating the main physical processes (i.e. accumulation, density dynamics, melting and sublimation processes, radiative balance, heat and mass exchanges). The explicit energy and mass balances framework requires several input forcing meteorological data: air temperature, wind velocity, relative air humidity, precipitation and incident shortwave solar radiation. While a full description of model is extensively explained in Piazzi et al. (2018~~7~~), some details on model parameterizations are given below.

- *Precipitation phase*: When total precipitation rate is provided, the partitioning between rain- and snowfall is based on both air temperature and relative humidity, according to the approach proposed by Froidurot et al. (2014).
- *Snow compaction*: Snow density is updated considering both the compaction and the destructive thermal metamorphism according to the physically-based parameterization proposed by Anderson (1976).
- *Fresh snow density:* In case of snowfall, the fresh snow density is evaluated as a function of the air temperature (Hedstromand and Pomeroy, 1998).

- *Snow albedo:* With respect to the original version of the model scheme described in Piazzi et al. (201~~7~~8), the empirical snow albedo parameterization proposed by Douville et al. (1995) has been introduced. Following this formulation, the surface albedo (α) is predicted as prognostic variable:

  Albedo for cold snow: $\alpha_s(t+\delta t) = \alpha_s(t) - {\tau_\alpha}^{-1}\delta t$ (1a)

  Albedo for melting snow: $\alpha_s(t+\delta t) = [\alpha_s(t) - \alpha_{min}]\exp(-{\tau_m}^{-1}\delta t) + \alpha_{min}$ (1b)

  Albedo update (snowfall event): $\delta\alpha_s = (\alpha_{max} - \alpha_s)\frac{S_f\delta t}{S_0}$ (1c)

  where:

  S_f is the snowfall rate [kg m^{-2}]; α_{max}= 0.85; α_{min}= 0.5; S_0 = 10 kg m^{-2}; τ_α = 10^7 s; τ_m = $3.6 \cdot 10^5$ s.

  The albedo dynamics is described by a linear decay over time under cold snow conditions (Eq. 1a) and an exponential decay in the presence of melting snow (Eq. 1b). When a snowfall event occurs, the albedo is consistently updated (Eq. 1c).
- *Turbulent heat fluxes:* Sensible and latent heat fluxes are evaluated following the bulk formulation. The atmospheric stability is evaluated as a function of the Richardson Bulk number, according to the empirical scheme of Caparrini et al. (2004).

2.3 Particle filter data assimilation scheme

The PF technique relies on the Monte Carlo approach to solve the Bayesian recursive estimation problem. Consider the state vector (X_t) including all the prognostic variables:

$$X_t = M[X_{t-1}, \theta, U_t, \Omega_t] \quad (2)$$

where *M* is the dynamic model operator, which calls for the ~~model~~ meteorological input vector (U_t), the vector of model parameters (θ), and the model error (Ω_t). Whenever a set of observations is available, the analysis procedure allows to update the a priori state according to the observation vector (Y_t), which requires an observation operator (*H*) enabling to generate the model equivalents of the observations:

$$Y_t = H[X_t, \Psi_t] \quad (3)$$

where Ψ_t is the observation~~al noise~~ error, which is generally assumed to be Gaussian and independent of the model error. The sequential filtering problem aims at finding the maximum of the conditional probability density function (pdf) of the model state $P(X_t|D_t)$, where $D_t = \{Y_t; t = 1, \ldots, t\}$ encompasses all the available observational information on the time step *t*.

Given the posterior pdf at time *t-1* $p(X_{t-1}|D_{t-1})$, it is possible to obtain the pdf of the current state $p(X_t|D_t)$ in two stages, namely the prediction of the prior density $p(X_t|D_{t-1})$ (Eq. 4) and the update of the forecast pdf according to new observations (Eq. 5).

$$p(X_t|D_{t-1}) = \int p(X_t|X_{t-1})p(X_{t-1}|D_{t-1})\, dX_{t-1} \quad (4)$$

$$p(X_t|D_t) = p(X_t|Y_t, D_{t-1}) = \frac{p(Y_t|X_t)p(X_t|D_{t-1})}{\int p(Y_t|X_t)p(X_t|D_{t-1})dX_t} \quad (5)$$

where $p(X_t|X_{t-1})$ is the known transition pdf, $p(Y_t|X_t)$ measures the likelihood of a given model state with respect to the observations.

When dealing with high-~~-~~dimensional and nonlinear systems, an analytical solution of the problem is unfeasible (Moradkhani et al., 2005). The implementation of ensemble methods ~~using a~~ (e.g. Monte Carlo sampling) allows to fully approximate the posterior density $p(X_t|D_t)$ through a set of *N* independent randomly drawn samples, called particles (Arulampalam et al., 2002; Moradkhani et al. 2005; Weerts and El Serafy, 2006):

$$p(X_{0:t}|Y_{1:t}) \approx \sum_{i=1}^{N} W_t^i \delta(X_{0:t} - X_{0:t}^i) \quad (6)$$

where $\{X_t^i, W_t^i\}$ denote the i-th particle drawn from the posterior distribution and its associated weight, $\delta(\cdot)$ is the Dirac delta function. 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, ~~P~~particles (X_t^i) are drawn from a known proposal distribution $q(X_{0:t}^i|Y_{1:t})$, according to the Sequential Importance Sampling (SIS) approach (Moradkhani et al., 2005; Guingla et al., 2012). The importance weights of the particles are recursively defined according to the following formula:

$$W_t^i \propto W_{t-1}^i p(Y_t|X_t^i) \tag{7}$$

A well-known common issue with SIS-PF is the sample degeneracy, which prevents particles from properly approximating the posterior distribution. Arulampalam et al., (2002) explained that whenever the effective number of particles (N_{eff}) falls below a fixed threshold value, the impact of the degeneracy needs to be mitigated by increasing the number of particles, where:

$$N_{eff} \approx \frac{1}{\sum_{i=1}^{N} (W_t^i)^2} \tag{8}$$

Since this approach is often unfeasible due to the increase in computational demand (Salamon and Feyen, 2009), a resampling procedure is frequently introduced to restore the sample variety through a Markov chain ~~chaotic~~ Monte Carlo (Moradkhani et al., 2005).

2.3.1 Sequential Importance Resampling

Gordon et al. (1993) proposed the Sequential Importance Resampling (SIR) technique, which introduces a resampling procedure within the SIS procedure. At each time step, the additional resampling step discards particles having low importance weights while replicating particles having high importance weight, while the total number of particles N is maintained unchanged (Figure 2a, b). As exhaustively explained by Weerts and Serafy (2006), the SIR algorithm relies on the generation of an empirical cumulative distribution (cdf) of the particles according to their weights W_t^i (Figure 2c) and the projection of a discrete set of N samples $\{X_t^i, i = 1, \dots, N\}$ with probabilities $\{W_t^i, i = 1, \dots, N\}$ uniformly drawn within the domain of the distribution. The resulting set contains replications of the particles having high importance weight, which are the most likely to be drawn (Figure 2d).

2.3.2 Likelihood function

When dealing with a multivariate SIR-PF scheme, it is necessary to take into account the different uncertaint~~y~~ies affecting each observed variable. Therefore, the likelihood function is a N_{obs}-dimensional normal distribution, where N_{obs} is the varying number of the effectively assimilated variables. The likelihood function is therefore defined as:

$$p(Y_t|X_t) = N\{(Y_t - X_t^i), \mu, R\} \tag{9}$$

where μ and R are respectively the null mean vector and the error covariance matrix of observations characterizing the multivariate Gaussian distribution. Thus, at each assimilation time step the particles weights are updated according to the following equation:

$$W_t^i = \frac{exp\left(-\frac{1}{2R}\left[Y_t - H\left(X_t^i\right)\right]^2\right)}{\sum_{i=1}^{N} exp\left(-\frac{1}{2R}\left[Y_t - H\left(X_t^i\right)\right]^2\right)} \tag{10}$$

2.4 Generation of ensemble particles

2.4.1 Perturbation of meteorological input data

Meteorological forcings are one of the major sources of uncertainty affecting snowpack simulations (Raleigh et al., 2015). Therefore, an ensemble of possible model realizations is generated by perturbing the model inputs, namely precipitation, air temperature and relative humidity, solar radiation, wind speed. The ensemble of perturbed inputs allows to take into account a well-representative range of weather conditions at the experimental sites, which result in an ensemble of possible snowpack states standing for the uncertainty of model predictions (Charrois et al., 2016). A meteorological ensemble of 100 members is generated by perturbing the in-situ meteorological data with an additive stochastic noise applied (in a log-scale for precipitation and wind speed) at each time step (i.e. 15 minutes). Following the methodology proposed by Magnusson et al. (2017), the random perturbations are provided through a first-order autoregressive model in order to guarantee a physical consistency and a temporal correlation of the time-variant forcings. Perturbations are generated considering the error statistics evaluated at the CDP site (**Table 1**) (Magnusson et al., 2017), which result from the comparison between SAFRAN reanalysis data (Vidal et al., 2010) and the observations supplied by the French station (Charrois et al., 2016). Even though this approach ensures to take account of the actual meteorological errors affecting the quality of the model predictions, ~~T~~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).

It is also noteworthy that the same ~~error~~ statistics of the meteorological analysis error specifically derived ~~from the observations supplied by~~ at the CDP station are used for the generation of the meteorological ensembles at all the snow experimental sites. As highlighted by Magnusson et al. (2017), this approach is likely to reduce the filter performance at the Italian and Swiss sites.

~~2.4.2 Model parameters resampling and perturbations generation~~

2.4.2 Perturbation of model parameters

Alongside the meteorological forcing, the parameterization of the physical processes occurring within the snowpack contributes to greatly increasing the uncertainty affecting model predictions (Essery et al., 2013; Lafaysse et al., 2017). The perturbation of key model parameters allows to take account of the uncertainty resulting from their empirical estimation.

Furthermore, ~~Since~~ since the introduction of stochastic noise plays a major role in reducing the effect of the sample impoverishment (Moradkhani et al., 2005), the uncertainty of model parameters is supposed to contribute to restore the ensembles spread between two following assimilation time steps. Following the methodology proposed by Moradkhani et al. (2005), the resampling procedure is carried out both in the parameters and the state variables spaces. Therefore, at each assimilation time step, after the particles resampling the parameters are perturbed through an additive noise before being used at the successive time step. Following Salamon and Feyen (2009), the parameters variance is restricted between upper and lower limits in order to avoid model instabilities and to also assure a minimum process noise, in order to prevent any variance collapse. The variance ranges are set according to the results of several tests carried out by varying their limits and evaluating the impact on filter performance. Unlike the study of Moradkhani et al. (2005), who applied the dual SIR-PF scheme to estimate model parameters, in this case the main aim is to succeed in enlarging the parameters ensemble through a consistent perturbation variance to ensure a significant spread of the particles.~~. Therefore, tuning parameters are properly set to guarantee a significant variance of the parameters distribution.~~

When considering the uncertainty of model parameters, a preliminary sensitivity analysis is of key importance for a twofold reason. Firstly, it is intended to properly identify the parameters mostly affecting the model simulations. Secondly, an accurate selection accordingly enables to neglect those parameters whose perturbation would demand for a larger computational requirement without resulting in a significant improvement of the ensemble spread.

Consistently with the study of Piazzi et al. (2018), snow roughness and snow viscosity are assumed to critically condition the model snowpack dynamics (Figure 3). Indeed, several analyses revealed a high sensitivity of model simulations to the perturbation of these mass-related parameters, especially in terms of SWE and snow density. Snow roughness strongly affects the snowpack energy balance by ruling the turbulent heat fluxes. As a consequence, the perturbation of this parameter mainly impacts the SWE ensembles by providing each particle with different snow melting fluxes. The effects of the perturbation of snow viscosity are prominent on the snow density evolution, especially on the snow compaction dynamics. As shown in Figure 3, it is noteworthy to observe the gradual increase of the ensemble spread of snow viscosity values throughout the melting period, suggesting that the perturbation of this parameter allows an offsetting effect of model melting issues.

Since the main criterion for selecting the model parameters focuses on identifying those whose perturbation allows to increase the ensemble spread of the state variables only slightly affected by the meteorological uncertainty, the three parameters describing the dynamics of the surface albedo are also considered (Table 2). The perturbation of the albedo parameters mainly guarantees a significant enlargement of the ensembles spread of this prognostic variable, with an impact on the snow mass balance especially during the melting period.

2.5 Experimental setup

2.5.1 Snow data

The multivariate DA scheme has been designed to consistently update the system state by jointly assimilating ground-based observations of surface temperature, albedo, snow depth, SWE and snow density. ~~Table 2~~Table 3 lists the datasets of the experimental sites.

Automatic in-situ measurements of surface temperature, albedo, and snow depth are supplied on an hourly or sub-hourly basis by all the selected stations throughout the whole datasets.

Even though direct SWE measurements are generally widely lacking, the snow experimental sites are one of the main sources of consistent measures of this variable. Daily automatic measures are provided at CDP since the winter season 2001/02. With a lower measurement frequency, at WFJ SWE observations are available every two weeks over the whole dataset period. Unlike these two sites, the TGN station supplies in-situ 6-hours~~rs~~ automatic SWE measurements during the snow seasons 2013/14 and 2015/16. In order to be able to properly use these measures within the analysis of simulations on seasonal/annual scale, the raw observational data have been smoothed from possible inconsistent oscillations and anomalies (e.g. rain-on-snow events) through their daily average (~~Table 3~~Table 4).

Although an exhaustive knowledge of snow density is needed to properly define the snowpack state and its dynamics, direct continuous observations of snow density are generally lacking. An exception is the TGN site, where biweekly manual measurements provide useful information throughout the whole dataset period. However, thanks to the relation among snow depth, SWE and snow density (Jonas et al., 2009), observations of at least two of these variables are enough to indirectly estimate the third one with a roughly tolerable degree of uncertainty. According to this approach, at both the Swiss and the French sites bi-weekly and daily snow density measurements have been derived, respectively (threshold value at 550 kg/m^3). At TGN daily snow densities have been indirectly estimated during the two winter seasons when SWE measures are available. Conversely, bi-weekly SWE estimates have been derived during the other two snow seasons, when snow densities measurements are otherwise supplied. Because the two sensors measuring snow depth and SWE can be not located at exactly the same point, however, it is noteworthy that possible inconsistencies can arise due to the spatial variability in snow cover, especially under shallow snow conditions (Essery et al., 2013; Lafaysse et al., 2017).

2.5.2 Multivariate DA experiments

Several experiments have been carried out with the aim of assessing the performances of the multivariate SIR-PF scheme under different configurations, as listed in Table 5.

In the first experiment [M_Exp] the efficiency of the DA scheme in updating the model snowpack states is tested by assuming the meteorological data as the only source of uncertainty. The snow observations are jointly assimilated every 3 hours to ensure an efficient exploitation of the high frequency in-situ measurements supplied by the automatic stations at the analysed snow experimental sites.

The second experiment [MP_Exp (1)] aims at assessing the impact of the perturbation of the model parameters on the filter performance. Indeed, the introduction of the parameters uncertainty is supposed to contribute to limiting the sample impoverishment by enlarging the ensembles spread, whose size strongly impacts the filter performance. Therefore, along with the perturbation of meteorological inputs, each particle independently evolves according to its own specific set of parameters ruling the model physical dynamics, whose equations remain unchanged, however. Thanks to this approach, the degeneracy of the model ensembles is limited via resampling and the sample impoverishment is prevented through the parameters perturbation. However, as stated by Salamon and Feyen (2009), when dealing with parameters uncertainty it is important to consider that the model response to a change in parameters does not have an immediate effect on the simulated state. This issue can be overcome by giving the model a sufficiently large response time between following system updates. The assimilation frequency is therefore reduced to every 24 hours. This choice is intended to guarantee a higher model response time without omitting a large number of observed snow data.

The sensitivity of this last configuration of the DA scheme to the measurement frequency is investigated by assessing how the filter performance is affected when limiting the observational dataset of the CDP site from daily to bi-weekly SWE measurements [MP_Exp (2)].

With the aim of reducing the system sensitivity to the availability of snow mass observations, the fourth experiment [MPP_Exp] tests the potential of using indirect information on SWE and snow density, whose measurements are generally time-consuming and often not available for real-time applications. According to the methodology proposed by Jonas et al. (2009), an additional empirical snow density model is introduced to reliably determine indirect sampling of SWE state from snow depth measurements through a parameterization of snow density, depending on four main factors: seasonality, observed snow depth, site altitude and location. With the aim of evaluating the reliability of the resulting estimates, a qualitative comparison analysis is performed with respect to the observations available at the Swiss and Italian measurement sites, as shown in Figure 4. Except for some sporadic winter seasons, generally the estimates of SWE and snow density well fit the observed snowpack dynamics, as demonstrated by a good agreement with the ground-based measurements. However, it is noteworthy that the estimate of snow density features is more challenging for shallow snow depths, since a high variability can range from low-density new snow in early winter to high-density slush during springtime (Jonas et al., 2009).

When dealing with a multivariate assimilation of several observed variables, it is of key importance to investigate whether the ensemble size can be sufficient to efficiently describe the high-dimensional assimilation scheme. With the aim of addressing this critical issue, after identifying the most proper DA configuration for each experimental site, according to the local features and the availability of observed data, the system sensitivity to the ensemble size is investigated by testing 100-, 200- and 500-particles ensemble simulations [nP_Exp]. This last experiment is performed by considering a sample of one winter season for each experimental site: snow seasons 2007/08 at CDP, 2001/02 at WFJ, and 2012/13 at TGN site. The impact of the variation in the particles number on the filter performance is analysed by evaluating both the ensembles spread and the ensemble effective size at the resampling step, namely the number of selected particles having significant likelihood values with respect to the total ensemble size.

2.5.3 Open loop ensemble simulations

With the aim of properly analysing the skill of the multivariate DA scheme, each experiment is evaluated through the comparison with control open loop (without DA) ensemble simulations forced by perturbed meteorological data (Ens_OL) (Sect. 2.4.1).

After verifying that the introduction of the stochastic noise does not affect the observed inputs on average at any site, firstly the aim is to assess the impact of the meteorological perturbation on the ensemble snowpack simulations, without considering the assimilation of snow data. Indeed, since the strong system nonlinearities make the model response to the inputs perturbation hardly predictable, it is important to verify that no unexpected biases occur with respect to the deterministic control run. Secondly, it is important to consider how the perturbation of the meteorological data succeeds in realistically depicting the uncertainty of snow model simulations.

To investigate the impact of the meteorological stochastic perturbations, 100-ensemble snowpack simulations forced by as many different meteorological conditions are analysed. For the sake of concision and clarity, a representative winter season is shown for each site (Figure 5). The ensembles spread reveals possible over- and underestimation of the ensemble model simulations as direct consequence of the perturbation of the forcing data. Since the meteorological perturbations are unbiased, this issue is mainly due to the nonlinearity of the involved physical processes. Therefore, it is noteworthy to observe that, even though the time series of the deterministic control open loop run are generally included within the ensemble envelop, they differ from the ensemble mean simulations. The variance of the mass-related ensembles is generally the largest at the end of the winter season, when the perturbation of energy-related forcing variables (i.e. air temperature, shortwave radiation) leads to well-spread melt-out scenarios resulting from the difference in melt timing (i.e. some particles have just started to melt and some others have already disappeared). During the winter season, the spread of SWE ensembles is increased whenever a snowfall event occurs due to the uncertainties in the precipitation rates allowing to provide the mass balance of each model realization with different input of snowfall rate. Of course, sites climatology (e.g. frequency of snowfall events) strongly impacts the resulting ensemble variance. A significant variance of the surface temperature ensembles is ensured by its high sensitivity to the inputs uncertainty, namely the perturbation of the air temperature and shortwave radiation, which directly impacts the snowpack energy balance. However, it is important to consider that some threshold processes involved within the snow dynamics model (e.g. disappearance of the surface snow layer, limitation of state variable within physical ranges) can be counter-productive in enlarging the ensembles spread. Figure 6 shows the Talagrand diagrams (Hamill, 2001) of the Ens_OL SWE simulations at CDP site. The deterministic open loop predictions are properly included within the ensemble envelop. However, the shape of the Talagrand distribution reveals an underestimated trend of the Ens_OL simulations, on average, with respect to the deterministic ones. This issue is more prominent when considering the SWE observations, as proven by the peak in the rank histogram.

2.5.~~2~~ 4 Evaluation metrics

~~T~~The results of this study are shown and discussed in terms of SWE, snow depth and surface temperature. ~~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 (Zappa et al., 2003). Along with this variable, the~~ The assessment of both SWE and snow depth simulations allows to indirectly evaluate also the model dynamics of snow density (Jonas et al., 2009). Furthermore, the impact of the filter updating on the system energy balance is analysed ~~by considering~~through the evaluation of the simulations of the surface temperature.

~~I~~In order to properly assess the filter performance, each ~~analysis~~ experiment is evaluated through a deterministic statistical index related to the ensemble mean simulations, and an ensemble-based probabilistic skill metrics.~~, by considering the whole datasets of measurements~~ These multi-year evaluation metrics are computed by considering the whole datasets of measurements (snowless periods are neglected)~~:~~.

~~The experiments are evaluated~~

~~Correlation coefficient:~~ $\cancel{R = \frac{cov(Obs,Exp)}{\sigma_{obs} \cdot \sigma_{exp}}}$ ~~(11)~~

~~Root Mean Square Error:~~ $\cancel{RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(Obs_k - Exp_k)^2}}$ ~~(12)~~

~~Normalized Error Reduction (Chen et al., 2011):~~ $\cancel{NER = \left(1 - \frac{RMSE_{Exp}}{RMSE_{OL}}\right) \cdot 100}$ ~~(13)~~

~~Assimilation efficiency (Brocca et al., 2012) :~~ $\cancel{Eff = \left(1 - \frac{\sum_{k=1}^{N}(Exp_k - Obs_k)^2}{\sum_{k=1}^{N}(OL_k - Obs_k)^2}\right) \cdot 100}$ ~~(14)~~

wh~~ere~~ ~~OL are the Open Loop simulations (without DA) of the control run.~~

The Kling-Gupta Efficiency (KGE) coefficient (Gupta et al., 2009) allows to analyse how the assimilation of snow observations succeeds in properly updating the model simulations, on average:

$$KGE = 1 - \sqrt{(r-1)^2 + (a-1)^2 + (b-1)^2} \qquad (11)$$

where:

- r is the linear correlation coefficient between the mean ensemble simulations and observed values;
- a is the ratio of the standard deviation of mean ensemble simulations to the standard deviation of the observed ones, i.e. an estimate of the relative variability between simulated and observed quantities;
- b is the ratio of the mean of mean ensemble simulations to the mean of observed ones, i.e. a measure of the overall bias.

The optimal KGE value is ideally equal to 1, revealing that the ensemble mean simulations succeed in well catching the observed values (theoretically, r=1, a=1, b=1).

The Continuous Ranked Probability Skill Score (CRPSS) is evaluated to assess changes to the overall accuracy of the ensemble simulations of each experiment (CRPS) by considering the open loop ensemble (Ens_OL) control run as the reference one ($CRPS_{ref}$), according to the formula:

$$CRPSS = 1 - \frac{CRPS}{CRPS_{ref}} \qquad (12)$$

The Continuous Ranked Probability Score (CRPS) measures the error in the cumulative probability distribution computed from the ensemble members relative to observations (Hersbach 2000):

$$CRPS = \int_{+\infty}^{+\infty}\left(P_{ens}(x) - P_{obs}(x)\right)^2 dx \qquad (13)$$

The smaller the CRPS value, the better the probabilistic simulation is (perfect score equal to 0). Conversely, the optimal value of CRPSS is equal to 1 and negative values indicate poorer performances with respect to the reference control run.

3. RESULTS AND DISCUSSION

~~**3.1.1 Impact of the meteorological uncertainty on ensemble simulations without DA**~~

~~After verifying that the introduction of the stochastic noise does not affect the observed inputs on average at any site, firstly the aim is to assess the impact of the meteorological uncertainty on the ensemble snowpack simulations, without considering the assimilation of snow data. Indeed, since the strong system nonlinearities make the model response to the inputs perturbation hardly predictable, it is important to verify that no spurious trends occur with respect to the deterministic control run. To investigate the impact of the meteorological stochastic perturbations, 100-ensemble snowpack simulations forced by as many different meteorological conditions are analyzed. For the sake of concision and clarity, a representative winter season is shown for each site (Figure 3).~~

~~The ensembles spread reveals possible over- and underestimation of the ensemble model simulations as direct consequence of the perturbation of the forcing data. Nevertheless, the time series of the deterministic control run (i.e. open loop simulations) are generally included within the ensemble envelop.~~

~~The variance of the mass-related ensembles is generally the largest at the end of the winter season, when the perturbation of energy-related forcing variables (i.e. air temperature, shortwave radiation) leads to well-spread melt-out scenarios resulting from the difference in melt timing (i.e. some particles have just started to melt and some others have already disappeared). During the winter season, the spread of SWE ensembles is increased whenever a snowfall event occurs due to the uncertainties in the precipitation rates allowing to provide the mass balance of each model realization with different input of snowfall rate. Of course, sites climatology (e.g. frequency of snowfall events) strongly impacts the resulting ensemble variance. A significant variance of the surface temperature ensembles is ensured by its high sensitivity to the inputs uncertainty, namely the perturbation of the air temperature and shortwave radiation, which directly impacts the snowpack energy balance. However, it is important to consider that some threshold processes involved within the snow dynamics model (e.g. disappearance of the surface snow layer, limitation of state variable within physical ranges) can be counter-productive in enlarging the ensembles spread.~~

~~**3.1.2 Multivariate DA simulations**~~

~~The SIR-PF scheme is implemented to jointly assimilate several observed snow data within the snow model (Sect. 2.4). The aim is to assess the performance of the multivariate DA system and investigate its efficiency in updating snowpack states.~~

~~The assimilation frequency is set to every 3 hours in order to ensure an efficient exploitation of the high frequency in-situ measurements supplied by the automatic stations at the analyzed snow experimental sites.~~

~~**3.1 Multivariate DA simulations with perturbed meteorological input data**~~

With respect to the control run (Ens_OL), the multi-year KGE values of the multivariate DA simulations relying on the perturbation of the meteorological data [M_Exp] reveal the filter effectiveness in updating snow depth simulations (Figure 7).

Conversely, the update of SWE model predictions is more challenging. At the French station, the assimilation of snow data actually leads to a worsening of the quality of the SWE simulations with respect to the probabilistic control run. A slight improvement is observed at the Swiss and Italian sites, where the filter updating benefits from a larger spread of the SWE simulations ensured by a higher frequency of snowfall events, on average. Even though the filter well succeeds in enhancing the simulations of surface temperature at CDP site, the snowpack thermal state at TGN and WFJ is poorly affected by the assimilation of snow data.

To better understand and properly assess the results, it is important to stress some key conditions exerting the most influence on the filter effectiveness. One of the main ruling issues is the scale of the model ensemble spread. A well-spread ensemble makes the filter efficient in weighting the particles, since they are properly discriminated through different likelihood values (Ades and Van Leeuwen, 2013). ~~When the ensemble is squeezed, the resampling procedure is more challenging since all the particles are close to each other with resulting similar likelihood values. In this undesired case, the filter might not succeed in well discriminating the more likely ensemble members since they all are assigned almost the same weight. On the other hand, if the particles are well spaced, their resampling is more straightforward since each particle is properly discriminated through a specific weight proportional to its likelihood.~~ Alongside this issue, it is of critical importance how the particles are placed on average with respect to ~~their measures~~the measure of the corresponding variable. The most conducive condition calls for well-spread ensembles enclosing their corresponding observations. ~~In this situation, the filter is favoured in selecting the most likely particles and properly weighting them.~~ However, the spread of the model ensembles turns out to be the overriding condition. Indeed, even if the model predictions are biased, a large ensemble spread can allow to progressively stretch the simulations towards the observed system state through subsequent proper updates.

When dealing with a multivariate DA scheme, the fulfilling of these conditions is even more challenging. In such an application, the filter is designed to select the particles best describing the observed system state with respect to all the available observations at the assimilation time step. Therefore, with respect to a univariate DA scheme, here the filtering procedure is more heavily constrained, depending on how many observations are provided.

Even though the effects of the ensemble degeneracy can be reduced through the resampling procedure, the perturbation of the meteorological data turns out to be not sufficient to prevent the sample impoverishment within two following assimilation time steps (e.g. SWE ensemble when no snowfall event occurs). It is noteworthy that a decrease in the ensemble spread, even just of one variable, can affect the overall resampling procedure. As previously explained, this limitation is even intensified by the physics of the snowpack model, whose threshold processes can weaken the effect of inputs perturbation.

Another further issue is the difference in the measurements frequency of the variables to assimilate. At each assimilation time step the Gaussian likelihood function is n-dimensional depending on the number of the observed variables. Thus, the particles weighting is carried out considering their likelihood in relation to the available measurements at that time. This dynamic entails that the resampling procedure can be more strongly conditioned by the observations having a higher measurement frequency (e.g. hourly or sub-hourly measurements of snow depth). Thus, possible misleading updates of the variables less frequently observed can occur, since they are updated without taking into account particles likelihood with respect to their own lacking observations (e.g. daily or bi-weekly measurements of SWE). For instance, when a measure of snow depth is provided, no observational information on SWE can be properly retrieved ~~unless~~ except its indirect estimate if snow density data are available. Otherwise, the filter can fail in consistently updating the overall snow mass-related state, since a lot of possible combinations of SWE and snow density can well fit the observed snowpack depth. In terms of filter efficiency, this means that when only a snow depth observation is provided the filter looks for particles having the higher likelihood with respect to this snow quantity, regardless the SWE and snow density states. Nevertheless, it is not unlikely that several particles have the same likelihood because the combination of even strongly-biased values of SWE and snow density can well fit the observed snow depth through an offsetting effect among these variables. This is the main reason explaining the higher filter performance in terms of snow depth with respect to the SWE simulations.

The multivariate DA simulations allow to point out two main limitations of this application. Firstly, the ensemble spread needs to be enlarged in order to improve the filter efficiency in well weighting and resampling the ensemble particles. Moreover, the effect of the difference in measurements frequency of the assimilated variables has to be limited in order to consistently update the snow mass balance.

~~3.2 Multivariate DA with perturbed model parameters~~

3.2 Multivariate DA simulations with perturbed model parameters

~~With the aim of overcoming the sample impoverishment, the uncertainty of model parameters is introduced in order to succeed in enlarging the ensembles spread, whose size strongly impacts the filter performance. Therefore, along with the perturbation of meteorological inputs, each particle independently evolves according to its own specific set of parameters ruling the model physical dynamics, whose equations remain unchanged, however.~~

~~3.2.1 Sensitivity analysis of model parameters~~

~~When considering the uncertainty of model parameters, a preliminary sensitivity analysis is of critical importance for a twofold reason. Firstly, it is intended to properly identify the most ruling parameters affecting the model simulations. Secondly, an accurate selection accordingly enables to neglect those parameters whose perturbation would uselessly demand for a larger computational requirement. The sensitivity analysis is carried out by making the parameters vary within proper ranges and analyzing the impact on the model predictions. Parameters ranges are estimated in order to both avoid model numerical instabilities and comply with possible physical constraints. The criterion for selecting the model parameters is mainly focused on identifying those allowing to increase the ensemble spread of the state variables more slightly affected by the meteorological uncertainty, namely SWE, snow density and surface albedo. The sensitivity analysis allows to identify 5 model parameters listed in Table 4.~~

~~While the three albedo parameters are perturbed in order to guarantee a significant enlargement of the ensembles spread of this prognostic variable, the perturbation of snow roughness and snow viscosity strongly impacts the snowpack dynamics. Indeed, several analyses reveal a high sensitivity of model simulations to the resampling of these mass-related parameters, especially in terms of SWE and snow density. Snow roughness strongly affects the snowpack energy balance by ruling the turbulent heat fluxes. As a consequence, the perturbation of this parameter mainly impacts the SWE ensembles by providing each particle with different snow melting fluxes. The effects of the perturbation of snow viscosity are prominent on the snow density evolution, especially on the snow compaction dynamics.~~

~~Thanks to this approach, the degeneracy of the model ensembles is avoided via resampling and the sample impoverishment is prevented through the parameters perturbation (Figure 5).~~

~~3.2.2 Multivariate DA simulations with parameters resampling~~

~~As stated by Salamon and Feyen (2009), when dealing with parameters uncertainty it is important to consider that the model response to a change in parameters has not immediate effect on the simulated state. This issue can be overcome by giving the model a sufficiently large response time between following system updates. The assimilation frequency is therefore reduced to every 24 hours. This choice is intended to guarantee a higher model response time without omitting a large number of observed snow data.~~

Figure 7 shows the ~~statistical metrics~~multi-year KGE values of the multivariate DA simulations resulting from the implementation of the model parameters ~~resampling~~perturbation [MP_Exp]. With respect to the previous ~~analysis~~ experiment considering the meteorological data as the only source of uncertainty (Sect. 3.1~~.2~~), the introduction of the parameters perturbation~~resampling~~ allows to heavily ~~increase~~ improve the filter efficiency at updating the model SWE simulations at CDP site. Table 6 shows this significant enhancement in terms of CRPSS, whose negative value for the M_Exp SWE simulations at the French station increases up to 0.74. It is noteworthy that the parameters ~~resampling~~ perturbation does not only ensure a sizeable enlargement of the ensembles spread but it also allows to better estimate the model parameters on average. Indeed, while the resampling of the state variables allows to consistently update the system state at the assimilation time step, the modelling of snowpack dynamics between two following assimilation time steps benefits from the parameters resampling, which enables to take better account of the parameters seasonality (e.g. melting period). Figure 8 provides a case in point, showing how the parameters perturbation impact on the filter effectiveness in terms of both the ensemble spread and its positioning with respect to the observation, at the same assimilation time step. At the French site, the daily SWE measurement frequency ensures an effective resampling of the mass-related parameters, namely snow roughness and snow viscosity. Furthermore, the retention of satisfying performance of the filter in terms of snow depth on average suggests a beneficial impact on the snow density dynamics. Conversely, the SWE simulations at the Swiss station do not benefit from the introduction of the parameters ~~resampling~~perturbation (Table 6) ~~does not always result in a significant improvement of SWE simulations at the Swiss station~~. This limitation is mainly due to the lower biweekly frequency of the SWE measurements, with respect to the French case study. At WFJ site, at the daily assimilation time steps when no SWE observation is available, the estimate of the particles likelihood cannot rely on observational information on the snow mass-related parameters (e.g. SWE, snow density). Therefore, it is not unlikely that the resampling procedure leads to suboptimal values of the mass-related parameters. Moreover, the enlargement of the ensembles spread ensured by the parameters ~~resampling~~ perturbation entails a higher probability of selecting particles having SWE values even farther from the actual state with respect to the simulations of the deterministic control run, when no direct SWE observed data are provided. This thesis is supported by the annual results obtained at the TGN site (not shown here), where the multivariate DA scheme allows to consistently update the SWE simulations when the daily average SWE measurements are available, namely throughout the winters 2013-14 and 2015-16. During the other two snow seasons, when biweekly SWE observations are assimilated, the filter does not succeed in improving model predictions.

The filter updating is not as effective for the simulation of surface temperature, especially at the Swiss and ~~French~~ Italian sites (Table 6). This suboptimal performance is mainly addicted to the concurrence of several factors. Firstly, the quicker dynamics of the daily thermal cycle make the temperature simulations more sensitive to the reduction in assimilation frequency, with respect to the other variables. Secondly, even though the filter succeeds in daily updating the system thermal state, the parameters values resulting from the resampling procedure do not ensure a long-lasting effect on the temperature trend between two following assimilation time steps. Indeed, since the parameters are resampled according to their representativeness at the assimilation time step, their values are not likely to be the optimal ones to well catch the succession of diurnal and nocturnal peaks.

Although the parameters ~~resampling~~ perturbation ensures an enlargement of the ensembles spread, which is one of the constraining conditions to ensure the filter effectiveness, the quality of the multivariate DA simulations strongly depends on the reliability of the parameters resampling, which requires direct observational information to properly estimate the more likely parameters values.

3.3 Sensitivity analysis of the multivariate DA scheme to the SWE measurement frequency

~~3.2.3 Sensitivity of parameters resampling to the SWE measurement frequency~~

The Italian case ~~study provides evidence of the impact of the difference in SWE measurement frequency on the parameters resampling. In order to further~~ With the aim of investigat~~e~~inge the system sensitivity to the SWE measurement frequency, an experiment is performed at the CDP station ~~with the aim of assessing~~ to assess how the reduction from daily to biweekly SWE observed data affects the 24-hours multivariate DA simulations [MP_Exp (2)]. Obviously, a reduction in measurement frequency is expected to reduce the impact of the filter updating on the model simulations. However, when dealing with a multivariate DA scheme, the imbalance among the measurement frequency of the assimilated variables can lead to a further side-effect hindering the parameters estimate due to the largest impact of the more frequently observed variables (e.g. snow depth, surface temperature) on the particles weighting. ~~Figure 7~~Figure 9 shows the ensembles of snow viscosity and snow roughness resulting from the assimilation of daily and biweekly SWE observations throughout the winter season 2001-2002. A divergence between the two ensemble time series is clearly detectable on average, especially in terms of snow viscosity. The suboptimal estimate of the mass-related parameters based on biweekly SWE measurements leads to a worsening of model predictions with respect to the control run, as shown in ~~Figure 8~~Figure 10. Conversely, the filter effectiveness is not affected in terms of snow depth thanks to offsetting effects between SWE and snow density simulations.

~~3.3 Proxy information of snow mass-related variables~~

~~3.3.1 Additional snow density model~~

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

Even though the introduction of the parameters uncertainty well succeeds in enlarging the ensembles spread, the resampling procedure of both states and parameters turns out to be even counter-productive when it is not properly conditioned by observed data of ruling snow mass-related variables. ~~Nevertheless, since SWE and snow density measurements are time-consuming and thus often not available for real-time applications, the aim is to reduce the system sensitivity affecting the filter performance by deriving indirect information on these snow variables. According to the methodology proposed by Jonas et al. (2009), an empirical snow density model is introduced in order to reliably determine indirect sampling of SWE state from snow depth measurements through a parameterization of snow density, depending on four main factors: seasonality, observed snow depth, site altitude and location. With the aim of evaluating the reliability of the resulting estimates, a qualitative comparison analysis is performed with respect to the observations available at the Swiss and Italian measurement sites, as shown in Figure 9.~~

~~Except for some sporadic winter seasons, generally the estimates of SWE and snow density well fit the observed snowpack dynamics, as demonstrated by a good agreement with the ground-based measurements.~~

~~However, it is noteworthy that the estimate of snow density features is more challenging for shallow snow depths, since a high variability can range from low-density new snow in early winter to high-density slush during springtime (Jonas et al., 2009).~~

~~3.3.2 Multivariate DA simulations with proxy information of snow mass-related variables~~

The implementation of the additional snow density model providing proxy information on the mass-related snow variables at the Swiss and Italian sites allows to optimize the parameters resampling, as revealed by the outperforming statistical scores of the SWE simulations, especially at WFJ station (~~Figure 10~~Figure 7, Table 6).

~~Generally, this approach allows for a consistent improvement of the snow depth predictions, as evidence of a proper estimate of snow density, except for the winter season 1999-2000 at the WFJ site.~~

Conversely, no prominent effects are observed in terms of surface temperature and snow depth.

The reduction in assimilation frequency necessarily leads to omitting large quantities of observed data. With the aim of preventing this limitation, the approach proposed by Salamon and Feyen (2009) has been tested. According to this method, each particle is assigned the median of the weights evaluated at all observation time steps within the 24-hrs response time interval. Although this approach allows to make full use of the available measurements, a more intensive use of proxy information on the snow mass-related variables makes the filter effectiveness more affected by the quality of the estimates, with resulting heterogeneous filter performance over the analysed datasets.

3.5 Sensitivity analysis of the multivariate DA scheme to the ensemble size

Figure 11 shows the main results of the experiment aiming at assessing the system sensitivity to the ensemble size [nP_Exp]. When evaluating the effective sample size after each filter update, expressed as percentage of selected particles with respect to the total number of the ensemble size (i.e. 100, 200, and 500 particles), it is noteworthy that an increase in the particles number generally does not result in a significant increment of the percent effective sample size. On average, this quantity ranges around 50% for the CDP and WFJ station, and up to 60% at the Italian experimental site. Furthermore, when assessing the impact of the variation in the particles number on the filter performance in updating the model simulations (Table 7), the resulting multi-year KGE values do not reveal a systematic improvement in the simulations reliability as the ensemble size increases.

A low system sensitivity to the ensemble size is also clearly proven by the slight impact of the change in the particles number on the ensembles spread (Figure 11). Indeed, the increase of the ensemble size generally does not ensure a proportional enlargement of the particles spread, expect for the snow depth, whose 500-particles ensemble simulations reveal a slightly larger spread at CDP and TGN site.

Despite being a multivariate application of the PF-based scheme, the results of this experiment mainly demonstrate that a 100-particles ensemble can be assumed as sufficiently representative for a point-scale application.

4. CONCLUSIONS

This study investigated the potentials of a SIR-PF scheme for a multivariate assimilation of snow data at three experimental sites in the Alps. Even though PF technique proved its capability of properly handling the strong system nonlinearities of snow modelling, several challenging issues need to be addressed.

When dealing with a multivariate DA application, the sample impoverishment is more likely to occur with respect to the univariate case, since the filter is designed to strictly select the particles having the highest likelihood with respect to all the observed state variables. The perturbation of the meteorological forcing data has turned out not to be sufficient to restore the ensembles spread within two following 3-hrs assimilation time steps, with resulting poor filter performance, especially in terms of SWE. In order to prevent this undesired condition, further stochastic noise has been introduced through the parameters ~~resampling~~perturbation, with a reduction of the assimilation frequency to every 24 hours to ensure a sufficient model response time. At CDP site, where SWE observations are available with a comparably high (daily) frequency with respect to the other assimilated variables, this filter setup outperforms the PF-based simulations considering the meteorological data as the only source of uncertainty. Conversely, at the Swiss and Italian stations, the benefit of introducing additional stochastic noise through the parameters perturbation is overcome due to the lower (biweekly) measurement frequency of the snow mass-related quantities. ~~Although this approach~~ Indeed, even though the parameters perturbation ~~has~~ succeed~~sed to enlarge~~ in enlarging the model ensembles, the system has revealed a prominent sensitivity to the difference in measurement frequency of the assimilated variables, which hinders the filter effectiveness in consistently updating the modelled snow quantities. ~~Indeed~~Actually, the more frequently measured snow variables (e.g. snow depths) strongly condition the mass-related parameters resampling (i.e. snow viscosity and snow roughness), which can result in possible suboptimal values.

Where snow mass-related observations are less frequently available, the assimilation of indirect estimates has revealed a remarkable potential to make up for the lack of information within the resampling procedure. At WFJ and TGN sites, the introduction of an additional model providing proxy data of snow density and SWE has allowed to improve the consistency of the filter updating.

The filter has turned out to be less effective in updating the simulation of surface temperature, mainly due to the quick dynamics of the daily thermal cycle which entails a short-lasting representativeness of the parameters values to properly describe the diurnal and nocturnal peaks.

The use of proxy information on snow mass-related variables supplied by an additional snow density model has allowed ~~to improve the parameters resampling, which results in outperforming filter updating of snow depth and SWE predictions.~~

In this point-scale application of the multivariate SIR-PF scheme, the system has revealed a low sensitivity to the ensemble size, thereby proving that 100 particles can be suited to represent the high dimensionality of the system. However, when modelling at larger scale, the sensitivity to the ensemble size needs to be deeply investigated, especially for multilayer snowpack models.

Another critical issue of key importance for spatialized applications is the combined perturbation of meteorological data and model parameters, since the high dimensionality of the modelling systems necessarily requires a consistently large stochastic noise to ensure the effectiveness of the filter updating. Especially during periods when no snowfall event occurs, the meteorological uncertainty is assumed not to be sufficient to ensure a well-representative range of possible snowpack states.

Furthermore, in multivariate PF-based applications for snow modelling, it is noteworthy to investigate the potential of introducing indirectly estimated information on those state variables not directly measured. Indeed, the lack of measures in ungauged or poorly gauged areas can hinder the application of a PF-based scheme, whose resampling procedures highly benefit from comprehensive information on the snowpack state to properly detect those particles having a higher overall likelihood value. This issue is even more compelling when dealing with a detailed physically-based snowpack model.

Nevertheless, several issues require further detailed analysis. The physical consistency of the meteorological ensembles needs to be improved. Indeed, the methodology does not take into account the correlations among the perturbed forcing variables and, moreover, the specific error statistics characterizing the perturbations are to be specifically evaluated at each analysed site. Even though the evaluation of the likelihood function for high-dimensional systems becomes more challenging (Margulis et al., 2015), the potential of using empirical likelihood variants should be extensively investigated to assess the impact of using a Gaussian distribution (Leisenring and Moradkhani, 2011; Thirel et al., 2013). Furthermore, an interest is focused on testing other resampling techniques with the aim of analysing how the resampling procedure affects the filter effectiveness (Moradkhani et al., 2005; Weerts and El Serafy, 2006; Salamon and Feyen, 2009).

Data availability

The snow and meteorological data from Col de Porte site are made freely available by Météo-France both on the PANGAEA depository (doi:10.1594/PANGAEA.774249) and on the public ftp server ftp://ftp-cnrm.meteo.fr/pubcencdp/. The Weissfluhjoch dataset provided by the WSL Institute for Snow and Avalanche Research SLF can be obtained via IDAWEB (https://gate.meteoswiss.ch/idaweb) as well as from the Environmental Data Portal ENVIDAT (envidat.ch/dataset/10-16904-1). The dataset of Torgnon site is available on the European Fluxes Database (www.europe-fluxdata.eu/).

Author contributions.

This work is part of the G. Piazzi's PhD thesis, supervised by S. Gabellani, L. Campo. G. Thirel supervised the research activities during the visiting period of G. Piazzi at the Catchment Hydrology Research Group of the ~~HBAN~~ HYCAR Research Unit (IRSTEA). All the authors have collaborated to the technical development of the modelling system and the manuscript writing.

Acknowledgements.

The research has been conducted in the framework of the collaboration between the ~~Hydrosystems and Bioprocesses~~ HYCAR Research Unit ~~(HBAN)~~ of the National Research Institute of Science and Technology for Environment and agriculture (IRSTEA) and CIMA Research Foundation. We wish to thank both institutions for supporting and promoting these research activities. The work has been also supported by Regione Autonoma Valle d'Aosta and the Next-Data Project. The authors would like to acknowledge Météo-France, the WSL Insitute, and the Environmental Protection Agency of Aosta Valley for providing the datasets used in this study. We are grateful to all the staff working at Col de Porte, Weissfluhjoch and Torgnon experimental sites, who ensure high-quality snow observations for many years. We would like to address special thanks to all the colleagues for the fruitful discussions on this matter. We are also grateful to Matthieu Lafaysse and the anonymous reviewer for helping to significantly improve this manuscript through their constructive remarks.

References

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

Anderson, E. A.: A point of energy and mass balance model of snow cover, Tech. Rep., Office of Hydrology – National Weather Service, 1976.

Andreadis, K. M., and Lettenmaier D. P.: Assimilating remotely sensed snow observations into a macroscale hydrology model. Advances in Water Resources 29.6, 872-886, 2005.

Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T.: A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Transactions on signal processing, 50(2), 174-188, 2002.

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Muñoz-Sabater, J., Pappenberger, F., de Rosnay, P., Stockdale, T. and Vitart, F.: ERA-Interim/Land: a global land surface reanalysis data set. Hydrology and Earth System Sciences, 19(1), 389-407, 2015.

Bergeron, J. M., Trudel, M., and Leconte, R.: Combined assimilation of streamflow and snow water equivalent for mid-term ensemble streamflow forecasts in snow-dominated regions. Hydrology and Earth System Sciences, 20(10), 4375, 2016.

Boni, G., Castelli, F., Gabellani, S., Machiavello, G., Rudari, R.: Assimilation of MODIS snow cover and real time snow depth point data in a snow dynamic model, Geoscience and Remote Sensing Symposium (IGARSS), IEEE International (pp. 1788-1791), 2010.

Bormann, K. J., Westra, S., Evans, J. P., and McCabe, M. F.: Spatial and temporal variability in seasonal snow density. Journal of hydrology, 484, 63-73, 2013.

Brasnett, B.: A global analysis of snow depth for numerical weather prediction. Journal of Applied Meteorology 38.6, 726-740, 1999.

Brocca, L., Moramarco, T., Melone, F., Wagner, W., Hasenauer, S., Hahn, S.: Assimilation of surface-and root-zone ASCAT soil moisture products into rainfall–runoff modelling. IEEE Transactions on Geoscience and Remote Sensing, 50(7), 2542-2555, 2012.

Caparrini F., Castelli F., and Entekhabi D.: Estimation of surface turbulent fluxes through assimilation of radiometric surface temperature sequences. Journal of Hydrometeorology 5.1, 145-159, 2004.

Charrois, L., Cosme, E., Dumont, M., Lafaysse, M., Morin, S., Libois, Q., & Picard, G.: On the assimilation of optical reflectances and snow depth observations into a detailed snowpack model. The Cryosphere, 10, 1021-1038, 2016.

Chen, Z.: Bayesian filtering: From Kalman filters to particle filters, and beyond. Statistics, 182(1), 1-69, 2003.

Chen, F., Crow, W. T., Starks, P. J., Moriasi, D. N.: Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture. Advances in Water Resources, 34(4), 526-536, 2011.

Clark, M. P., Slater, A. G., Barrett, A. P., Hay, L. E., McCabe, G. J., Rajagopalan, B., Leavesley, G. H.: Assimilation of snow covered area information into hydrologic and land-surface models. Advances in water resources, 29(8), 1209-1221, 2006.

Cortés, G., Girotto, M., and Margulis, S.: Snow process estimation over the extratropical Andes using a data assimilation framework integrating MERRA data and Landsat imagery, Water Resour. Res., 52, 2582–2600, doi:10.1002/2015WR018376, 2016.

Cressman, G. P.: An operational objective analysis system. Mon. Wea. Rev 87.10, 367-374, 1959.

De Lannoy, G. J. M., Reichle, R. H., Arsenault, K. R., Houser, P. R., Kumar, S., Verhoest, N. E. C., Pauwels, V. R. N.: Multiscale assimilation of Advanced Microwave Scanning Radiometer-EOS snow water equivalent and moderate resolution imaging spectroradiometer snow cover fraction observations in northern Colorado. Water Resour. Res., 48, 2012.

Dechant, C., and Moradkhani, H.: Radiance data assimilation for operational snow and streamflow forecasting. Advances in Water Resources, 34(3), 351-364, 2011.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P. Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L. Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hòlm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J. , Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Quarterly Journal of the royal meteorological society, 137(656), 553-597, 2011.

Dettinger, M.: Climate change: Impacts in the third dimension. Nature Geoscience, 7(3), 166-167, 2014.

Douville, H., Royer, J. F., and Mahfouf, J. F. : A new snow parameterization for the Meteo-France climate model. Climate Dynamics, 12(1), 21-35, 1995.

Dong, J., Walker, J. P., Houser, P. R., Sun, C.: Scanning multichannel microwave radiometer snow water equivalent assimilation. Journal of Geophysical Research: Atmospheres, 112, 2007.

Drusch, M., Vasiljevic, D., and Viterbo, P.: ECMWF's global snow analysis: assessment and revision based on satellite observations, J. Appl. Meteorol., 43, 1282–1294, 2004.

Durand, M., Margulis, S. A.: Feasibility test of multifrequency radiometric data assimilation to estimate snow water equivalent. Journal of Hydrometeorology, 7(3), 443-457, 2006.

Durand, M., Margulis, S. A.: Effects of uncertainty magnitude and accuracy on assimilation of multiscale measurements for snowpack characterization. J. Geophys. Res. Atmos. 113, D02105, 2008.

Dziubanski, D. J., and Franz, K. J.: Assimilation of AMSR-E snow water equivalent data in a spatially-lumped snow model. Journal of Hydrology, 540, 26-39, 2016.

Essery, R., Morin, S., Lejeune, Y., and Ménard, C. B.: A comparison of 1701 snow models using observations from an alpine site. Advances in Water Resources, 55, 131-148, 2013.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research: Oceans, 99(C5), 10143-10162, 1994.

Evensen, G.: The ensemble Kalman filter: Theoretical formulation and practical implementation. Oceandynamics, 53(4), 343-367, 2003.

Fayad, A., Gascoin, S., Faour, G., López-Moreno, J. I., Drapeau, L., Le Page, M., and Escadafal, R.: Snow hydrology in Mediterranean mountain regions: a review. Journal of Hydrology, 2017.

Filippa, G., Cremonese, E., Galvagno, M., Migliavacca, M., Di Cella, U. M., Petey, M., Siniscalco, C.: Five years of phenological monitoring in a mountain grassland: inter-annual patterns and evaluation of the sampling protocol. International journal of biometeorology, 59(12), 1927-1937, 2015.

Froidurot, S., Zin, I., Hingray, B., Gautheron, A.: Sensitivity of precipitation phase over the Swiss Alps to different meteorological variables. Journal of Hydrometeorology, 15(2), 685-696, 2014.

Galvagno, M., Wohlfahrt, G., Cremonese, E., Rossini, M., Colombo, R., Filippa, G., Julitta, T., Manca, G., Siniscalco, C., Migliavacca, M., Morra di Cella, U.: Phenology and carbon dioxide source/sink strength of a subalpine grassland in response to an exceptionally short snow season. Environmental Research Letters, 8(2), 025008, 2013.

Gascoin, S., Lhermitte, S., Kinnard, C., Bortels, K., and Liston, G. E.: Wind effects on snow cover in Pascua-Lama, Dry Andes of Chile. Advances in Water Resources, 55, 25-39, 2013.

Gelb, A.: Optimal linear filtering, in Applied Optimal Estimation, edited by A. Gelb, pp. 102– 155, MIT Press, Cambridge, Mass, 1974.

Guingla, P., Antonio, D., De Keyser, R., De Lannoy, G., Giustarini, L., Matgen, P., & Pauwels, V.: The importance of parameter resampling for soil moisture data assimilation into hydrologic models using the particle filter. Hydrology and Earth System Sciences, 16(2), 375-390, 2012.

Gordon, N. J., Salmond, D. J., and Smith, A. F.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In IEE Proceedings F (Radar and Signal Processing) (Vol. 140, No. 2, pp. 107-113).IET Digital Library, 1993.

Griessinger, N., Seibert, J., Magnusson, J., and Jonas, T.: Assessing the benefit of snow data assimilation for runoff modelling in Alpine catchments. Hydrology and Earth System Sciences, 20(9), 3895, 2016.

Gupta, H. V., H. Kling, K. K. Yilmaz and G. F. Martinez: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling." Journal of Hydrology 377(1-2): 80-91, 2009.

Hamill, T. M.: Interpretation of rank histograms for verifying ensemble forecasts. Monthly Weather Review, 129(3), 550-560, 2001.

Hedstrom, N. R., Pomeroy, J. W.: Accumulation of intercepted snow in the boreal forest: measurements and modelling. Hydrological Processes, 12, 1611-1625, 1998.

Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather and Forecasting, 15(5), 559-570, 2000.

Huang, C., Newman, A. J., Clark, M. P., Wood, A. W., &Zheng, X.: Evaluation of snow data assimilation using the ensemble Kalman filter for seasonal streamflow prediction in the western United States. Hydrology and Earth System Sciences, 21(1), 635, 2017.

Jonas, T., Marty, C., and Magnusson, J.: Estimating the snow water equivalent from snow depth measurements in the Swiss Alps. Journal of Hydrology, 378(1), 161-167, 2009.

Kalman, R. E.: A new approach to linear filtering and prediction problems. Journal of basic Engineering, 82(1), 35-45, 1960.

Lafaysse, M., Cluzet, B., Dumont, M., Lejeune, Y., Vionnet, V., and Morin, S.: A multiphysical ensemble system of numerical snow modelling. The Cryosphere, 11(3), 1173, 2017.

Leisenring, M., and Moradkhani, H.: Snow water equivalent prediction using Bayesian data assimilation methods. Stochastic Environmental Research and Risk Assessment, 25(2), 253-270, 2011.

Liston, G. E., Pielke, R. A., Greene, E. M.: Improving first-order snow-related deficiencies in a regional climate model. Journal of Geophysical Research: Atmospheres, 104(D16), 19559-19567, 1999.

Liston, G. E., Hiemstra, C. A.: A simple data assimilation system for complex snow distributions (SnowAssim).Journal of Hydrometeorology, 9(5), 989-1004, 2008.

López-Moreno, J. I., and Nogués-Bravo, D.: A generalized additive model for the spatial distribution of snowpack in the Spanish Pyrenees. Hydrological Processes, 19(16), 3167-3176, 2005.

Luce, C. H., Lopez-Burgos, V., and Holden, Z.: Sensitivity of snowpack storage to precipitation and temperature using spatial and temporal analog models. Water Resources Research, 50(12), 9447-9462, 2014.

Magnusson, J., Gustafsson, D., Hüsler, F., and Jonas, T.: Assimilation of point SWE data into a distributed snow cover model comparing two contrasting methods. Water resources research, 50(10), 7816-7835, 2014.

Magnusson, J., Winstral, A., Stordal, A. S., Essery, R., and Jonas, T.: Improving physically based snow simulations by assimilating snow depths using the particle filter. Water Resources Research, 53(2), 1125-1143, 2017.

Malik, M. J., Van Der Velde, R., Vekerdy, Z., Su, Z.: Assimilation of satellite-observed snow albedo in a land surface model.Journal of hydrometeorology, 13(3), 1119-1130, 2012.

Margulis, S. A., Girotto, M., Cortés, G., and Durand, M.: A particle batch smoother approach to snow water equivalent estimation. Journal of Hydrometeorology, 16(4), 1752-1772, 2015.

Miller, R. N., Ghil, M., Gauthiez, F.: Advanced data assimilation in strongly nonlinear dynamical systems.Journal of the atmospheric sciences, 51(8), 1037-1056, 1994.

Molotch, N. P., and Meromy, L.: Physiographic and climatic controls on snow cover persistence in the Sierra Nevada Mountains. Hydrological processes, 28(16), 4573-4586, 2014.

Moradkhani, H.: Hydrologic remote sensing and land surface data assimilation. Sensors, 8(5), 2986-3004, 2008.

Moradkhani, H., Hsu, K. L., Gupta, H., and Sorooshian, S.: Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. Water resources research, 41(5), 2005.

Morin, S., Lejeune, Y., Lesaffre, B., Panel, J. M., Poncet, D., David, P., and Sudul, M.: An 18-yr long (1993-2011) snow and meteorological dataset from a mid-altitude mountain site (Col de Porte, France, 1325 m alt.) for driving and evaluating snowpack models. Earth System Science Data, 4(1), 13, 2012.

Piazzi, G., Campo, L., Gabellani, S., Castelli, F., Cremonese, E., Morra di Cella, U., Stevenin, H., Ratto, S. M.: An EnKF-based scheme for snow multivariate data assimilation at an Alpine site. J. Hydrol. Hydromech, 2018~~7~~ (Accepted).

Quéno, L., Vionnet, V., Dombrowski-Etchevers, I., Lafaysse, M., Dumont, M., and Karbou, F.: Snowpack modelling in the Pyrenees driven by kilometric-resolution meteorological forecasts. The Cryosphere, 10(4), 1571-1589, 2016.

Raleigh, M. S., Lundquist, J. D., and Clark, M. P.: Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework. Hydrology and Earth System Sciences, 19(7), 3153-3179, 2015.

Revuelto, J., López-Moreno, J. I., Azorin-Molina, C., and Vicente-Serrano, S. M.: Topographic control of snowpack distribution in a small catchment in the central Spanish Pyrenees: intra-and inter-annual persistence. The Cryosphere, 8(5), 1989-2006, 2014.

Rodell, M., Houser, P. R.: Updating a land surface model with MODIS-derived snow cover. Journal of Hydrometeorology, 5(6), 1064-1075, 2004.

Salamon, P., and Feyen, L.: Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter. Journal of Hydrology, 376(3), 428-442, 2009.

~~Schmid, L., Heilig, A., Mitterer, C., Schweizer, J., Maurer, H., Okorn, R., & Eisen, O.: Continuous snowpack monitoring using upward-looking ground-penetrating radar technology. Journal of Glaciology, 60(221), 509-525, 2014.~~

Schmucki, E., Marty, C., Fierz, C., and Lehning, M.: Evaluation of modelled snow depth and snow water equivalent at three contrasting sites in Switzerland using SNOWPACK simulations driven by different meteorological data input. Cold Regions Science and Technology, 99, 27-37, 2014.

Slater, A. G., Clark, M. P.: Snow data assimilation via an ensemble Kalman filter. Journal of Hydrometeorology, 7(3), 478-493, 2006.

Stauffer, D. R., Seaman, N. L.: Use of four-dimensional data assimilation in a limited-area mesoscale model. Part I: Experiments with synoptic-scale data. Monthly Weather Review, 118(6), 1250-1277, 1990.

Su, H., Yang, Z. L., Niu, G. Y., Dickinson, R. E.: Enhancing the estimation of continental-scale snow water equivalent by assimilating MODIS snow cover with the ensemble Kalman filter. Journal of Geophysical Research: Atmospheres, 113, 2008.

Su, H., Yang, Z. L., Dickinson, R. E., Wilson, C. R., Niu, G. Y.: Multisensor snow data assimilation at the continental scale: The value of Gravity Recovery and Climate Experiment terrestrial water storage information. Journal of Geophysical Research: Atmospheres, 115, 2010.

Sun, C., Walker, J. P., Houser, P. R.: A methodology for snow data assimilation in a land surface model. Journal of Geophysical Research: Atmospheres, 109, 2004.

Thirel, G., Salamon, P., Burek, P., and Kalas, M.: Assimilation of MODIS snow cover area data in a distributed hydrological model using the particle filter. Remote Sensing, 5(11), 5825-5850, 2013.

Vidal, J. P., Martin, E., Franchistéguy, L., Baillon, M., and Soubeyroux, J. M.: A 50-year high-resolution atmospheric reanalysis over France with the Safran system. International Journal of Climatology, 30(11), 1627-1644, 2010.

Viviroli, D., Dürr, H. H., Messerli, B., Meybeck, M., and Weingartner, R.: Mountains of the world, water towers for humanity: Typology, mapping, and global significance. Water resources research, 43(7), 2007.

Viviroli, D., Archer, D. R., Buytaert, W., Fowler, H. J., Greenwood, G., Hamlet, A. F., ..., and Lorentz, S.: Climate change and mountain water resources: overview and recommendations for research, management and policy. Hydrology and Earth System Sciences, 15(2), 471-504, 2011.

Weerts, A. H., and El Serafy, G. Y.: Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. Water Resources Research, 42(9), 2006.

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.
WSL Institute for Snow and Avalanche Research SLF: Biweekly manual snow profiles from Weissfluhjoch, Davos, Switzerland, Dataset, doi:10.16904/2, 2015a.
WSL Institute for Snow and Avalanche Research SLF: Meteorological and snowpack measurements from Weissfluhjoch, Davos, Switzerland, Dataset, doi:10.16904/1, 2015b.
Zappa, M., Pos, F., Strasser, U., Warmerdam, P., Gurtz, J.: Seasonal water balance of an Alpine catchment as evaluated by different methods for spatially distributed snowmelt modelling. Hydrology Research, 34(3), 179-202, 2003.
Zheng, Z., Kirchner, P. B., and Bales, R. C.: Topographic and vegetation effects on snow accumulation in the southern Sierra Nevada: a statistical summary from lidar data. The Cryosphere, 10(1), 257-269, 2016.

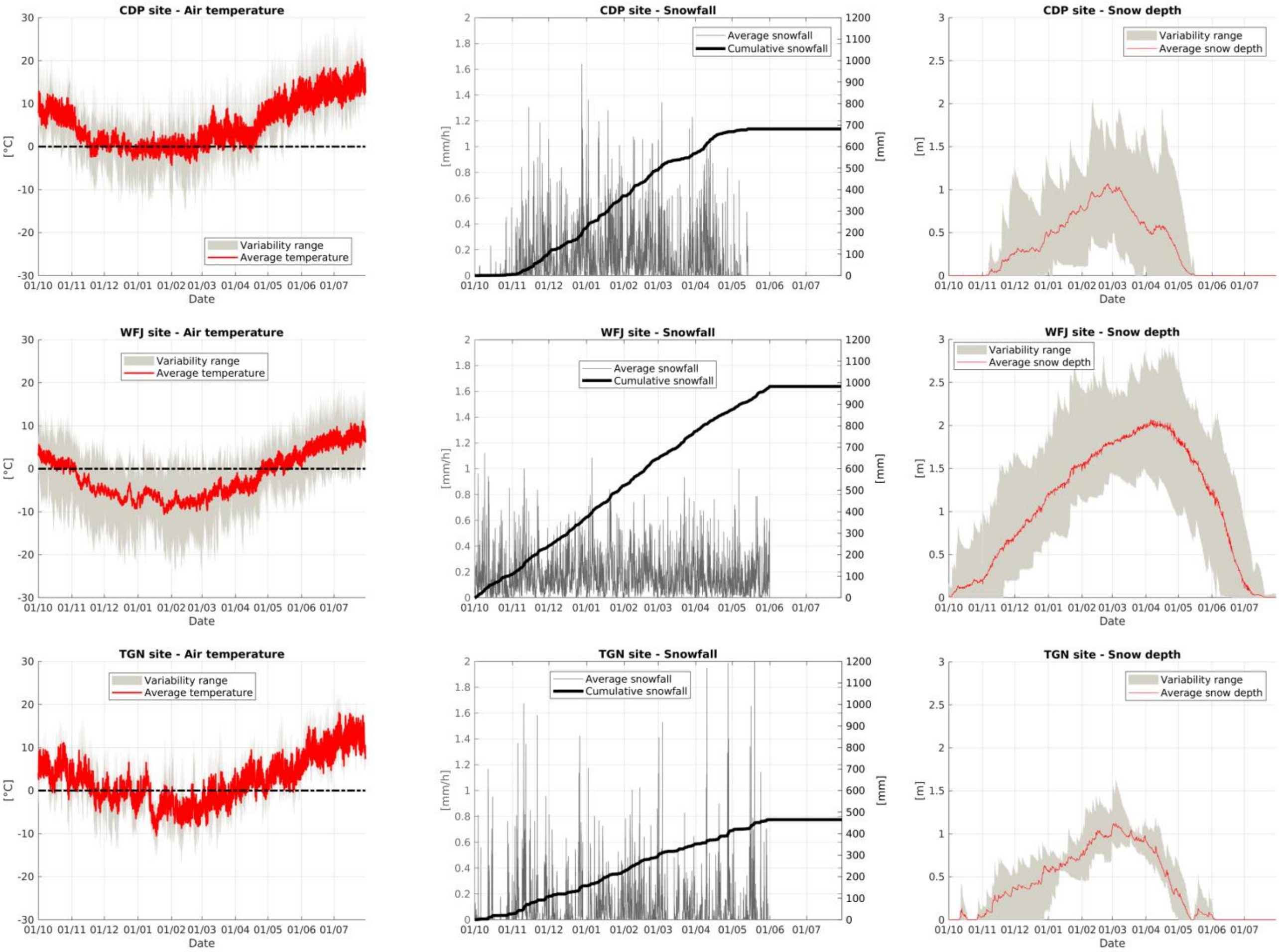


Figure 1: Meteorological characterization of CDP site (first row), WFJ site (second row), TGN site (third row) - Air temperature (left column), snowfall rates (middle column; at WFJ and TGN sites snowfall rates have been estimated according to Froidurot et al., 2014) and snow depth (right column) throughout an average snow season (early October – early July) throughout the overall datasets.

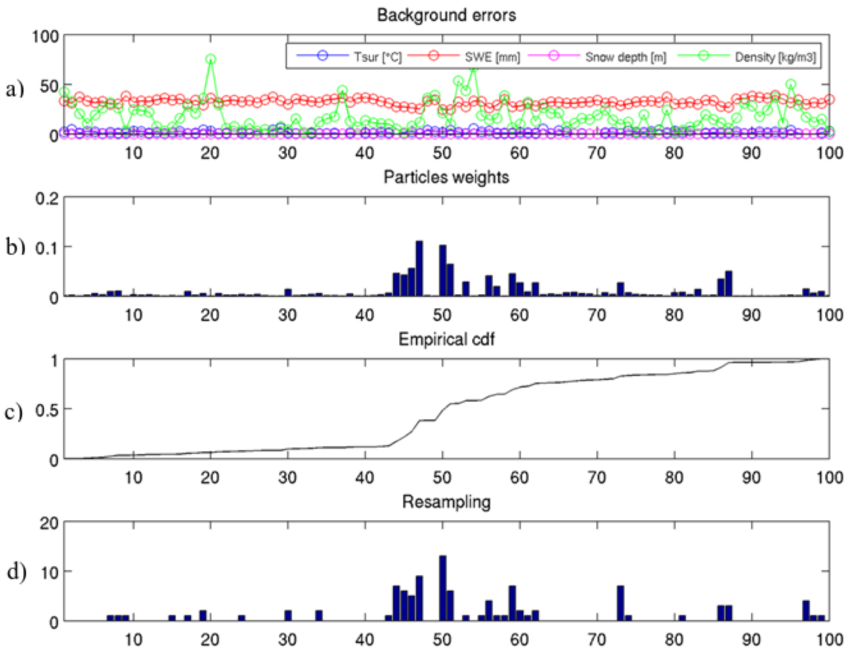


Figure 2: SIR-PF scheme for multivariate DA: (a) Open circles are the background errors for surface temperature (in blue), SWE (in red), snow depth (in magenta) and snow density (in green). (b) Importance weights as a function of the particles indices. (c) Empirical cdf of the weights. (d) Number of resampled particles as a function of the particles indices.

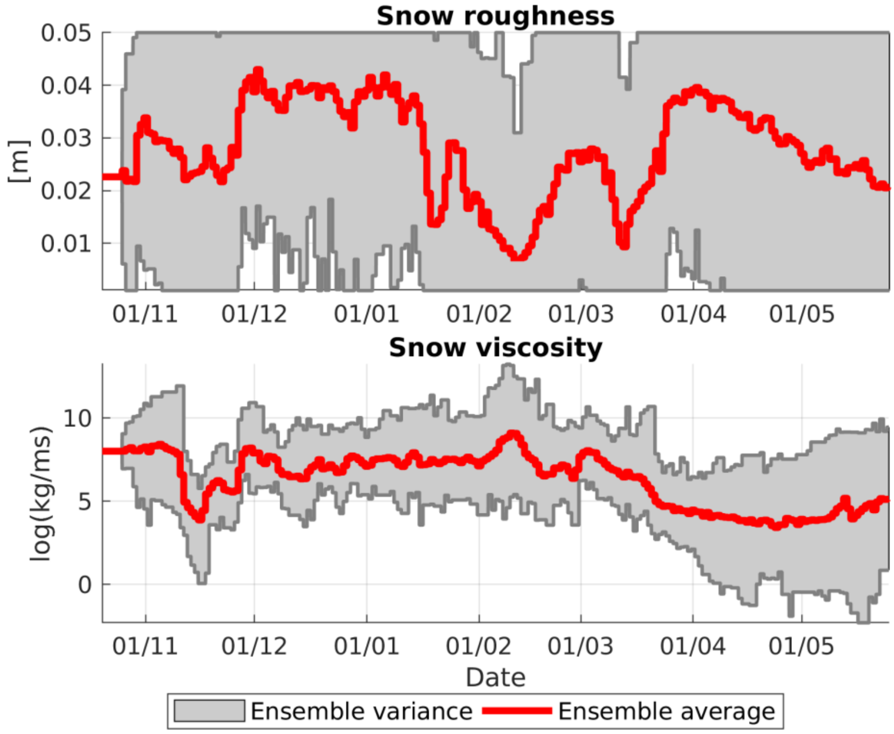
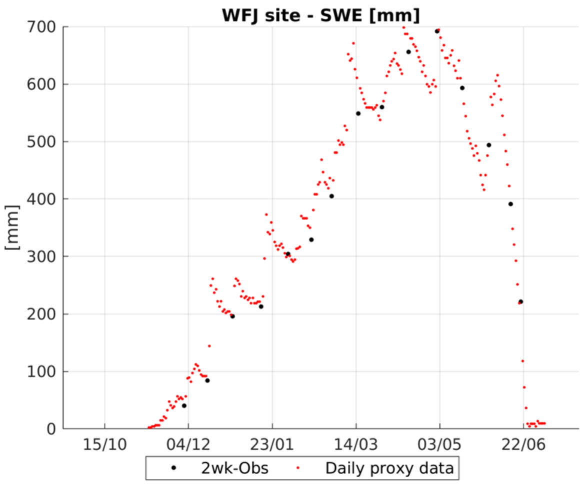


Figure 3: Uncertainty of model parameters – CDP site – Winter season 2010-2011. The panel shows the ensembles seasonality of snow roughness and snow viscosity.



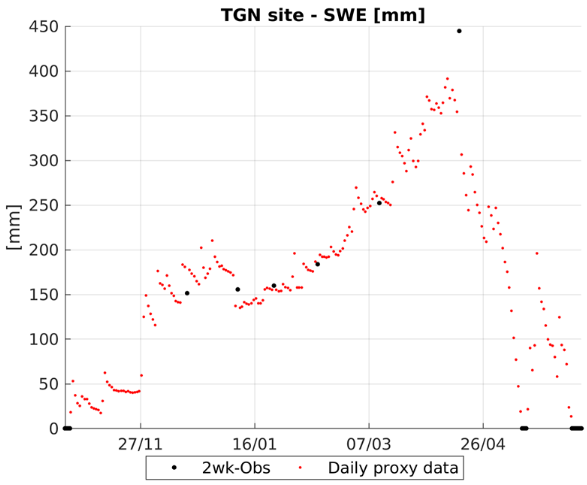


Figure 4: Additional snow density model –Comparison between SWE measurements and indirect proxy estimates - WFJ site, winter season 2005/06 (on the left); TGN site, winter season 2012/13 (on the right).

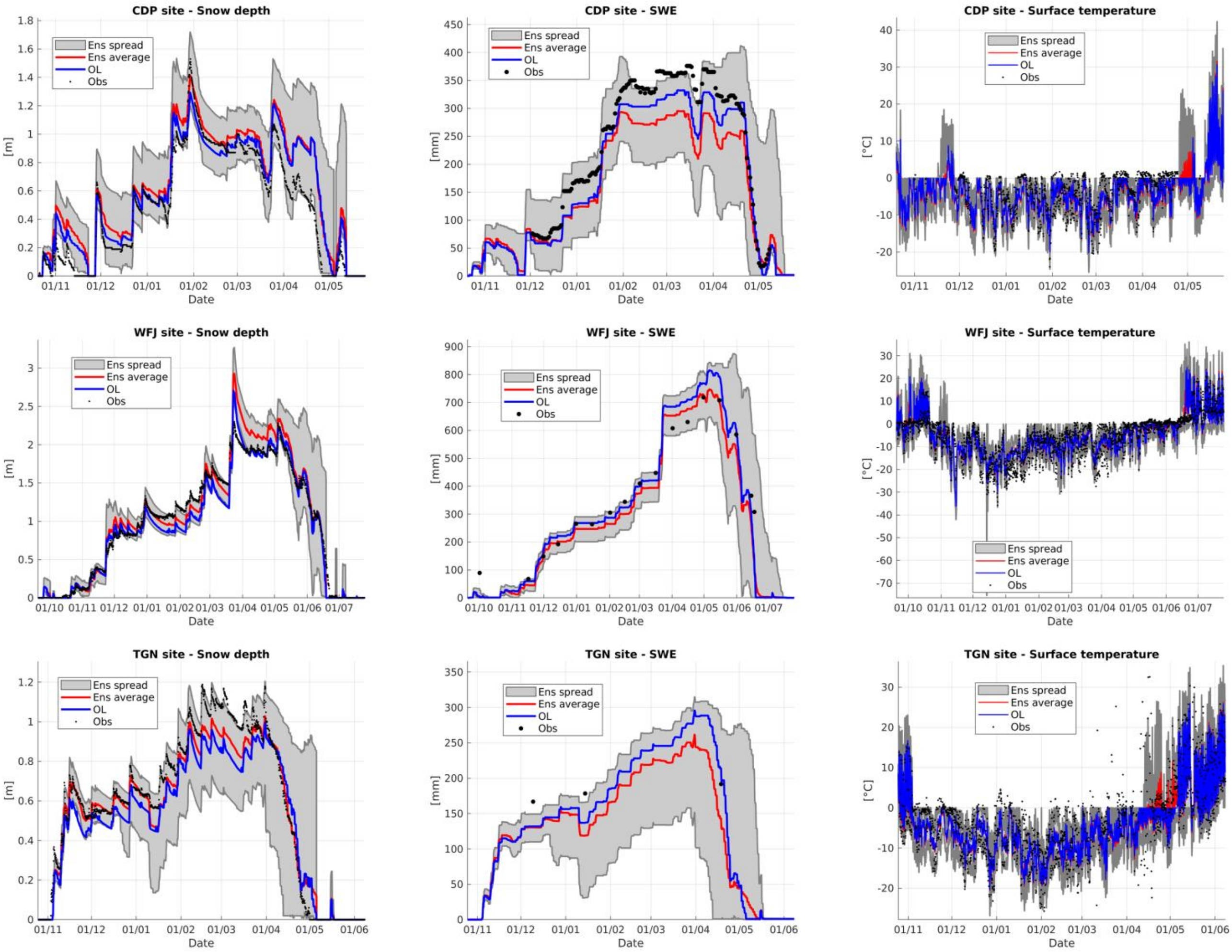
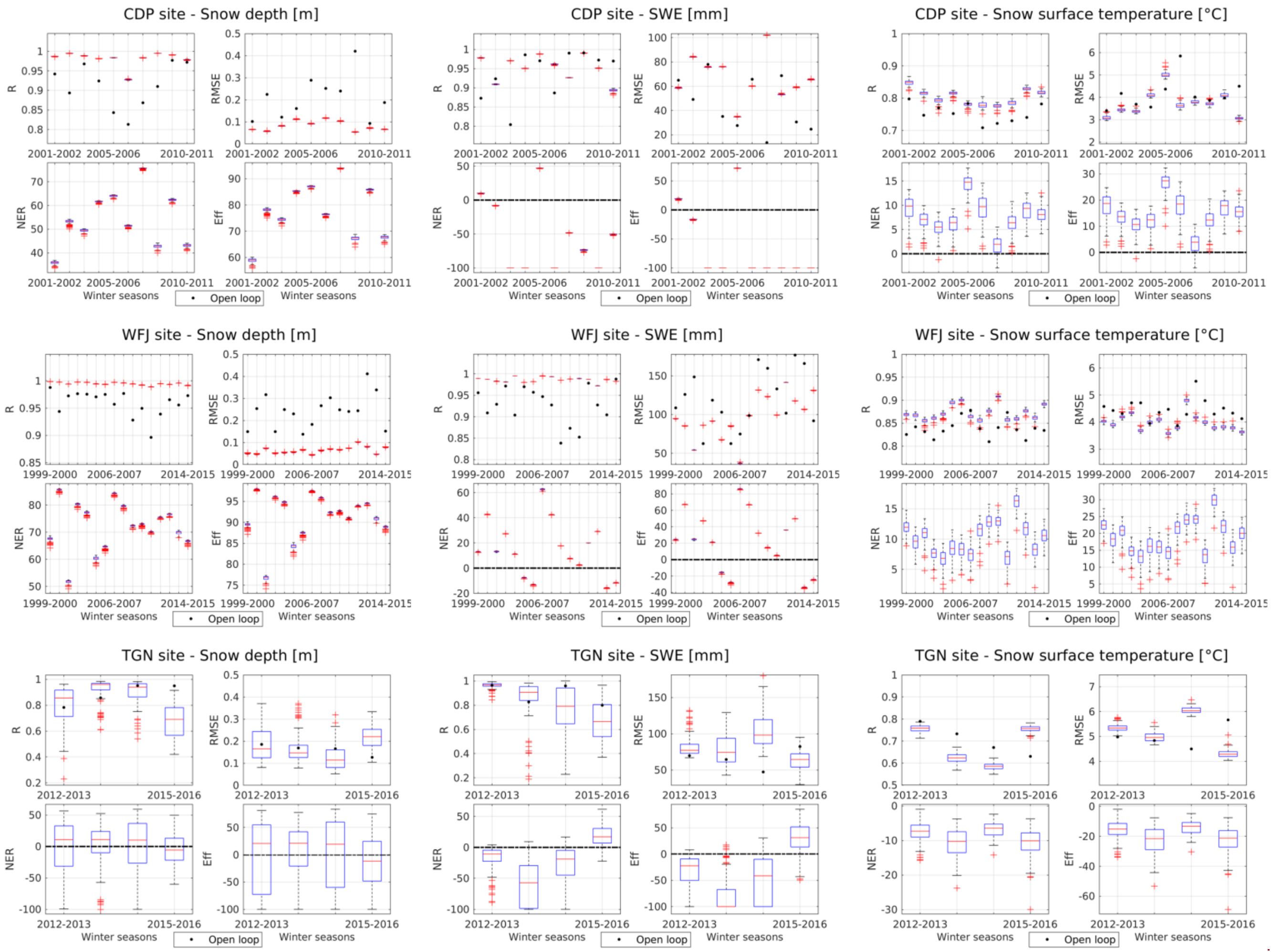


Figure ~~3~~5: Impact of the meteorological uncertainty - Ensemble simulations of snow depth (left column), SWE (middle column), and surface temperature (right column) – CDP, winter season 2003-2004 (first row); WFJ, winter season 2001-2002 (second row); TGN, winter season 2014-2015 (third row).



~~**Figure 4: Multivariate DA scheme with perturbed meteorological data – Statistical scores of snow depth (left column), SWE (middle column) and surface temperature (right column). The bottom and top edges of each box indicate the 25th and 75th percentiles, respectively.**~~

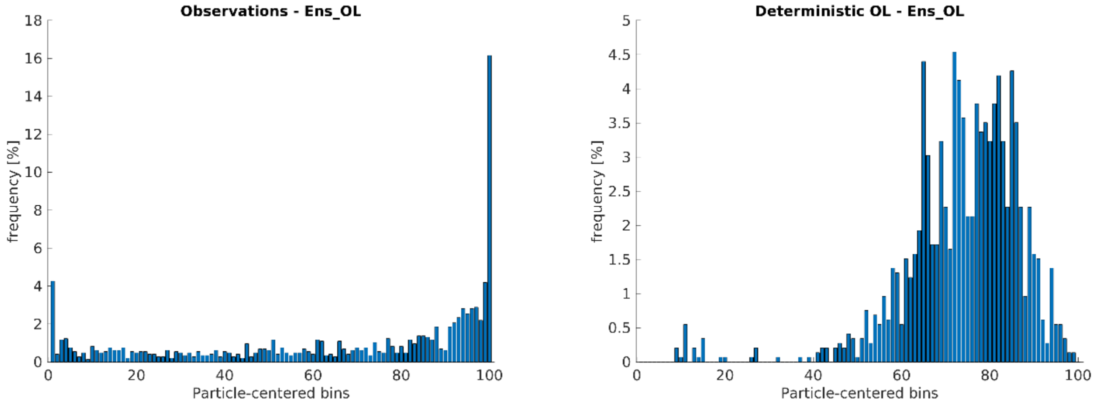


Figure 6: Talagrand diagram – Analysis rank histogram of SWE ensemble open loop (Ens_OL) simulations considering SWE observations (on the left) and the SWE deterministic open loop predictions. CDP dataset (10 winter seasons).

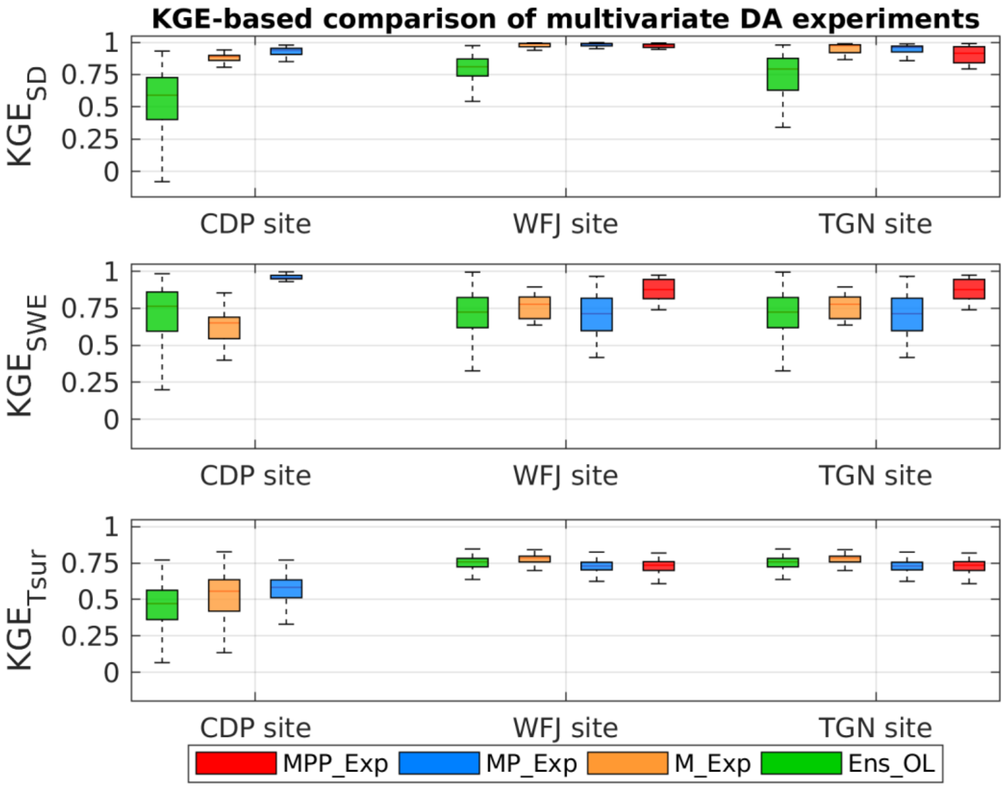
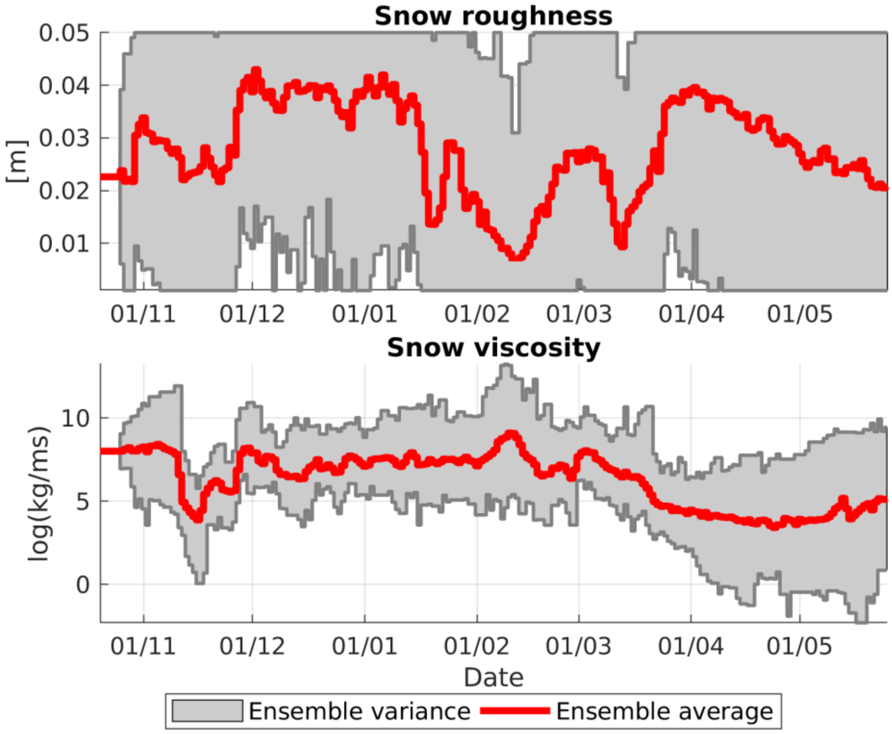
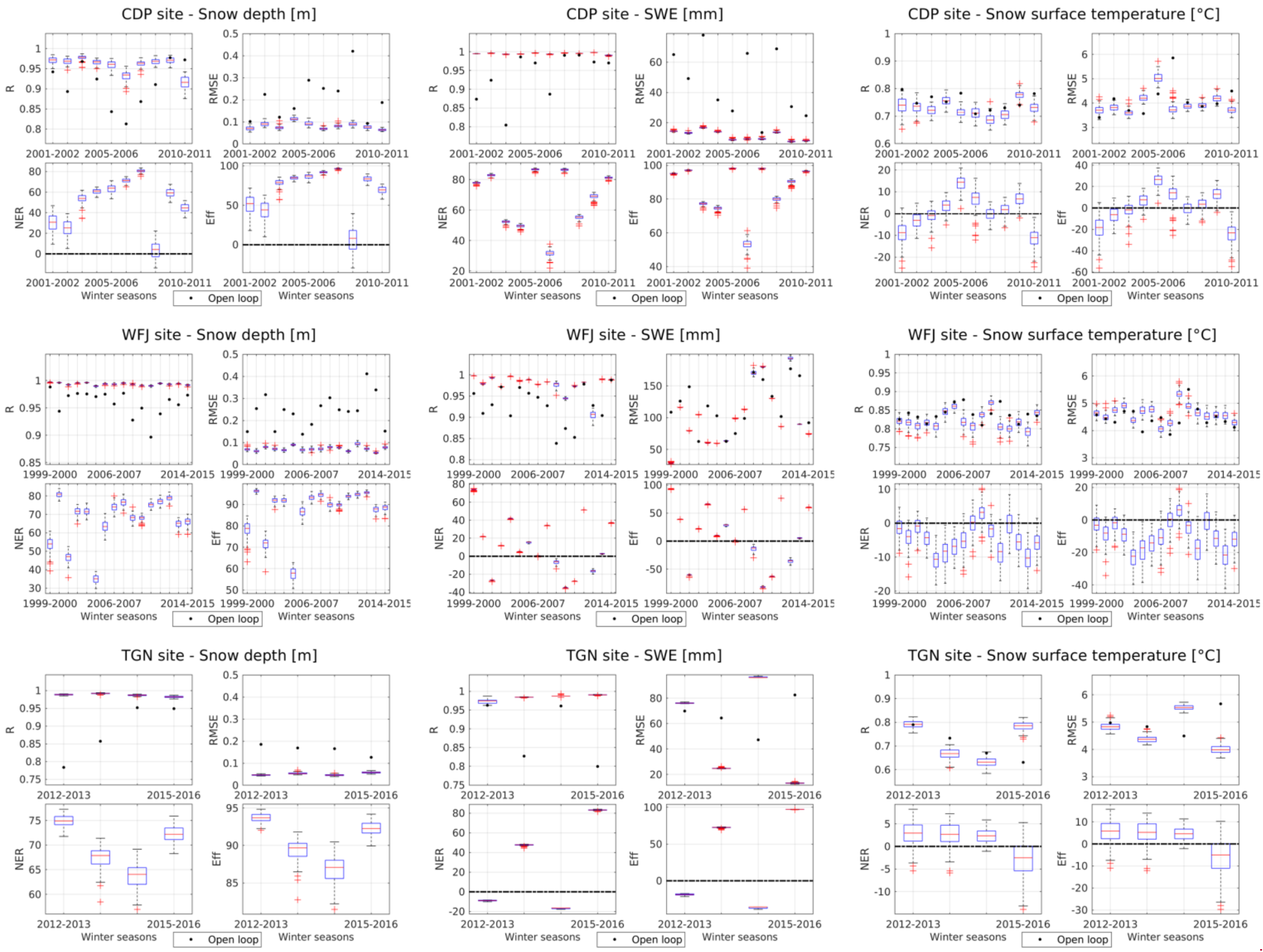


Figure 7: Multivariate DA experiments – Multi-year KGE values of snow depth (SD), SWE, and surface temperature (T_{sur}) simulations. The Ens_OL, M_Exp, MP_Exp (1), and MPP_Exp statistical scores are in green, orange, blue and magenta, respectively. The bottom and top edges of each box indicate the 25th and 75th percentiles, respectively. The line in the middle of each box is the median. The whiskers extending above and below each box indicate the 5th and 95th percentiles.



~~Figure 5: Uncertainty of model parameters – CDP site – Winter season 2010-2011. The panel shows the ensembles seasonality of snow roughness and snow viscosity.~~



~~Figure 6:Multivariate DA scheme with perturbed meteorological data and model parameters resampling – Statistical scores of snow depth (left column), SWE (middle column) and surface temperature (right column) simulations – CDP site (first row); WFJ site (second row); TGN site (third row).~~

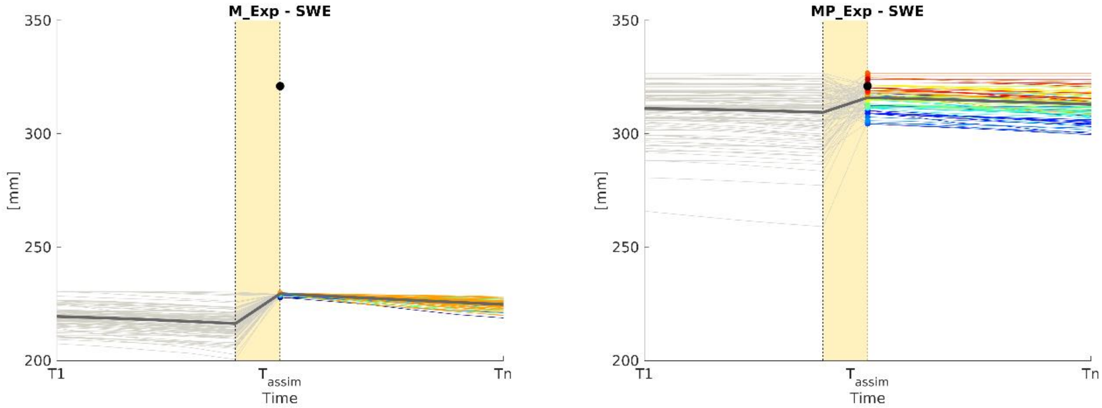


Figure 8: Particles resampling at an assimilation time step of SWE observation – M_Exp vs MP_Exp – In grey are the ensemble members, whose ensemble mean is represented thicker; the black dot is the observation; the resampled particles are multicoloured.

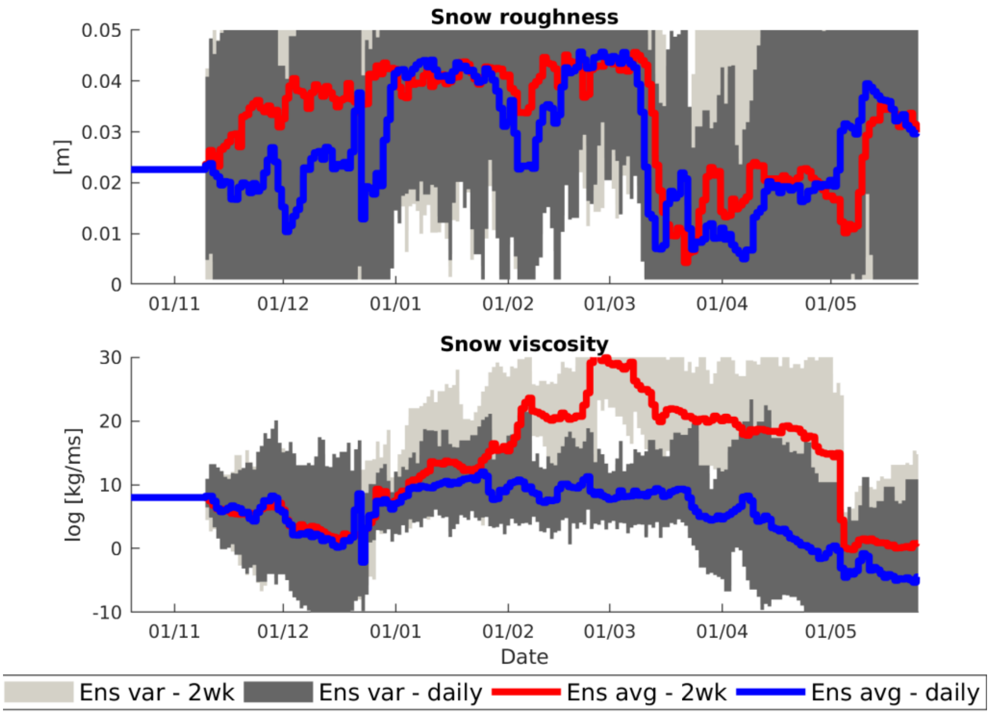


Figure ~~7~~9: Sensitivity analysis of the multivariate DA scheme to SWE measurement frequency at CDP site, winter season 2001/02 – Parameters ensembles: snow roughness (on top) and snow viscosity (second row) resulting from the assimilation of daily (average trend in blue) and biweekly (red) SWE observations.

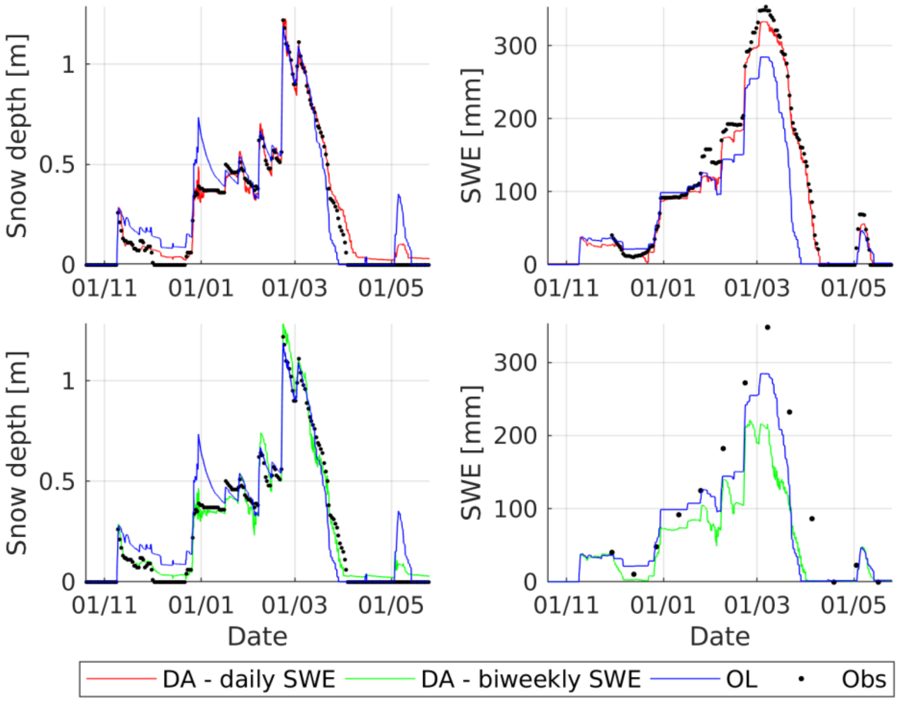
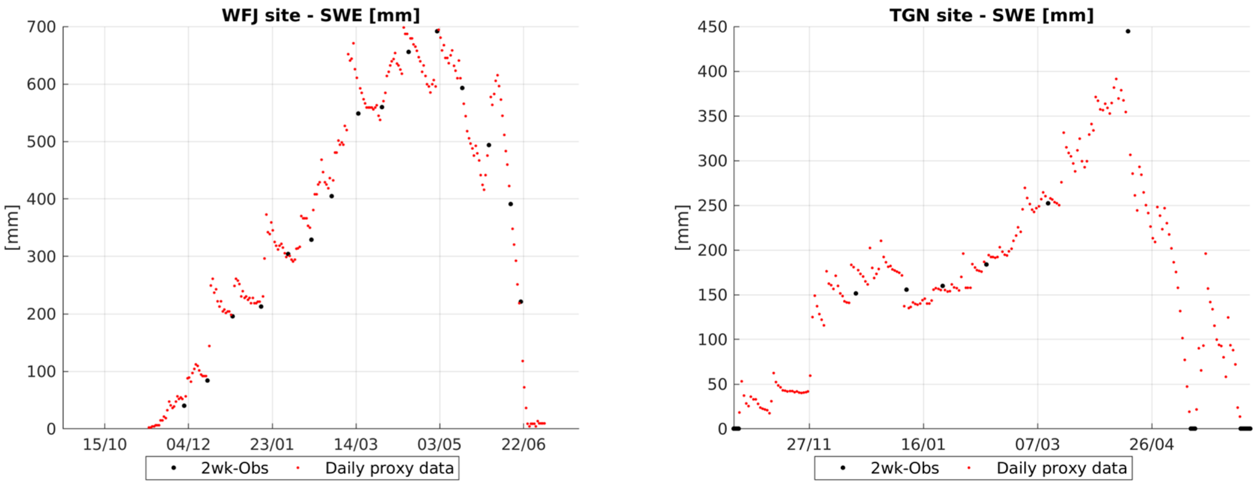
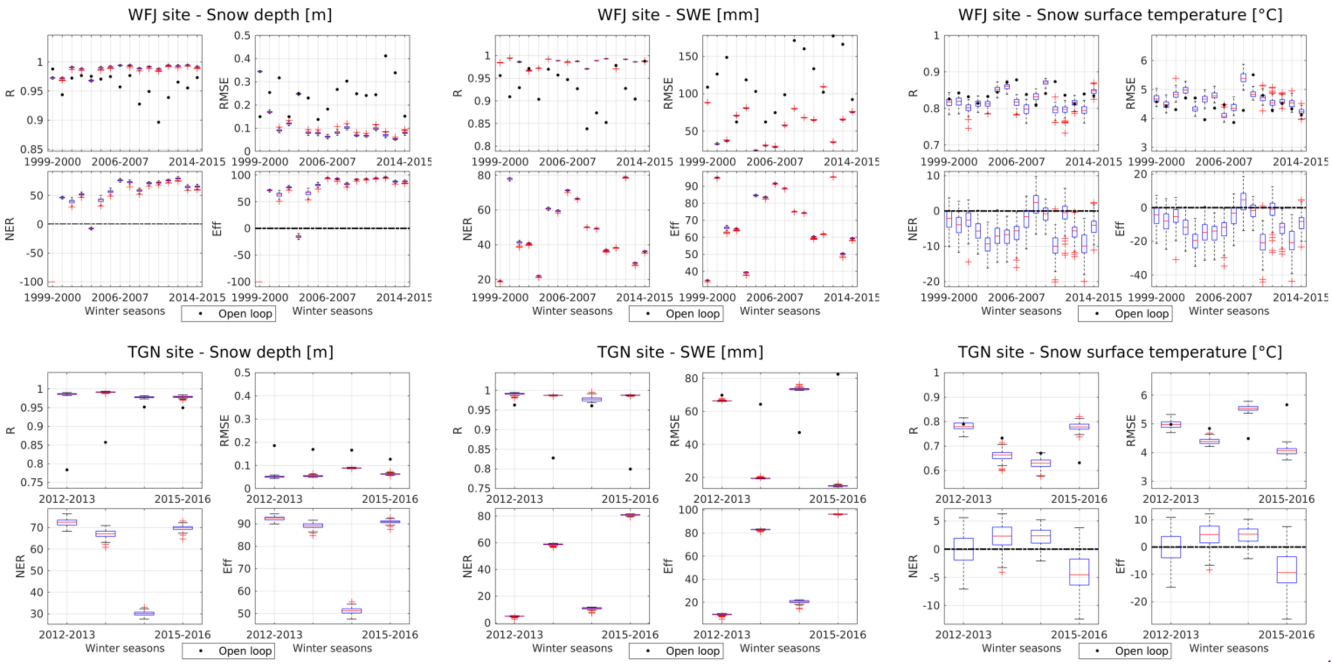


Figure ~~8~~10: Sensitivity analysis of the multivariate DA scheme to SWE measurement frequency at CDP site, winter season 2001/02 – Mean ensemble s~~S~~imulations of snow depth (left column) and SWE (right column) for daily (first row) and biweekly (second row) SWE measurements.



~~**Figure 9: Additional snow density model - Comparison between SWE measurements and indirect proxy estimates - WFJ site, winter season 2005/06 (on the left); TGN site, winter season 2012/13 (on the right).**~~



~~Figure 10: Multivariate DA scheme with auxiliary snow model – Statistical scores of snow depth (left column), SWE (middle column) and surface temperature (right column) simulations – WFJ site (first row); TGN site (second row).~~

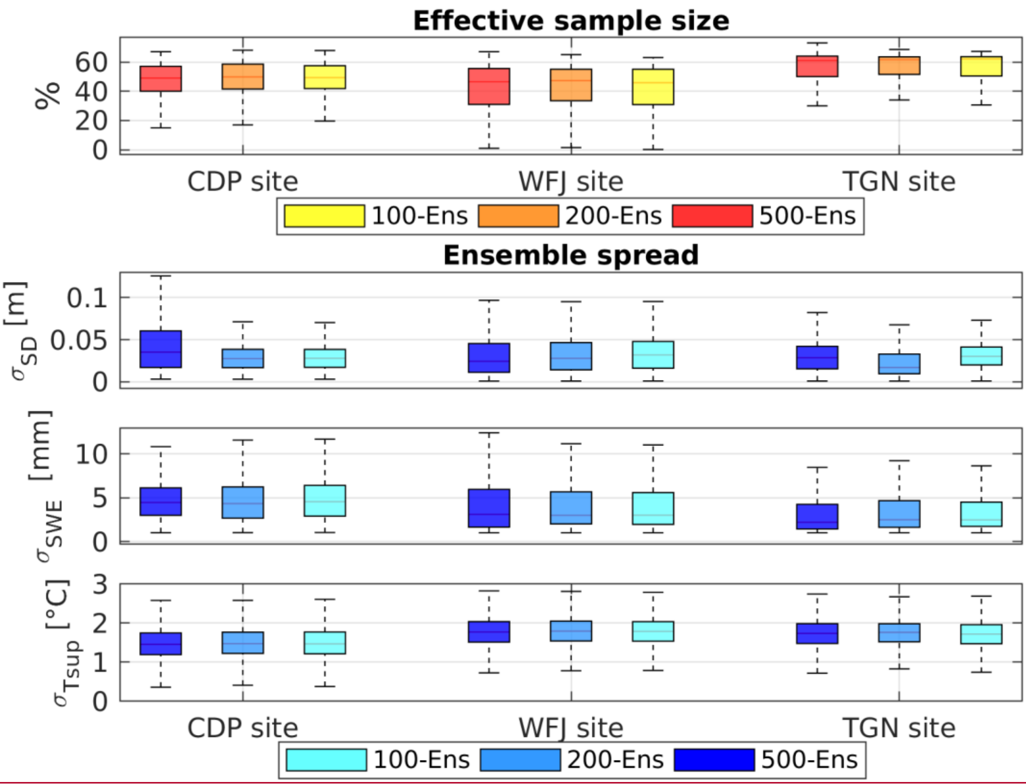


Figure 11: nP_Exp – Effective sample size and ensemble spread as a function of the ensemble size.

Variable		Unit	Distribution	μ	σ	τ [h]	Lower limit	Upper limit
Air temperature	T_a	[°C]	Normal	0	0.9	4.8	-	-
Relative humidity	RH	[%]	Normal	0	8.9	8.4	0	100
Solar radiation	SW	[W/m^2]	Normal	0	min(SW, 109.1)	3.0	0	-
Precipitation	P	[mm]	Lognormal	-0.19	0.61	2.0	0	-
Wind speed	V	[m/s]	Lognormal	-0.14	0.53	8.2	0.5	25

Table 1: Error statistics for the generation of the meteorological ensembles (Magnusson et al., 2017).

Site	Dataset size	Snow seasons		Reference
		from	to	
CDP	10-years	2001/2002	2010/2011	Morin et al., 2012
WFJ	16-years	1999/2000	2014/2015	WSL Institute for Snow and Avalanche Research SLF, 2015b
TGN	4-years	2012/2013	2015/2016	Galvagno et al., 2013

Table ~~2~~3: *Snow datasets.*

	Assimilated variables				
Site	T_{snow}	SWE	α	SD	ρ_{snow}
	[°C]	[mm]	[-]	[m]	[kg/m^3]
	σ_{obs} = 2°C	σ_{obs} = 10 mm	σ_{obs} = 0.15	σ_{obs} = 0.05 m	σ_{obs} = 50 kg/m^3
CDP	hourly	daily	hourly	hourly	**daily**
WFJ	30-min	bi-weekly	daily	30-min	**bi-weekly**
TGN	30-min	daily (2013/14, 2015/16); **bi-weekly (2012/13, 2014/15)**	daily (2012/13)	30-min	**daily (2013/14, 2015/16)**; bi-weekly (2012/13, 2014/15)

Table ~~3~~4: Measurement frequency of the assimilated variables at each experimental site: snow surface temperature (T_{snow}), SWE, albedo (α), snow depth (SD) and snow density (ρ_{snow}) – Observational uncertainties (σ_{obs}) are reported - Variables indirectly estimated are highlighted in bold.

	Parameter		Nominal value	Range
1.	Snow roughness	[mm]	0.0226	$[0.001 - 0.05]$
2.	Snow viscosity	[kg/ms]	10^8	$[7 \cdot 10^7 - 10^{12}]$
3.	Albedo parameter τ_α	[s]	10^7	$[0.52 - 1.55 \cdot 10^7]$
4.	Albedo parameter τ_m	[s]	$3.6 \cdot 10^5$	$[1.73 - 5.2 \cdot 10^5]$
5.	Albedo parameter S_0	[mm]	10	$[2 - 15]$

Table 42: Sensitivity analysis - Selected model parameters.

Experiment	Ensemble generation	Assimilation period	Specific details
M_Exp	Perturbation of input data	3 hours	-
MP_Exp (1)	Perturbation of input data & model parameters	Daily	-
MP_Exp (2)	Perturbation of input data & model parameters	Daily	Sensitivity to SWE measurement frequency
MPP_Exp	Perturbation of input data & model parameters	Daily	Additional snow density model
nP_Exp	Perturbation of input data & model parameters	Daily	Sensitivity to the particles number

Table 5: List of multivariate DA experiments.

	Surface temperature [°C]			SWE [mm]			Snow depth [m]		
	CDP	WFJ	TGN	CDP	WFJ	TGN	CDP	WFJ	TGN
M_Exp	0,097	0,128	0,112	-1,040	0,007	0,365	0,651	0,764	0,557
MP_Exp (1)	0,108	0,062	0,111	0,746	0,007	0,583	0,813	0,833	0,694
MPP_Exp	-	0,050	0,092	-	0,467	0,639	-	0,701	0,565

Table 6: Comparison of DA performance under different configurations: values of CRPSS score.

	Surface temperature [°C]			SWE [mm]			Snow depth [m]		
	CDP	WFJ	TGN	CDP	WFJ	TGN	CDP	WFJ	TGN
OL-Ens	0,369	0,774	0,591	0,893	0,864	0,361	0,182	0,855	0,798
100-Ens	0,494	0,744	0,616	0,995	0,937	0,450	0,937	0,959	0,812
200-Ens	0,484	0,741	0,636	0,991	0,959	0,466	0,925	0,990	0,814
500-Ens	0,485	0,743	0,617	0,993	0,962	0,537	0,933	0,968	0,819

Table 7: nP-Exp - Values of KGE coefficient as a function of the ensemble size.