GMD review

We would like to thank both referees for their extensive analysis of our manuscript which we believe helps a lot improving our paper. All the comments have been addressed and point by point response is provided below each comment. Note that some slight changes were made in the manuscript in order to improve its clarity, and are visible in the track changes. In the following, the reviewer initial comments are written in black, our answer in blue and the corrections in the paper are highlighted in red. Line references for modifications correspond to the initial submitted version of the manuscript, not the modified.

Reviewer 1

In this study, the authors developed two variants of the particle filter (PF), named the global PF and the klocal PF, to assimilate snow depth and reflectance for snow water equivalent (SWE) estimation. The global PF assimilation all observations in the domain while the klocal PF is a localized PF that assimilate only a subset of observations. To prevent the degeneracy of PF, the global PF inflates the observation error covariance until a sufficient number of replicas are available, while the klocal approach applies the maximum of "k" observations to maintain a sufficiently large observation-state variable variation. Some notable assumptions include the observations are free of noise, error, and correlation in space and time, and the prior estimates and the observations are generated from the same model (identical twin). The results prove that the inflations and the k-localization effectively prevent the degeneracy, and the PF systems are able to spread the observed snow signal to non-observed areas.

This is a nice contrition to the existing PF literature and has the potential to significantly extend the applicability of PF. The study fits the scope of the journal. I hope the authors consider the following comments in the revision:

The authors would like to thank Reviewer 1 for his/her thorough review and his/her questions on several subjects (the semi-distributed geometry, the methodology and assumptions, and the potential shortcomings of the different assimilation algorithms) which deserved more details and rigor in the formulation. We would like also to thank Reviewer 1 for expressing his/her need for more physical explanation on the ensemble correlation patterns of Fig. 4. We believe that these comments helped a lot improving the clarity of the manuscript, and we hope that the corrections fully address the reviewer comments.

1. The domain is divided into classes based on elevation band, aspect and slope, but there is no information regarding the geographic distribution of these classes. The PF's performance is generally good in high-elevation areas, but performance variations still exist among these areas. Could this be a result that the observation improve the more local classes more than the class that is farther away from the observation?

The semi-distributed framework does not allow to define a horizontal euclidian distance between topographic classes. Therefore, we do not consider any variability of the horizontal proximity between classes. However, in mountainous environments, topographic conditions often more directly drive snowpack variability than distance. As the reviewer points out, there is indeed a difference in performance between the observed classes and the unobserved classes, the former achieving better improvements, in general (see Figs. 6-7 and Sec. 4.2.1), and locations that are farther away (in model space) from the observations achieve the lowest performance (e.g. Fig. 8b 2100m, North, 40 degrees). Furthermore as the reviewer notes, there is also a notable variability of performance even among the observed classes, in particular for the reflectance assimilation.

According to this comment, the end of Sec. 4.2.2 (l. 331 of the manuscript) was amended to be more precise and descriptive:

Fig. 8. shows the spatial performance of the different algorithms for member 2016_p60. Spatial patterns similar to the HS assimilation are found. rlocal performance is limited to the observed classes, while global and klocal manage to improve the simulations across aspects and slopes. However, skill scores are lower than for HS (0.2-0.5), and the performance of all algorithms is poor in the classes that are the farther away from the observations, i.e. at lower elevations (600-900 m) and in some of the high altitude steep Northern classes (e.g. 2100_N_40 on Figs. 8b-c). Finally note that slight degradations of performance can sometimes be evidenced even in the observed classes for all the algorithms (e.g. in flat conditions at 3300 m on Fig. 8}a for the rlocal}, not evidenced by this example for the other algorithms).

2. Some discussions on the assumption and the feasibility-testing nature of the system is needed in the abstract or be acknowledged in the introduction section.
The feasibility-testing nature and the identical twin setup of this experiment were indeed not acknowledged enough in the abstract. This is now corrected on (L 13.14):
...based on background correlation patterns. Feasibility-testing experiments are carried out in an identical twin experiment setup, with synthetic observations of HS and VIS-NIR reflectances available in only a 1/6th of the simulation domain. …

Another notable assumption, the fact that observations are not corrupted, needed to be underlined and justified. We are actually conscious of this limitation, and a recent study has been submitted (Revuelto et al., submitted) in which we assimilate synthetic corrupted observations at the point scale. In our situation we did not corrupt the observations because little is known about the spatial structure of errors of reflectance (e.g. Cluzet et al., 2020): we know that assuming independent errors (i.e. diagonal R) is a very rough approximation of the reality which has strong consequences on the propagation of information. Corrupting the observations with such random structures would be theoretically more consistent, but would not yield much more insight on the potential for information from real observations to be spatially propagated as real spatial correlation of observation errors might be very different from this hypothesis. Future efforts should concentrate in better characterizing these spatial structures of errors. Consistently, the following sentence was modified at the beginning of Sec. 3 (L. 206)
Synthetic observations are extracted from a model run and assimilated without adding any noise. These observations mimic…

and a paragraph was added in the end of Sec. 5.2:
Regarding the observations, our study has some methodological limits, however. Observation errors are very roughly prescribed, and the assimilated observations are not corrupted as usually done in synthetic experiments (e.g. Durand et al., 2006). These choices were motivated by the fact that very little is known about the spatial correlation of reflectance observation errors in the semi-distributed setting (Cluzet et al., 2020). In a recently submitted paper, the impact of random and systematic errors of reflectance observations on point-scale assimilation experiments is thoroughly investigated (Revuelto et al., in prep). Efforts to better characterize these observation errors should be conducted in future work

Lastly, the synthetic nature of the experiments should be stated in the conclusions. The sentence on L490 was changed to:
In the framework of synthetic experiments, we have shown in particular that:…

In addition to the assumptions mentioned above, the depth observation error is assumed to be 0.1m (error covariance is 1e-2m^2), which is quite a high-bar for existing observation techniques, especially when used on space-borne platform for large-scale measurements.

We agree that the prescribed observation error is a high-bar for space-borne sensors. Indeed, results from recent studies such as Eberhard et al., (2020) could be used to provide a more accurate estimate of HS retrieval errors from satellites. Conversely, it could be considered as a low value for other sources of HS observations (e.g. stereo satellite imagery, Deschamps-Berger et al., 2020; local measurements with a high spatial representativeness error). As our work is a feasibility-testing experiment based on synthetic observations, an arbitrary observation error was chosen but indeed it may be important to adjust this value when applying the algorithm to real observations. This is now mentioned in the discussion on line 371:

Global and klocal algorithms exhibit strong performances when assimilating HS (Fig. 5). HS is closely linked with the SWE (by the bulk density) and the interest of this variable for data assimilation is clear (Margulis et al., 2019). Here, it should be kept in mind that HS assimilation is used as a baseline experiment to evaluate the algorithms and put reflectance assimilation into perspective. The prescribed HS observation errors ($\sigma_0=0.1m$) are not necessarily realistic. They should be adapted to the nature of the HS sensor. For example, space-borne HS observation errors are typically larger (e.g. Eberhard et al., 2020; Deschamps-Berger et al., 2020). The assimilation of such observations would probably yield lower improvements.

Though the performance is lower for Reflectance than in our HS experiments, it remains considerable and in line with previous results on point simulations (Charrois et al., 2016), with an average score improvement of 20-40\%…

Finally, note that the inflation procedure inside the global and rlocal approaches modifies the observation error which is assumed to be poorly known, reducing the impact of the prescribed value as mentioned in Sec. 2.3.1.

3. Line 27: panel a of Figure 1 does not look like flat âĂŤ the surface does seem to make an angle with the level surface (the brown triangle)
This panel is actually flat, but we agreed the perspective view might be misleading. For this reason, we changed the background color of Fig. 1 in order to reinforce the perspective view, hoping that it helps.
Changed Fig. 1.

4. Line 128: it would be useful to include more details of the perturbation for each key forcing variable, like what perturbation models and error statistics are used, and whether spatial correlations are considered.
We agree tat this part was too elusive. Spatial correlations are not considered (i.e. equal to one) this is what we meant with "spatially homogeneous" (l. 128 of the manuscript), but this formulation could be misleading and more details were added.. For the sake of clarity, we also add the mention that perturbations are temporally correlated. The sentence was therefore modified accordingly
Before the beginning of the simulation, spatially homogeneous stochastic perturbations (e.g. at a given date, the same perturbation parameter is applied across the whole domain) with temporal auto-correlations are applied to this forcing to generate an ensemble of forcings.

In addition, an appendix was added giving more details on the perturbation procedure and parameters.
Added appendix A.

5. Figure 2: how do the forcing particle (Fi) and the model particle (Mi) get paired? Is it random or does it follow some protocol?
Yes, the pairing is random and keeps the same during the whole simulation. For the sake of clarity, the line 130-131 was changed to:

At the beginning of the simulation, each forcing F_i is associated with a random M_i ESCROC configuration and this relation is fixed during the whole simulation.

6. Line 180: can posterior estimates form the klocal approach show spatial discontinuity, since each area is updated independently by different measurements?

Thanks for underlining this point. Yes indeed klocalisation generates spatial discontinuities, it is one of the common drawbacks of localised approaches (see Farchi and Bocquet, 2018, already cited, for a thorough review). However, we expect the k-localisation to produce similar analyses (i.e. PF samples) for similar locations because their analyses will be based on similar sets of observations, thereby reducing the discontinuities compared to the r-local approach. In our setup, this has no direct consequence on the simulation as simulation points are independent, but it can hamper the interpretation of the spatial patterns of individual members. We changed the following lines inside the introduction (l. 69 of the manuscript):

It makes it possible to constrain the model in locations that are not directly observed, but with nearby observations. Contrary to global approaches, localisation has the disadvantage of producing spatially discontinuous analyses (each point receives a different analysis). This issue can be mitigated in various ways (Poterjoy, 2016; Farchi and Bocquet, 2018; Van Leeuwen et al., 2019). The underlying hypothesis…

Furthermore, we discussed this point in the end of Sec. 5.3:

Finally, in the case of modeled coupling between simulation points (e.g. snow drift), which was not the case here, the spatial discontinuities of the klocal analyses (see Sec. 1) might be a drawback compared to the global approach. Spatial discontinuities may reveal impractical for the interpretation of individual simulations outputs by snow forecasters too. The klocal approach is likely to reduce these discontinuities compared to the rlocal}, because similar locations will receive similar analyses (i.e. based on similar sets of observations). This issue could be partly mitigated by e.g. state-block-domain approaches (Farchi and Bocquet., 2018).

7. Line 195: how are the 10% and 0.3 here determined? Are they from previous literature or are there sensitivity test?

These parameters were adjusted during preliminary design experiments. As reflectance is not defined in the absence of snow, the number of pairs available to compute correlations between two locations varies for reflectance, and spurious high correlations are found when there is a very low number of common members. Regarding the 0.3 value, it is also an adjustment, based on the idea that if correlations are too low, it does not make sense to try to propagate information, as there will likely be a negative impact or no impact. The correlations exhibited on Fig. 4 enables the reader to realize typical (open-loop) correlation values with 40 members. We agree that a most rigorous definition based on significance levels would probably be a better option, and we will investigate this in future works. The following sentence on L196. was modified:

…, and match the following criteria: which were adjusted in preliminary experiments:
\begin{itemize}
\item in $\bm{x}^i_v$, there are at least 10\% of members defined in both points. As reflectance is not defined when there is no snow, spurious high correlations can be obtained when the computation of correlations is based on a very low number of pairs.
\item $\lvert \mathbf{B}_{\bm{v}}(n,p) \rvert >0.3$. If the absolute correlation is too low, it is likely that there is a poor potential for the distant observation to constrain the ensemble locally. In such a situation, it is better to reject the observation from the local analysis. Negative ensemble correlations can be physically sound, e.g. after a rain-on-snow event between the HS of two points separated by the rain-snow line. In such a situation, an HS observation on either point can hold information on precipitation rates at both locations. At the observed location, the PF will select the members with the

most appropriate precipitation rates. This sample is likely to perform well at both locations, so it can be used to constrain the unobserved location.
\end{itemize}

8 Line 239: the PF performance with band4 and band5 observations are quite different (as in Figure 4), what could be the reason?
 Note that Fig. 4 does not present the skill of an assimilation experiment, it is an example of open-loop ensemble background correlation patterns for band 4 and band 5 on a specific date. Regarding the interpretation of these results, there was a lack of physical explanations to help interpret the correlations of Band4 and Band5. These observations are sensitive to the snowpack surface properties, namely the specific surface area (SSA, m²/kg) and light absorbing impurities content (LAP, g/g_snow). This is now stated in the introduction (line 30-31 of the manuscript):
For instance, snowpack VIS-NIR reflectances from moderate resolution (250-500 m) satellites such as MODIS or Sentinel-3 can help constraining the snowpack surface properties such as microphysical properties (characterized by the specific surface area, SSA (m^2kg^{-1}) and light absorbing particles content (LAP, (gg_{snow}^{-1}))) (Durand et al., 2006; Dozier et al., 2009).

The individual sensitivity of the spectral reflectances is now further detailed (l. 232-233).
Reflectance is sensitive to the surface SSA and LAP (see Sec. 1). A minimal set of two different bands is used, corresponding to MODIS sensor band 4 (555 nm, sensitive to SSA and LAP) and 5 (1240 nm, mainly sensitive to SSA) (e.g. Fig. 2. of Cluzet et al., 2020).

A slight adjustment of the interpretation of Fig. 4 was performed to point negative correlations for Band5:
...being substantially correlated with the considered class. Note that negative correlations are evidenced with some lower altitude South-oriented topographic classes (e.g. 1500_S_40 on Fig. 4b). Finally, these patterns...

Indeed, the reasons why the correlation patterns of the different variables are different were already exposed in Sec. 5.2 & 5.3 but in a way too elusive way. This comment shows that the physical interpretation is very important to understand the paper and its motivations, and its absence might have been somewhat frustrating. In short (see track changes and Fig. 4): Band 4 is sensitive to SSA and LAP. LAP forcings are spatially uniform, partly explaining the rather constant and high spatial correlation of Band4. The spatial homogeneity of meteorological forcings also explains the strong HS correlations. Band 5 is sensitive to changes in surface micro-structural properties. Differential metamorphism can sometimes occur (between southern and northern aspect) causing a de-correlation in band 5, potentially explaining what is observed on Fig. 4b. Negative correlations can also happen for the same reason between e.g. two elevations separated by the rain-snow line.

See the track-change throughout 5.2&5.3

Finally, investigating the skill of the PF as a function of the selected spectral bands is beyond the scope of this paper but note that this important topic is investigated by Revuelto et al., (submitted to Journal of Hydrology). This reference was clearly missing (because this reference was only in preparation when this manuscript was submitted).
We now refer to Revuelto et al., (submitted) in the last paragraph of Sec. 5.2.

9. Line 279: Figure 3c
Corrected
10. Line 367: remove one "because of".

11. Figure 1: panel a is not "flat", as it has an elevation gradient.

We addressed this comment in the response to the referee's comment 3.

Making c the same size with b so their slope difference is more clear.

 We understand that the different horizontal extent between (b) and (c) might be confusing but in this schematic representation, it is important that (a), (b) and (c) reach the same elevation. (c) appears smaller than (b) because it is steeper, but indeed they reach the same altitude. If (c) had the same basal area as (b) as suggested by the reviewer, it would have a similar size, but it would be twice as high, and unfortunately we believe that this would be detrimental to the description of the geometry.

Reviewer 2 Kristoffer Aalstad

General comments

This manuscript presents a new ensemble-based snow data assimilation framework, Crocus-Observations (CrocO), to assimilate observations into the Crocus snowpack model in a semi-distributed geometry with a particle filter (PF). To address the issue of degeneracy, different variants of the PF are tested in a series of synthetic experiments where spatially sparse observations of height of snow (HS) or reflectance are assimilated for a massif (group of mountains) that is discretized into topographic classes. The sparsity of observations is meant to mimic the real situation where in-situ HS observations are usually only available for a handful of locations in a massif while clouds, shadows, and canopies can cause spatial gaps in (useful) reflectance retrievals. The objective is to use the PF to propagate information in space; i.e. to constrain the model ensemble not just in the observed classes, but also in the unobserved classes. The issue, compared to a completely local approach (called rlocal), is that this requires the assimilation of a larger number of observations which may trigger degeneracy. Through a series of 16 synthetic scenarios the authors demonstrate that it is possible to achieve such a propagation of information without degeneracy, both in the case of HS and reflectance assimilation, using either a global PF with inflation (called global) or a PF that is localized based on background correlations (called klocal). This work fits well within the scope of GMD, and it is certainly of interest to the growing snow data assimilation community where the PF is gaining popularity. To my knowledge, it is also the first snow data assimilation study to demonstrate how the PF could be used in a spatialized context (non-local analyses) while avoiding degeneracy. The technical level of the work is also high with the framework being built up to eventually be run for operational purposes in an HPC environment. I therefore recommend this paper for publication pending minor revisions with a few technical concerns as outlined below.

The authors would like to thank Kristoffer Aalstad for this exhaustive review. We believe that comments helped to improve the clarity of several essential points (e.g. the formulation of the PF implementation, the statement of the degeneracy problem, and the non intuitive behavior of the PF in case of negative correlations). Regarding the motivations, and methodology, there was a lack of justification for the choice of the PF over the EnKF and obviously, the question of the SCF, an essential variable, was overlooked. Theoretical limitations of our work were shed to light, an issue which had to be acknowledged, even though we believe that we agree on the fact that it might not be severely detrimental to the applicability of our method. Finally, there were significant theoretical inputs on the bases of the PF and on potential avenues. These contributions were beneficial to the authors much beyond what will appear in the manuscript. The authors wanted to express their gratitude for this as well.

Specific comments

L1 Consider changing "the snowpack" to just "snowpack" since not all snowpack properties are crucial.
Corrected

L2 Change "on the snowpack" to "on the state of the snowpack".
Corrected

L4 Change "inform on" to "provide information about".
Corrected

L5 Change "enables to estimate" to "enables the estimation of". It is not clear who or what is "enabled to".
Corrected

L7 Consider changing "non observed" to "unobserved".
Thanks for the suggestion, changed throughout the text.

L10 Change "known" to "prone" and "a too large number of" to "too many".
Corrected

L34 It could be worth mentioning that higher resolution optical satellites (e.g. Landsat, Sentinel-2) are better able to resolve fractional snow cover at the MODIS scale (e.g. Aalstad et al., 2020, and references therein).

Thanks for this suggestion, this statement actually makes sense. Even though SCF is not the main focus in this work, we for sure consider assimilating it in future work, and it is worth mentioning it. In order not to loose track on our objectives (i.e. assimilating reflectances), we propose the following formulation, which acknowledges that SCF has more or less the same spatio-temporal limitations of reflectance, and mention that it saturates for deep snowpack.

The higher resolution offered by products from Landsat or Sentinel-2 might be an avenue to this issue (e.g. Masson et al., 2018; Aalstad et al., 2020) but at these resolution, reflectance retrievals are quite noisy due to e.g. digital elevation model errors (Cluzet et al., 2020). Finally, note that pixel fractional snow cover (snow cover fraction, SCF) can be accurately retrieved even from noisy reflectances (Sirguey et al., 2009; Aalstad et al., 2020), but it inherits the same spatio-temporal limitations as reflectances. SCF informativeness might also be limited in deep snowpack conditions (De Lannoy et al., 2012).

L38 Change "enable to" to (e.g.) "enable us to".

The sentence was changed to:

Detailed snowpack models are the only ones able to assess avalanche hazard and monitor water resources alike (Morin et al, 2020), but these applications are limited by their considerable errors and uncertainties (Essery et al., 2013; Lafaysse et al., 2017).

L41 Change "enables to" to "lets us".

The sentence was changed to:

Indeed, data assimilation combines the spatial and temporal coverage of snowpack models with the available information from observations in an optimal way.

L51-53 To be more precise I would suggest stating more explicitly that the two steps in the SIR PF analysis are importance sampling of the (unnormalized) posterior, with the prior as the proposal (or importance) density, followed by resampling to reduce the variance in the weights. In that way, it is also easier to understand the origin of the name "SIR". van Leeuwen (2009), who is already cited, explains these steps clearly for curious readers.

Thanks for this nice suggestion for improving this paragraph, explaining this two-step is a plus. However, as the referee understood, we are not familiar with Bayesian terms such as "importance sampling" and "proposal" and we are wondering whether using such terms would confuse readers without a background in Bayesian theory. We propose an alternative formulation, keeping the spirit of the two-steps and the term of "importance sampling", and helps understanding the "Sequential Importance Resampling" formulation.

The analysis of the PF-SIR (later on "PF") works in two steps. In a first step, so-called "importance sampling", the particles are weighted according to their distance to the observations (relative to the observation errors). Then, a resampling of the particles is performed in order to reduce the variance in the weights.

L55 When you say "i.e.. . . " I expected a brief definition or explanation of what degeneracy is. Instead you state a consequence (or remedy) to degeneracy. It may be better to define degeneracy (as you do later on L163), after which you can mention solutions.

Moreover, degeneracy is only mentioned in the context of assimilating a large number of observations; which is seemingly what you try and deal with in this study. This problem can arise even in low dimensional states and is often a result of the likelihood (and thus posterior) becoming more peaked and harder to resolve with the available particles. An arguably broader issue that causes degeneracy with the PF (and importance sampling in general) is the curse of dimensionality where the required ensemble size (to avoid degeneracy) scales exponentially with the dimension of the state. This is also discussed in the studies of Snyder et al. (2008); Bengtsson et al. (2008) that are already cited. I would suggest introducing the curse of dimensionality explicitly, since it can help explain why one expects that using a global (rather than local) PF algorithms, where the state space becomes much larger, is quite difficult.

This is a very interesting input, as it sheds light on the reasons why we expect the localised approach to be more suited to large scale problems, it was accounted for.

It is also surprising that the EnKF is barely mentioned, one of its strengths and the reason it is widely used in many applications is that it is more robust to this curse.

We agree that further discussion was needed on the PF vs. EnKF. Indeed, the main reason why we cannot consider using the EnKF is the Lagrangian formulation of our model which makes the computation of ensemble mean and updates impractical . This is thoroughly explained in Charrois et al. (2016) (already cited in the paper). On the contrary, as you say it is important to state that while applying the EnKF in spatialised application is quite easy, degeneracy/ curse of dimensionality are a severe drawbacks for the PF.

The concerned paragraph and the previous one were therefore modified:

The Particle Filter with sequential importance resampling (PF-SIR, Gordon 1993; van Leeuwen 2009 is a Bayesian ensemble data assimilation technique well suited to snowpack modeling (Dechant and Moradkhani, 2011; Charrois et al., 2016; Magnusson et al., 2017; Piazzi et al., 2018; Larue et al., 2018). The PF-SIR is a sequential algorithm relying on an ensemble of model runs (particles) which represents the forecast uncertainty. At each observation date, the prior (or background) composed of the particles is evaluated against the observations. The analysis of the PF-SIR (later on "PF") works in two steps. In a first step, so-called "importance sampling", the particles are weighted according to their distance to the observations (relative to the observation errors). Then, a resampling of the particles is performed in order to reduce the variance in the weights. The Ensemble Kalman Filter (EnKF, Evensen 2003), has also been widely used for snow cover data assimilation (e.g. Slater et al., 2006; De Lannoy et al., 2012; Magnusson et al., 2014). However, the PF is more adapted to models with a variable number of numerical layers such as detailed snowpack models (Charrois et al., 2016).

The PF could be used in a spatialised context to propagate the information from observation as suggested by Largeron et al., (2020) and Winstral et al., (2019). Contrary to the EnKF, such applications are rare to date (e.g. Thirel et al., 2013; Baba et al., 2018; Cantet et al., 2019). Indeed, spatialised data assimilation with the PF is not straightforward because of the degeneracy issue, i.e. only a few particles are replicated in the analysis, often resulting in a poor representation of the forecast uncertainties. Degeneracy can be mitigated by increasing the number of particles, but the required population scales exponentially with the number of observations simultaneously assimilated (Snyder et al., 2008). Furthermore, an accurate representation of spatial error statistics by the ensemble is essential for the success of the assimilation system. To achieve that, the required ensemble size also scales exponentially with the system dimension, an issue known as the curse of dimensionality (Bengtsson, 2008). These issues are severe drawbacks when considering applications of the PF on large domains

(i.e. implying a large number of observations and/or simulation points) with a reasonable number of particles (Stigter et al., 2017).

L60 While it is probably true that observation error variances are often underestimated, it is (in terms of Bayes' rule) strictly speaking incoherent to keep inflating these variances outside of certain frameworks such as likelihood tempering (see van Leeuwen et al., 2019, and references therein). Tempering of the likelihood explains the coherency of the ensemble smoother with multiple data assimilation (ES-MDA), used in Aalstad et al. (2018) for snow DA, which also inflates the observation error covariance matrix. It is not necessarily a big problem that the use of inflation here is incoherent, but the fact that it is a heuristic approach should be mentioned explicitly and potential solutions such as tempering could be proposed.

Thank you for the very interesting input on tempering methods. We didn't realize that Aalstad et al., (2018) was performing inflation, which we interpret as conceptually closer to the tempering presented in van Leeuwen et al., (2019) than to our approach. Our understanding of this literature and of the present comments is that tempering mitigates sampling issues but does not alter the extraction of information from observations since tempering/inflation coefficients sum to one. In contrary, our method does, and is therefore theoretically sub-optimal if not inconsistent. As you say, the inflation method we propose, as introduced by Larue et al., (2018) is a heuristic method aiming at mitigating mis-specified observation and representativeness errors. We acknowledge that this fact is worth underlining here (see changes). Meanwhile, we understand that tempering might be suited to tackle badly specified observation errors, but not in its present form. This is for sure an interesting lead to investigate. The following change is proposed for lines 58-62:

Several solutions exist to tackle the PF degeneracy. A first approach is to inflate the observation errors in the PF. The tolerance of the PF is increased, leading to more particles being replicated. This heuristic approach is based on the fact that observation error statistics (including sensor, retrieval and representativeness errors) are usually poorly known and underestimated. It can also be used as a safeguard to prevent the PF to degenerate on specific dates, when observations are not compatible with the ensemble. PF inflation was successfully implemented in point scale simulations of the snowpack (Larue et al., 2018).

see also change line 168 of the manuscript:

A first approach to mitigate degeneracy is to use inflation. This heuristic method iteratively...

L69 Change "It makes" to "This makes".
Corrected
L79 Change "operationally used" to "used operationally".
Corrected
L81 Change "enables to" to "enables us to".
Corrected
L91 Change "reflectance" to "reflectance observations".
Corrected
L95 Change "Following" to "Subsequently".
Corrected
L101 Change "the model into" to "the model for".
Corrected
L102 Change "enables to" to "enables us to".
Corrected
L112 Change "enabling to represent the snowpack coupling" to "coupling the snowpack with".
Corrected
L117 Change "This way," to "As such,".
Corrected

L140 In general I would suggest to put the hat just above the variable and not the sub/superscript. Similarly, I don't think sub/superscripts should be in bold since b ib rather than they are not matrices or vectors. That means (for example) using x*** x b and X b rather than X b . This is a recurring issue throughout the math in the text. To conform with usual DA notation it might be better to not use a hat for the state (i.e. just x) and instead use a hat for the predicted observations (x or better yet y)

Thanks for this rigorous input which has been accounted for in the revised manuscript.

L146 Remove "supposed" since you state the independence assumption in the ensuing brackets.

Corrected

L148 Change "type of variable of observation" to "type of observation".

Corrected

L165 I didn't see N ef f defined or even mentioned in Doucet et al. (2001), but maybe I missed it.

Thanks for pointing this citation error. Correct reference is: Doucet, A.: On sequential simulation-based methods for Bayesian filtering, Tech. Rep., 1998, but it is not peer-reviewed, so we opted for Liu and Chen (1995)

L168 Change "sample population" (a mix of distinct terms) to "effective sample size".

Corrected

L175 Change "inspired on" to "inspired by".

Corrected

L180 Change "observations simultaneously assimilated" to "observations that are simultaneously assimilated".

Corrected

L195&L201 I don't really follow the procedure here. First you say reflectance is not defined when there is no snow, then you say it is set to 0.2 for snow-free ground. Which is it? Are the bare ground reflectance values set as undefined or actually considered? I would expect the residuals to also contribute important information in the assimilation also in the cases that an observation or particle is bare as opposed to snow-covered.

Thanks for this interesting remark which underlines the strong link between reflectance and SCF assimilation. We completely agree on the fact that snow/no snow holds precious information. Our choice was to not comment this question too much in order to focus on reflectances, but we agree it deserves clarifications. L.195 explains that TARTES optical scheme only provides snow reflectance (i.e. not a surface reflectance of a mixed soil-snow surface): this variable is not defined in the absence of snow. Some members being "undefined" is problematic for the PF. Conversely, in the observations, "no-snow" is an information, contrary to "no observation". For this reason, in L201-202 we force a default value in the computation of the weights. By putting a reflectance of 0.2, (which corresponds to the bare soil broadband albedo in ISBA) in the unmasked snow-free synthetic observations and snow-free members, we extract this binary information in a very rough way. Ideally, future work should jointly assimilate reflectance and SCF in order to better leverage this information.

According to this explanation, Sec. 2.3.3 was expanded (the fact that reflectance observations are bounded was dropped):

Assimilating reflectance with the PF requires some adaptations. In Crocus, TARTES optical scheme (see Sec. 2.2.1) only provides snow reflectance, not all-surface reflectance: no value for the surface reflectance is issued in the absence of snow. Conversely, the weights of the particles are not defined in Eq. 2 if the members are snow-free. These issues were roughly accommodated by setting the reflectances of snow-free members and observations to 0.2 (the value of bare soil broadband albedo in ISBA model) in the PF Eq. 2 (Sec. 2.2.3).

L196 Why are negative background correlations considered "significant"? If the prior ensemble is negatively correlated between the analysis point and the observed point then surely the residuals

(innovations) in the observed point should not necessarily be expected to carry over to the hypothetical residual at the analysis point? Is the reasoning that the hypothetical residual at the analysis point is in the perfectly negatively correlated case equal to minus the innovation at the observed point and that only the square of the innovation matters with a diagonal R?

Thanks for pointing the question of negative correlations. We are not sure to fully understand the question, so we try to answer but we might have missed something. Negative correlations can be physically sound. Consider HS and two points separated by the rain-snow line during a rain-on-snow event, an ensemble built by perturbations on the precipitation rates, and an observation available after the precipitation event. In the snowy (rainy) point, the members with the highest solid (liquid) precipitation will see their HS increase (decrease), resulting in a negative ensemble correlation between the HS of the two points. Now consider that only the HS of the snowy point is observed, and that the ensemble underestimated HS: it is likely that precipitation rates were underestimated at both locations: HS is likely overestimated in the rainy point. The PF will select the members with the highest precipitation rates at the snowy point, but this information is also valid for the rainy point, and therefore this information should be transferred by using the same PF sample there.

The correspond item was therefore modified:

\item $\lvert \mathbf{B}_{\bm{v}}(n,p) \rvert >0.3$. If the absolute correlation is low, it is likely that there is a poor potential for the distant observation to constrain the ensemble locally. In such a situation, it is better to reject the observation from the local analysis. Negative ensemble correlations can be physically sound, e.g. after a rain-on-snow event between the HS of two points separated by the rain-snow line. In such a situation, an HS observation on either point can hold information on precipitation rates at both locations. At the observed location, the PF will select the members with the most appropriate precipitation rates. This sample is likely to perform well at both locations, so it can be used to constrain the unobserved location.

Also, perhaps use another term then "significant" which unfortunately still has strong statistical (null hypothesis significance testing) connotations.

Thanks for this remark, this was modified throughout the text.

L205 Change "openloop" to "open-loop".

Corrected (multiple changes).

L206 The sentence "These observations allow to mimic real observations with a perfect knowledge of the true state" can easily be misunderstood to mean that real observations capture the true state. If anything, perfect observations are quite unrealistic and do not perfectly mimic reality at all. The fact that observations are not perfect is central to the Bayesian origins of ensemble-based DA in general and particle filtering in particular. With perfect observations DA just becomes an optimization problem. Ironically, you would end up with a sure-thing hypothesis (Jaynes, 2003; Schöniger et al., 2015), your likelihood would be a Dirac-delta function, and your particle weights would be nonsensical. In practice you do use a non-zero σ k 2 in the analysis so this doesn't happen, but it is inconsistent to not perturb your synthetic observations.

Thanks for his thorough remark. Despite this is mentioned on L209, we acknowledge that the fact that we don't corrupt the observations should be pointed out more clearly as a limit of our methodological study compared to the literature (e.g. Durand et al., 2006) despite recent studies did not do so either (e.g. Charrois et al., 2016). We are actually conscious of this limitation, and a recent study has been submitted (Revuelto et al., submitted) in which we assimilate synthetic corrupted observations at the point scale. In our situation we did not corrupt the observations because little is known about the spatial structure of errors of reflectance (e.g. Cluzet et al., 2020): we know that assuming independent errors (i.e. diagonal R) is a very rough approximation of the reality which has strong consequences on the propagation of information. Corrupting the observations with such random structures would be theoretically more consistent, but would not yield much more insight on the potential for information

from real observations to be spatially propagated as real spatial correlation of observation errors might be very different from this hypothesis. Future efforts should concentrate in better characterizing these spatial structures of errors. Consistently, the following sentence was modified:

Synthetic observations are extracted from a model run and assimilated without adding any noise. These observations mimic...

and a paragraph was added in the end of Sec. 5.2:

Regarding the observations, our study has some methodological limits, however. Observation errors are very roughly prescribed, and the assimilated observations are not corrupted as usually done in synthetic experiments (e.g. Durand et al., 2006). These choices were motivated by the fact that very little is known about the spatial correlation of reflectance observation errors in the semi-distributed setting (e.g. Cluzet et al., 2020). In a recently submitted paper, the impact of random and systematic errors of reflectance observations on point-scale assimilation experiments is thoroughly investigated (Revuelto et al., 2021). Efforts to better characterize these observation errors should be conducted in future work

L207 Change "It allows" to "This allows us to". Linked to previous comment.
Corrected

L223 I guess by integral you really mean average? It is hard to imagine what the integral of SWE over time would represent physically unless it is normalized by the time period you are integrating over.
Indeed we computed the time integral, for the sake of computational simplicity, not the average. There is only a proportionality factor between the integral and the average, so SWE percentiles correspond to average SWE percentiles. We propose to simplify the statement by replacing the "integral" by "average", which makes it more sound, and does not change anything to the idea.
Changed "integral" to "average"

L224 On a first reading it was not clear why you extract percentiles of the open-loop ensemble to be used as synthetic observations. Perhaps you could make it clearer that you are effectively independently considering several different synthetic truth scenarios rather than a single truth run?
Thanks for pointing out this lack of clarity. The following sentence was added on L225:
...e.g. 2014_p80). This method enables us to evaluate the efficiency of data assimilation experiments under contrasted snow condition scenarios. Before any assimilation experiment...

Also, after you have extracted these different synthetic truth runs, what is in the way of perturbing the observed variables in these (for each scenario) to generate synthetic observations as is usually done in twin experiments? This would allow for a more realistic evaluation, since real observations are noisy and you would still have access to the synthetic true SWE (unobserved) that you use in your evaluation?
See previous answer to comment from L206.
L230 Change "date" to "dates".
Corrected
L233 Change "is set" to "are set".
Corrected
L235 Change "uses only" to "only uses".
Corrected
L247-265 When you compute your evaluation metrics you are using the corresponding truth not the corresponding (non-existent) observations. Your entire evaluation is based on how CrocO performs in terms of estimating the (unobserved) true SWE. As such, I suggest changing $o_{c,t}$ to $T_{c,t}$ (T for truth, or something similar) and similarly for $O_{c,t}$ to make this clearer. Alternatively, you could be more explicit that all your evaluation is SWE-based and instead use notation like $SWE_{m,c,t}$ for the SWE ensemble and $SWE_{?c,t}$ for the true SWE in a given scenario?

Thanks for this nice suggestion. We opted for the first option, substituting $o_{c,t}$ by $\tau_{c,t}$. (see changes).

L250 I suggest calling this the absolute error of the (ensemble) mean (AEM), to avoid confusion with the (ensemble) mean absolute error (MAE).

Thanks for this comment. We opted for the AEM name, which is unambiguous. Modifications were performed accordingly (including Fig. 3.)

For the caption of Figure 3, and when discussing this Figure (around L281) you call "AE" the RMSE which is incorrect. Judging by Fig.3a the RMSE would be considerably larger for the open-loop than for any of the analyses.

Thanks for this comment. We actually forgot to replace RMSE by AE (AEM) in the text, thanks for pointing this out !

L264 This could be understood to mean that this is Eq .8 in Hersbach (2000),which it is not, and it is unusual to enumerate an equation (your Eq. 8) before it is presented on the next line. Furthermore, I couldn't find such an equation in Hersbach (2000), the closest I could find was his Eq. 39 which had an extra uncertainty term and a sign reversal for the "Resol" term. Could you explain the discrepancy?

Thanks for pointing this out. There was an error in the reference, the appropriate one being Candille et al., 2015. While the interpretation of the Reli term is unambiguous, interpretation of the Resol term is more controversial (P. de Mey and G. Candille, personal communication). This is why we didn't focus on the Resol term. Recent publication from Leutbecher et al., (2020) might help understanding Resol for curious readers.

L277 Change "well representative" to just "representative".

Corrected

L292 Change "contrasted" to "contrasting".

Corrected

L294 Change "as for HS" to just "for HS"?

, we actually mean that for band4, spatial correlation patterns are similar to those of HS.

We replaced "as for HS" by "Similar results are obtained for HS."

L296 Again consider using another word than significant.

Corrected

Furthermore, are high background correlations that surprising given that, for a given ensemble member, you use the same multiphysics (M i) and forcing perturbations (F i) across the entire (semidistributed) domain? Isn't this mainly an indication that the SAFRAN forcing is quite spatially homogeneous (L128)?

Thanks for this remark. Indeed, intrinsic correlations come from the forcing variables and ESCROC members and, this point is actually discussed in Sec 5.3 of the manuscript. Nevertheless Fig.4a-b actually shows that despite same Mi and Fi are applied across the entire domain, some locations are completely decorrelated due to the combination of strong vertical gradients and some highly non-linear processes.

We amended L412 of the manuscript to also mention that ESCROC members were also spatially constant:

Strong Band 4 correlations might be due to the spatially homogeneous perturbations of LAP fluxes used to force the simulations (see Sec. 2.2.2), a key driver of this variable, and because the same snow model configuration is applied for a given member across the simulation domain. Several studies suggest...

Moreover, this comment points out a lack of interpretation of these background correlations in our manuscript. In line with our answer to comment 8 of Reviewer 1, physical interpretation of the Band 5 background correlations evidenced in Fig. 4b is a bit more complex than for Band 4 and HS.

Differential metamorphism can happen between the opposite sides of a mountain (because of different illumination conditions), or similarly, across the rain snow line, resulting in a de-correlation of band 5 reflectances. Details on these processes were added throughout sections 5.2 and 5.3, and we suggest to refer to the track-change.
See track change in Secs. 5.2 and 5.3

L301 Change "launched" to "conducted". In general, I would suggest referring to the SWE percentile-based sets of observations as "synthetic observation scenarios" rather than "synthetic members" to avoid confusion with the ensemble members.
Corrected, thanks for this suggestion.
L319 There are many examples in the literature of fractional snow-covered area (fSCA), which is retrieved from reflectance, constraining bulk variables like SWE quite well.
We agree and following your suggestion, several significant changes been made throughout the text (introduction discussion and conclusions, see in particular the answer to L.34, L. 391, L470. We hope these corrections are enough. Here, for the sake of clarity, we propose to correct to:
"raw reflectance products"

HS observations are also often not representative of the model scale.
This is an important point on which we completely agree, but we do not aim at discussing HS assimilation too much as the main focus of this study is reflectance.

L320 Change "all other things equal" to "all other things being equal". Perhaps make it clearer that you are not jointly assimilating HS and reflectance in this experiment.
Thanks for this suggestion. The sentence was changed to:
In order to assess this difference, we conduct assimilation of reflectance only, in the same setup as in Sec. 4.2.1}, all other things being equal.

L324 Change "well represent" to "properly represent". Also on the next line use (e.g.) "marked" instead of "significant".
Corrected
L327 Change "with respect to" to "compared to".
Corrected
L330 Why is "Skill" capitalized?
Corrected
L350 I would recommend switching "a right probability" to "the right frequency". Paraphrasing the discussion from the bottom of page 564 in Hersbach (2000): for the (average) CRPS, the reliability is similar to the rank histogram which can show if the frequency that the truth has a certain rank in the ensemble is equal for all ranks. In applications Bayesian (rather than frequentist) inference, which is what the PF is used for, there is an important distinction between the concept of frequency and probability; the latter is a measure of uncertainty (degree of belief, plausibility) (e.g. Lindley, 2000; Jaynes, 2003).
Corrected, accounted for, thanks a lot for this input.

L356 Change "conceptual" to "synthetic".
Corrected
L360 Change "on the" to "for the".
Corrected

L362 This is an interesting speculation, but these are ensemble correlations between two areas in your domain not real spatial correlations. Maybe the ensemble is similar in the eastern and western aspects of the domain because a rain shadow effect (or something else) is not captured in your open-loop.

The potential mismatch between ensemble correlations and real spatial correlations is discussed in L411 of the manuscript. We agree that as you mention, not accounting for the intra-massif variability of meteorological conditions (in the sense that you mean, e.g. Western slopes should lie preferentially in the windward side of the massif and receive more/less snow than those on the windward side). On L. 411, we added a mention to this:

Strong and almost uniform HS correlations (Fig. 4b) might be caused by the satial homogeneity of precipitation perturnations and because we do not account for e.g. wind drift, intra-massif variability of meteorological conditions and gravitational redistribution of snow (Wayand et al., 2018)....

L364 Change "such elevation" to "such elevations".

Corrected

L370 I would argue that the fSCA depletion is quite informative for any seasonal snowpack, it is not necessarily maximally informative for intermittent snowpacks below the rain-snow line.

Thanks again for this remark. We believe that answer to L319 comment is appropriate here too.

L372 Change "well linked" to "closely linked".

Corrected

L375 Change "outstanding" to "unexpected" and (next line) "between these" to "for these two".

Corrected, the sentence was changed to:

This study unexpectedly suggests that reflectance information can be spread from southern slopes to the northern ones,...

L379 Change the sentence "It is informative. . . " to "In our ensemble data assimilation framework, however, it does seem to be informative.". On the next line I also recommend removing "in this case".

Corrected

L382 Change "enabling to correct" to "enabling a correction of".

Corrected, changed to: … parametrisations, thus correcting the ensemble...

L387 Sentinel-2 and the Landsats should not be put in the same moderate resolution category as MODIS, VIIRS, and Sentinel-3.

Corrected, changed to: …the abundance of moderate to high resolution space-borne sensors (MODIS, Sentinel2-3, VIIRS, Landsat…)…

L391 Change "usually" to "often" to qualify this statement.

Corrected, changed to : "generally" a bit stronger than "often".

L394-395 In terms of the current status of remote sensing of snow using optical satellites, this sentence seems too pessimistic. Even though Warren (2013) states that retrieving BC content of snow from satellites is unlikely to be successful, it does not follow that reflectances retrievals from optical satellites are currently too inaccurate to be used to provide accurate information on snowpack properties.For example Aalstad et al. (2020) (and many other references therein) show that fractional snow-covered area (fSCA) can be estimated quite accurately from reflectances through a variety of methods using optical satellite sensors that are currently in orbit. These fSCA retrievals can, in turn, be used to constrain modeled estimates of other snowpack properties such as SWE through particle-based DA methods (see e.g. Alonso-González et al., 2020, for a recent example).

Thanks for pointing out this sentence whose formulation was inappropriate. Our purpose was to talk about surface properties (grain size, and light absorbing particle contents (LAP)). Thanks to a previous comment (L34), it is now acknowledged that SCF is accurately retrieved. Regarding surface properties, as you say, Warren (2013) statement only stands for LAP, while for snow microphysical properties, the

required accuracy might be reached. And of course, for SCF, it is already the case, this point was added in the introduction (see changes to comment on line 34).

In the near-infrared, the signal-to-noise ratio of reflectances observations might be sufficient to constrain the surface microphysical properties (Durand et al., 2007; Mary et al., 2013), whereas the required accuracy for visible reflectance retrievals to remain informative on the snowpack light absorbing particles content is high (Warren, 2013), and it is yet to prove whether either approach can achieve this requirement.

Regarding the interest of SCF for data assimilation, we agree also on its added value, and that it needed to be acknowledged in the discussion. A sentence was added at the beginning paragraph of Sec. 5.3 (L. 388)

Reflectance is an appealing variable for snowpack modelling because of its sensitivity to snowpack surface properties (Dozier, 2009) and the abundance of moderate to high resolution space-borne sensors (MODIS, Sentinel2-3, VIIRS, Landsat...) providing us with a handful of observations to assimilate, contrary to HS. The potential for assimilation of SCF, which is retrieved from reflectances, is clear (Margulis et al., 2016, Aalstad et al., 2018, Alonso-Gonzalez et al., 2020). This study demonstrates the potential of the PF to spread information and assimilate raw reflectances with a positive impact (Sec. 5.2). Yet, assimilating real observations of reflectance is another challenge, for two reasons.

L408 Change "informations" to "information".
Corrected
L410 How can a correlation pattern based on an ensemble be realistic? In Bayesian inference the ensemble represents a probability distribution: a measure of uncertainty which is in the mind, not real. Jaynes (2003) explains this well with what he calls a mind projection fallacy: confusing reality and states of knowledge about reality.
We acknowledge this is a bad formulation, thank you for pointing this out. We mean: based on the assumption that ensemble background correlations are a realistic representation of modeling errors. The sentence was changed to:
The klocal algorithm could be more suited to this situation, because it is looking for local optima, based on the assumption that background correlation are a realistic representation of modelling errors.

L414 Change "reliable model for that" to "reliable LAP model".
Corrected, changed to : ...no reliable model of such processes exists in complex terrain.

L418 Change "suffers from obvious" to "suffering from obvious" and "suffer for large" to "suffer from large".
Corrected
L421 As before, in Bayesian probability theory how can an ensemble correlation be real?
See answer to L410: changed to:
In the future, improving the ability of ensemble correlations to represent modelling errors correlations could make the spreading of information an even more challenging task with the klocal algorithm.

L424 Change "area" to "areas".
Corrected
L426 Change "into larger" to "for larger".
Corrected
L427 Change "take the best" to "outperform".
Corrected
L451 Change "in the way of a new" to just "in a new".

Corrected

L456 Change "spatialized" to "semi-distributed".

We acknowledge that this work has only been done in a semi-distributed geometry, which is a spatialized setting. We would like to stick with the use of "spatialized" because the term "semi-distributed" is quite obscure for the majority of the people, and it might confuse the audience especially if they only read the abstract/conclusion. We consider that it is clearly stated everywhere else in the paper that we work in a semi-distributed setting and that this will be clear for curious readers. Finally, as we mentioned later on, nothing specific to the semi-distrubited geometry was developed here: CrocO can be applied seamlessly on networks of in-situ stations and fully distributed frameworks.

Also mention somewhere in the conclusion that this is a synthetic experiment.

We completely agree that this should be mentioned. The sentence on L490 was changed to:

In the framework of synthetic experiments, we have shown in particular that:

L460 Capitalize the leading words in this enumeration.

Corrected

L469 Change "errors" to error".

Corrected

L470 Again, why would fSCA only be worth assimilating at lower elevations? The depletion of fSCA might provide useful information anywhere in your domain. For example, Margulis et al. (2016) assimilated fSCA with a particle batch smoother (equivalent to your rlocal PF without resampling) to produce a 30 year high resolution snow reanalysis for the Californian Sierra Nevada with unprecedented accuracy. This study and others like it surely indicate that fSCA is quite valuable also for a PF even at higher elevations.

We agree that this statement was too pessimistic regarding the SCF. The sentence was changed to:

Snow cover fraction would be a good companion variable to jointly assimilate with reflectances, requiring the use of an appropriate observation operator.

L490 Change "softwares" to "software".

Corrected

L500 Change "enabling to" to "enabling us to".

Corrected, changed to: necessary to

Fig. 1 caption: Change "elevation bands altitudes" to "altitudes of the elevation bands". Also change "40 ◦ degrees slopes" to "40 ◦ slopes" since the ◦ symbol is shorthand for degrees.

Corrected

Fig. 2: Why is the superscript of the fourth prior particle at t 1 3 and not 4? As suggested earlier for L140, consider changing the use of hats in your math notation.

Corrected

Table 1: Change N eeff to N ef f . In the caption, change "setup of" to "Setup for" and change "snow depth" to "height of snow" to be consistent with the rest of themanuscript. The same applies to the title of subsection 4.2.1.

Corrected

Table 2: Change N eeff to N eff . In the caption, change "setup of" to "Setup for". Furthermore, change "second" to "first"; this is the first reflectance experiment.

Corrected

Table 3: Same problems as with the other Tables.

Corrected

Fig. 3: In the caption, change RMSE to AE (or AEM).

Corrected, changed to AEM

Fig. 4: In the caption, change "the denote" to "denote the".

Corrected

Fig. 4: In the caption, change "on the whole" to "for the whole".
Corrected
Fig. 6: In the caption, consider changing "synthetic members" to "synthetic scenarios" (since these are not ensemble members). Also, why is "Skill" capitalized?
Corrected accounted for, according to previous corrections.
Fig. 8&9: In the caption, consider changing "member" to "scenario" to avoid confusing the concept of your truth scenarios and the ensemble.
Corrected

References

Aalstad et al.: Ensemble-based assimilation of fractional snow-covered area satellite retrievals to estimate the snow distribution at Arctic sites, TC, https://doi.org/10.5194/tc-12-247-2018, 2018.

Aalstad et al.: Evaluating satellite retrieved fractional snow-covered area at a high-Arctic site using terrestrial photography, RSE, https://doi.org/10.1016/j.rse.2019.111618, 2020.

Alonso-González et al.: Snowpack dynamics in the Lebanese mountains from quasi-dynamically downscaled ERA5 reanalysis updated by assimilating remotely-sensed fractional snow-covered area, HESSD, https://doi.org/10.5194/hess-2020-335, preprint under review, 2020.

Bengtsson et al.: Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems, in: Probability and Statistics: Essays in Honor of David A. Freedman, https://doi.org/10.1214/193940307000000518, 2008.

Candille, G., Brankart, J.-M., and Brasseur, P.: Assessment of an ensemble system that assimilates Jason-1/Envisat altimeter data in a probabilistic model of the North Atlantic ocean circulation., Ocean Science, 11, 425–438, 2015.

Charrois, L., Cosme, E., Dumont, M., Lafaysse, M., Morin, S., Libois, Q., and Picard, G.:
On the assimilation of optical reflectances and snow depth observations into a detailed
snowpack model, The Cryosphere, 10, 1021–1038, 10.5194/tc-10-1021-2016], 2016.

Cluzet, B., Revuelto, J., Lafaysse, M., Tuzet, F., Cosme, E., Picard, G., Arnaud, L., and Dumont, M.: Towards the assimilation of satellite reflectance into semi-distributed ensemble snowpack simulations, Cold Regions Science and Technology, 170, 102 918, 2020

De Lannoy, G. J., Reichle, R. H., Arsenault, K. R., Houser, P. R., Kumar, S., Verhoest, N. E., and Pauwels, V. R.: Multiscale assimilation of Advanced Microwave Scanning Radiometer–EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado, Water Resources Research, 48, 2012.

Doucet et al.: An introduction to sequential Monte Carlo methods, in: Sequential Monte Carlo methods in practice, https://doi.org/10.1007/978-1-4757-3437-9_1, 2001.

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

Eberhard, L. A., Sirguey, P., Miller, A., Marty, M., Schindler, K., Stoffel, A., and Bühler, Y.: Intercomparison of photogrammetric platforms for spatially continuous snow depth mapping, The Cryosphere Discussions, pp. 1–40, 2020.

Hersbach: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems , WF, https://doi.org/10.1175/1520-0434(2000)015\%3C0559:DOTCRP\%3E2.0.CO;2, 2000.

Jaynes: Probability theory: The logic of science, https://doi.org/10.1017/CBO9780511790423, 2003.

Larue, F., Royer, A., Sève, D. D., Roy, A., and Cosme, E.: Assimilation of passive microwave AMSR-2 satellite observations in a snowpack evolution model over northeastern Canada, Hydrology and Earth System Sciences, 22, 5711–5734, 2018.

Leutbecher, M. and Haiden, T., Understanding changes of the continuous ranked probability score using a homogeneous gaussian approximation, QJRMS, 2020, 1-18

Lindley: The philosophy of statistics, The Statistician, https://doi.org/10.1111/1467-9884.00238, 2000.

Liu, J. S. and Chen, R.: Blind deconvolution via sequential imputations, Journal of the american statistical association, 90, 567–576, 1995.

Margulis et al.: A Landsat-Era Sierra Nevada Snow Reanalysis (1985–2015), JHM, https://doi.org/10.1175/JHM-D-15-0177.1, 2016.

Revuelto, J. et al. Assimilation of surface reflectance in snow simulations: impact on bulk snow variables (submitted to Journal of Hydrology).

Schöniger et al.: A statistical concept to assess the uncertainty in Bayesian model weights and its impact on model ranking, WRR, https://doi.org/10.1002/2015WR016918, 2015.

Snyder et al.: Obstacles to high-dimensional particle filtering, MWR, https://doi.org/10.1175/2008MWR2529.1, 2008.

van Leeuwen: Particle Filtering in Geophysical Systems, MWR, https://doi.org/10.1175/2009MWR2835.1, 2009.

van Leeuwen et al.: Particle filters for high-dimensional geoscience applications: A review, QJRMS, https://doi.org/10.1002/qj.3551, 2019.

Warren: Can black carbon in snow be detected by remote sensing?, JGR, https://doi.org/10.1029/2012JD018476, 2013.

CrocO_v1.0: a Particle Filter to assimilate snowpack observations in a spatialised framework

Bertrand Cluzet[1], Matthieu Lafaysse[1], Emmanuel Cosme[2], Clément Albergel[3],
Louis-François Meunier[3], and Marie Dumont[1]

[1]Univ. Grenoble Alpes, Université de Toulouse, Météo-France, CNRS, CNRM, Centre d'Études de la Neige, Grenoble, France
[2]Institut des Géosciences de l'Environnement, IGE, UGA-CNRS, Grenoble, France
[3]CNRM, Université de Toulouse, Météo-France, CNRS, 31057 Toulouse, France.

Correspondence: Bertrand Cluzet (bertrand.cluzet@meteo.fr)

Abstract. Monitoring the evolution of ~~the~~ snowpack properties in mountainous areas is crucial for avalanche hazard forecasting and water resources management. In-situ and remotely sensed observations provide precious information on the state of the snowpack but usually offer a limited spatio-temporal coverage of bulk or surface variables only. In particular, visible-near infrared (VIS-NIR) reflectance observations can ~~inform on~~ provide information about the snowpack surface properties but are limited by terrain shading and clouds. Snowpack modelling enables ~~to estimate~~ the estimation of any physical variable, virtually anywhere, but is affected by large errors and uncertainties. Data assimilation offers a way to combine both sources of information, and to propagate information from observed areas to non observed areas. Here, we present CrocO, (Crocus-Observations) an ensemble data assimilation system able to ingest any snowpack observation (applied as a first step to the height of snow (HS) and VIS-NIR reflectances) in a spatialised geometry. CrocO uses an ensemble of snowpack simulations to represent modelling uncertainties, and a Particle Filter (PF) to reduce them. The PF is ~~known~~ prone to collapse when assimilating ~~a too large number of~~ too many observations. Two variants of the PF were specifically implemented to ensure that observations information is propagated in space while tackling this issue. The *global* algorithm ingests all available observations with an iterative inflation of observation errors, while the *klocal* algorithm is a localised approach performing a selection of the observations to assimilate based on background correlation patterns. ~~Experiments~~ Feasibility-testing experiments are carried out in ~~a~~ an identical twin experiment setup, with synthetic observations of HS and VIS-NIR reflectances available in only a 1/6th of the simulation domain. Results show that compared against runs without assimilation, analyses exhibit an average improvement of snow water equivalent Continuous Rank Probability Score (CRPS) of 60% when assimilating HS with a 40-member ensemble, and an average 20% CRPS improvement when assimilating reflectance with a 160-member ensemble. Significant improvements are also obtained outside the observation domain. These promising results open a way for the assimilation of real observations of reflectance, or of any snowpack observations in a spatialised context.

1 Introduction

Seasonal snowpack is ~~a key~~ an essential element of mountainous areas. Monitoring the evolution of its physical properties is essential to forecast avalanche hazard (Morin et al., 2020), rain-on-snow related floods (Pomeroy et al., 2016; Würzer et al., 2016) and to monitor water resources (Mankin et al., 2015). Observations alone are too scarce to monitor snowpack conditions. In-situ observations provide precise observations of several key variables, but they lack spatial representativeness and have a poor spatial coverage. Remote sensing of snowpack variables such as the height of snow (HS, (m)), snow water equivalent (SWE, $(\mathrm{kg\,m^{-2}})$), visible-near infrared (VIS-NIR) reflectance, or surface temperature, provide comprehensive information over large areas but usually have a limited temporal resolution on a small set of variables. Furthermore, these observations are usually available in fractions of simulation domains only, even for space-borne data (Davaze et al., 2018; Veyssière et al., 2019; Shaw et al., 2019). For instance, snowpack VIS-NIR reflectances from moderate resolution (250-500 m) satellites such as MODIS or Sentinel-3 can help constraining the snowpack surface properties ~~(Dozier et al., 2009)~~such as microphysical properties (characterized by the specific surface area, SSA $(\mathrm{m^2 kg^{-1}})$ and light absorbing particles content (LAP, $(\mathrm{gg_{snow}^{-1}})$)) (Durand and Margulis, 2006; Dozier et al., 2009). However, in the areas covered by clouds, forests, or concerned by high sub-pixel variability (ridges, roughness, fractional snow cover) and shadows, satellite retrievals are less accurate (Masson et al., 2018; Lamare et al., 2020), and data should be filtered out (Cluzet et al., 2020). The higher resolution offered by products from Landsat or Sentinel-2 might be an avenue to this issue (e.g. Masson et al., 2018; Aalstad et al., 2020) but at these resolution, reflectance retrievals are quite noisy due to e.g. digital elevation model errors (Cluzet et al., 2020). Finally, note that pixel fractional snow cover (snow cover fraction, SCF) can be accurately retrieved even from noisy reflectances (Sirguey et al., 2009; Aalstad et al., 2020), but it inherits its spatio-temporal limitations. SCF informativeness is also limited in deep snowpack conditions (De Lannoy et al., 2012).

Snowpack models of different complexity offer an exhaustive spatial and temporal coverage (Krinner et al., 2018). They are applied within several spatial configurations, including collection of points, regular or irregular grids (Morin et al., 2020). In this paper, "spatialised" refers indistinctly to any of these configurations. ~~Only detailed snowpack models enable~~ Detailed snowpack models are the only ones able to assess avalanche hazard and monitor water resources alike (Morin et al., 2020), but ~~they suffer from~~ these applications are limited by their considerable errors and uncertainties (Essery et al., 2013; Lafaysse et al., 2017)~~, limiting their use~~. In that context, combining remote sensing observations with models through data assimilation is an appealing solution ~~(?). It enables to optimally combine~~ (Largeron et al., 2020). Indeed, data assimilation combines the spatial and temporal coverage of snowpack models with the available information from observations in an optimal way. Assimilation of optical reflectance could reduce modelled SWE errors by up to a factor of two ~~(Charrois et al., 2016)~~(Durand and Margulis, 2007; Charrois et al., 2, and preliminary studies showed its potential for spatialised assimilation (Cluzet et al., 2020). Assimilation of HS is very efficient in reducing modelled SWE errors (Margulis et al., 2019). However, the limited spatial coverage of observations is stressing the need for data assimilation algorithms able to ~~spread~~ propagate the snowpack observations information into the ~~non-observed areas (Winstral et al., 2019; Cantet et al., 2019; ?)~~unobserved areas (Winstral et al., 2019; Cantet et al., 2019; Largeron et al., 2020)

.

The Particle Filter with ~~Sequential Importance Resampling (PF, Gordon et al., 1993; Van Leeuwen, 2009)~~ sequential importance resampling (PF-SIR, Gordon et al., 1993; Van Leeuwen, 2009) is a Bayesian ensemble data assimilation technique well suited to snowpack modelling ~~(Magnusson et al., 2017). This sequential algorithm relies~~ (Dechant and Moradkhani, 2011; Charrois et al., 2016; M. The PF-SIR is a sequential algorithm relying on an ensemble of model runs (particles) ~~to represent~~ which represents the forecast uncertainty. At each observation date, the prior (or background) composed of the particles is evaluated against the observations. The ~~PF posterior (or analysis) is built by replicating the particles that are nearest to the observation (with respect~~ analysis of the PF-SIR (later on "PF") works in two steps. In a first step, so-called "importance sampling", the particles are weighted according to their distance to the observations (relative to the observation ~~error)and discarding the others.~~ errors). Then, a resampling of the particles is performed in order to reduce the variance in the weights. The Ensemble Kalman Filter (EnKF Evensen, 2003), has also been widely used for snow cover data assimilation (e.g. Slater and Clark, 2006; De Lannoy et al., 2012; M. However, the PF is more adapted to models with a variable number of numerical layers such as detailed snowpack models (Charrois et al., 2016).

The PF could be used in a spatialised context to ~~spread observations information~~ propagate the information from observation as suggested by ~~?~~ Largeron et al. (2020) and Winstral et al. (2019). ~~This~~ Contrary to the EnKF, such applications are rare to date (e.g. Thirel et al., 2013; Baba et al., 2018; Cantet et al., 2019). Indeed, spatialised data assimilation with the PF is not straightforward because of the ~~PF degeneracy~~ degeneracy issue, i.e. ~~the required~~ only a few particles are replicated in the analysis, often resulting in a poor representation of the forecast uncertainties. Degeneracy can be mitigated by increasing the number of particles, but the required population scales exponentially with the number of observations ~~(Bengtsson et al., 2008; Snyder et al., 2008). This issue is a severe drawback~~ simultaneously assimilated (Snyder et al., 2008). Furthermore, an accurate representation of spatial error statistics by the ensemble is essential for the success of the assimilation system. To achieve that, the required ensemble size also scales exponentially with the system dimension, an issue known as the curse of dimensionality (Bengtsson et al., 2008). These issues are severe drawbacks when considering applications of the PF on large domains (i.e. implying a large number of observations and/or simulation points) with a reasonable number of particles (Stigter et al., 2017).

Several solutions exist to tackle the PF degeneracy. A first approach is to inflate the observation errors in the PF. The tolerance of the PF is increased, leading to more particles being replicated. This approach is based on the fact that observation error statistics (including sensor, retrieval and representativeness errors) are usually poorly known and underestimated. It can also be used as a safeguard to prevent the PF to degenerate on specific dates, when observations are not compatible with the ensemble. PF inflation was successfully implemented in point scale simulations of the snowpack (Larue et al., 2018). When dealing with a large number of observations, inflation might lead to degeneracy or null analysis (posterior equal to the prior). In this work, we generalize over space the inflation of Larue et al. (2018), trying to ingest all the observations into a single analysis over the domain, in a so-called *global* approach.

PF localisation is a more widespread alternative, tackling degeneracy by reducing the number of observations that are simultaneously assimilated by the PF (Poterjoy, 2016; Poterjoy and Anderson, 2016; Penny and Miyoshi, 2016; Poterjoy et al., 2019,

3

italic notations are taken from the review of Farchi and Bocquet, 2018). In this method, the simulation domain is divided into *blocks* where different PF analyses are performed considering a local subset of observations (*domain*) based on a localisation radius. ~~It~~ This makes it possible to constrain the model in locations that are not directly observed, but with nearby observations. Contrary to *global* approaches, localisation has the disadvantage of producing spatially discontinuous analyses (each point

95   receives a different analysis). This issue can be mitigated in various ways (Poterjoy, 2016; Farchi and Bocquet, 2018; Van Leeuwen et al., 2

.

The underlying hypothesis of localisation is that model points are independent beyond a certain distance, i.e. constraining one point with the observation from a too distant point would be meaningless, and likely degrade the analysis performance (Houtekamer and Mitchell, 1998). However, in the case of small simulation domains or modelled systems driven by large-scale

100   coherent causalities, large scale correlations (relative to the domain size) may be physically sound, and defining a localisation radius may be a difficult task. In order to face this issue, we developed a new localisation approach called the k-localisation, where localisation *domains* are based on background correlation patterns.

These developments were implemented into CrocO (Crocus-Observations) an ensemble data assimilation system able to sequentially assimilate snowpack observations with a PF in a spatialised context. CrocO can be implemented in any geome-

105   try, (e.g. within a distributed (gridded) framework or any irregular spatial discretisation). Here, we apply CrocO in a semi-distributed framework, which is a conceptual spatialised geometry, ~~operationally used~~ used operationally by Météo-France for avalanche hazard forecasting (Lafaysse et al., 2013; Morin et al., 2020). This framework is similar to many topographic-based discretisation in hydrological models (e.g. Clark et al., 2015). This setup enables us to account for the snowpack variability induced by the topography at the scale of a mountain range, through meteorological conditions (elevation controls the air tem-

110   perature and precipitation phase) and the snowpack radiative budget (also dependent on the aspect and slope angle) (Durand et al., 1993).

CrocO uses an ensemble of stochastic perturbations of SAFRAN meteorological analysis (Durand et al., 1993; Charrois et al., 2016) to force ESCROC ~~(Ensemble System CROCus, Lafaysse et al. (2017)),~~ (Ensemble System CROCus, Lafaysse et al., 2017) , the multi-physical version of Crocus snowpack model (Vionnet et al., 2012). The ensemble setup accounts for the major

115   sources of uncertainties in snowpack modelling (Raleigh et al., 2015) and was formerly described and evaluated in the semi-distributed geometry by Cluzet et al. (2020).

Inflation and k-localisation were implemented into CrocO. Here, we present CrocO and evaluate how it addresses the issues of reflectance observation sparseness and PF degeneracy in the context of snowpack modelling. This problem is divided into two scientific questions: (1) Is CrocO PF able to efficiently spread the information from sparse observations in space without

120   degenerating? (2) Is the spatial information content of reflectance observations valuable for snowpack models? We assess these questions by evaluating the performance of CrocO to model the SWE when assimilating synthetic observations of HS and reflectance covering only a portion of the domain.

Section 2 presents the CrocO system, i.e. the ensemble modelling system and the PF algorithms. Section 3 introduces the evaluation methodology. ~~Following~~ Subsequently, Sec. 4 assesses the performance of CrocO and Sec. 5 discusses the results.

125   Finally, Sec. 6 provides perspectives and research directions.

2 Material and methods

2.1 Modelling geometry

Simulations are performed in the semi-distributed geometry. Mountain ranges such as the Alps are discretized into so-called *massifs* of about 1000 km² to account for regional variability of meteorological conditions. Within each massif, topographic-
induced variability is taken into account by running the model ~~into~~ for a fixed set of topographic classes, e.g. by 300 m elevation bands, for 0º, 20º and 40º slopes and 8 aspects (see Fig. 1). This set enables us to reproduce the main features of snowpack variability (e.g. Mary et al., 2013).

In this study, we focus on the Grandes Rousses, a single massif in the central French Alps. This area of about 500 km² is represented by ~~$N_{pts} = 187$~~ $N_{pts} = 187$ independent topographic classes (see Fig. 1). In the following, specific topographic classes are denoted as follows: *elevation_aspect_slope*, e.g. 1800_N_40 stands for a 40º slope, with a northern aspect at 1800 ~~m. a.s.l.~~ m.a.s.l.

2.2 CrocO Ensemble data assimilation setup

The ensemble data assimilation workflow of CrocO is represented in Fig. 2. In the following, only a short description of the system and its elements is provided. More details on the ensemble modelling setup are available in Cluzet et al. (2020). Information about its implementation into Météo-France HPC system can be found in Appendix B1.

2.2.1 Ensemble of snowpack models

Crocus is a detailed snowpack model, ~~enabling to represent the snowpack coupling~~ coupled with the ground and atmosphere in the ISBA land surface model (Interaction Soil-Biosphere-Atmosphere). It is embedded within the SURFEX_v8.1 modelling platform ~~(SURFace Externalisée, Masson et al. (2013))~~ (SURFace EXternalisée, Masson et al., 2013). The TARTES optical scheme (Libois et al., 2013, 2015) represents VIS-NIR spectral radiative transfer within the snowpack, driven by snow metamorphism (Carmagnola et al., 2014) and Light Absorbing Particles (LAP (gg_{snow}^{-1})) deposition fluxes (Tuzet et al., 2017). Moreover, TARTES computes the snowpack reflectance with a high spectral resolution, making the model directly comparable to the observations. ~~This way~~ As such, TARTES is both a physical component of Crocus and an observation operator.

ESCROC (Ensemble System CROCus, Lafaysse et al., 2017) multi-physical ensemble version of Crocus is used to account for snowpack modelling uncertainties. A random draw among 1944 ESCROC multi-physics configurations was performed and used in all the simulations and denoted ~~$(M_i)_{0 < i \leq N_e}$, N_e~~ $(M_i)_{0 < i \leq N_e}$, N_e being the ensemble size (e.g. 40 or 160 members, see Fig. 2). These configurations are considered equiprobable before any data assimilation.

2.2.2 Ensemble of meteorological forcings

Meteorological forcings are taken from SAFRAN (Durand et al., 1993) reanalysis, where forecasts from the ARPEGE Numerical Weather Prediction (NWP) model are downscaled and adjusted with surface observations within the massif area. They are combined with MOCAGE LAP fluxes (Josse et al., 2004) interpolated at Col du Lautaret (2058 ~~m.a.s.l~~m.a.s.l, inside the Grandes-Rousses) to constitute the reference forcing dataset. Before the beginning of the simulation, spatially homogeneous stochastic perturbations (e.g. at a given date, the same perturbation parameter is applied across the whole domain) with temporal auto-correlations are applied to this forcing to generate an ensemble of forcings ~~$(F_i)_{0<i\leq N_e}$~~ $(F_i)_{0<i\leq N_e}$ with the same procedure as described in Cluzet et al. (2020). ~~Each~~More details on the perturbations procedure can be found in Appendix A. At the beginning of the simulation, each forcing F_i is associated with ~~the corresponding~~a random M_i ESCROC configuration and this relation is fixed during the whole simulation.

2.2.3 The Particle Filter in CrocO

The PF is applied sequentially at each observation date on the background state vectors (soil and snowpack state variables, denoted *BG* on Fig. 2). Its analysis is an ensemble of initial conditions used to propagate the model forward. The algorithm is implemented into SODA ~~(SURFEX Offline Data Assimilation, Albergel et al. (2017)),~~ ((SURFEX Offline Data Assimilation, Albergel et al., , the data assimilation module of SURFEX_v8.1, enabling a continuous execution sequence between ensemble propagation and analysis, as depicted in Fig. 2.

2.3 The Particle Filter equations

At a given observation date, we consider a set of observed variables available at several locations, totalling N_y different observations.

- Each member $\widehat{\boldsymbol{x}_b^i}$ $\boldsymbol{x}_b^i$ of the background state $\widehat{\mathbf{X}_b}$ $\mathbf{X}_b$ is projected into the observation space using the observation operator h. In our case, h is just an orthogonal projection on the N_y observations since HS and reflectance are diagnosed within Crocus (see Sec. 2.2.1). The projection $\widehat{\boldsymbol{x}_b^i} = h\widehat{\boldsymbol{x}_b^i} = (x_k^i)_{0<k\leq N_y}$ $\widehat{\boldsymbol{x}}_b^i = h\boldsymbol{x}_b^i = (\widehat{x}_k^i)_{0<k\leq N_y}$, corresponds to the modelled values at each observed variable/point.

- these N_y observations are collected in the vector $\boldsymbol{y} = (y_k)_{0<k\leq N_y}$. The associated observation error covariance matrix $\mathbf{R}$ (Eq. 1) is ~~supposed~~diagonal (e.g. observation errors are assumed independent):

$$\mathbf{R} = \underline{\mathbf{diag}}(\mathbf{diag}(\sigma_k^2, 0 < k \leq N_y) \tag{1}$$

Where σ_k^2 stands for the observation error variance of observation k and depends only on the type of ~~variable of~~observation y_k (e.g. HS or reflectance).

The PF analysis usually works in two steps.

- (1) computing the particle weights w^i as the normalised observation likelihood for each particle (Eq. 2):

$$w^i = \frac{e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{x}_b^i)^T \mathbf{R}^{-1}(\boldsymbol{y}-\boldsymbol{x}_b^i)}}{\sum_{k=1}^{N_e} e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{x}_b^k)^T \mathbf{R}^{-1}(\boldsymbol{y}-\boldsymbol{x}_b^k)}} \quad \frac{e^{-\frac{1}{2}(\boldsymbol{y}-\widehat{\boldsymbol{x}}_b^i)^T \mathbf{R}^{-1}(\boldsymbol{y}-\widehat{\boldsymbol{x}}_b^i)}}{\sum_{k=1}^{N_e} e^{-\frac{1}{2}(\boldsymbol{y}-\widehat{\boldsymbol{x}}_b^k)^T \mathbf{R}^{-1}(\boldsymbol{y}-\widehat{\boldsymbol{x}}_b^k)}} \tag{2}$$

- (2) resampling the particles based on their weights to build the analysis vector $\widehat{\mathbf{X}_a}\mathbf{X}_a$. Here, we apply the PF resampling from Kitagawa (1996) which returns $\boldsymbol{s}=(s_i)_{0<i\leq N_e}, (s_i\in[1..N_e]) \boldsymbol{s}=(s_i)_{0<i\leq N_e}, (s_i\in[1..N_e])$ a sorted vector with duplications, representing the particles to replicate.

A sample reordering step was added for numerical optimisation with no expected incidence on the PF behaviour (see in Appendix B2 for more details).

Two simple variants of this algorithm can be identified in a spatialised context:

- *global* approach: perform one analysis over the domain, putting all the available observations in $\boldsymbol{y}$.

- *rlocal* approach: perform one analysis per model point, assimilating only local observations, if any. This corresponds to a localised PF with *block* and *domain* size of 1.

2.3.1 Particle Filter degeneracy

Degeneracy occurs when only a small fraction of the particles have non-negligible weights, resulting in a sample $\boldsymbol{s}$ where only a few different indices are present. It can be diagnosed from the weights using the effective sample size N_{eff} ~~(Liu and Chen, 1995; Doucet et al.~~ ~~:~~ (Liu and Chen, 1995):

$$N_{\text{eff}} = \frac{1}{\sum_{i=1}^{N_e}(w^i)^2} \quad \frac{1}{\sum_{i=1}^{N_e}(w^i)^2} \tag{3}$$

With a degenerate sample, $N_{\text{eff}} \gtrsim 1$, and with innocuous analysis (all particles are replicated) ~~$N_{\text{eff}}=N_e$~~ $N_{\text{eff}}=N_e$.

A first approach to mitigate degeneracy is to use inflation. This heuristic method iteratively inflates $\mathbf{R}$ values until the ~~sample population~~ effective sample size is large enough. Here, we develop a variant from Larue et al. (2018) method, which was not explicitly relying on N_{eff} (Eq. 3). Consider applying an inflation factor $\frac{1}{\alpha}$ to $\mathbf{R}$, ($0 < \alpha \leq 1$, $\alpha = 1$ being the value for no inflation), and update N_{eff} (Eqs. 2 and 3): N_{eff} is naturally a decreasing function of α (the more we inflate $\mathbf{R}$ the more different particles will be replicated). The idea of our method is to ensure that N_{eff} exceeds a target value, N_{eff}^*. If $N_{\text{eff}} < N_{\text{eff}}^*$ (degenerate case), we reduce α (inflate) until $N_{\text{eff}} = N_{\text{eff}}^*$ using Alg. 1. In the following, inflation is used in the *global* and *rlocal* PF (see Sec. 2.2.3).

The core of Alg. 1 is an hybrid bisection-secant method to find the zero of $f : \alpha \mapsto N_{\text{eff}}(\alpha) - N_{\text{eff}}^*$ in $[0,1]$. It is inspired ~~on~~ by **rtsafe** algorithm (Vetterling et al., 1992). The **guess** function computes a new guess α_2 to minimize f. Note that in the unlikely case where Alg. 1 does not converge, all the particles are replicated.

Algorithm 1 Weighting algorithm with inflation.

Input: ~~$\boldsymbol{x}^i, \boldsymbol{y}, \mathbf{R}, N_{\text{eff}}^*$~~ $\widehat{\boldsymbol{x}}^i, \boldsymbol{y}, \mathbf{R}, N_{\text{eff}}^*$

Output: ~~$\boldsymbol{w}^i$~~ w^i

1: $\alpha \leftarrow 1$

2: $\mathbf{R} \leftarrow \frac{1}{\alpha}\mathbf{R}$

3: ~~$\boldsymbol{w}^i \leftarrow$~~ $w^i \leftarrow$ **weights**(~~$\boldsymbol{x}^i, \boldsymbol{y}, \mathbf{R}$~~ $\widehat{\boldsymbol{x}}^i, \boldsymbol{y}, \mathbf{R}$) (Eq. 2)

4: $N_{\text{eff}} \leftarrow$ **eff_weights**(w^i) (Eq. 3)

5: **if** $N_{\text{eff}} < N_{\text{eff}}^*$ **then**

6: $\quad \alpha_1 \leftarrow 0$

7: $\quad$ ~~$N_{\text{eff}_1} \leftarrow N_e$~~ $N_{\text{eff}_1} \leftarrow N_e$

8: $\quad$ ~~$cond \leftarrow True$~~ $cond \leftarrow$ **True**

9: $\quad i \leftarrow 0$

10: $\quad$ **while** $cond$ **do**

11: $\quad\quad \alpha_2 \leftarrow$ **guess**($\alpha_1, \alpha, N_{\text{eff}_1}, N_{\text{eff}}, N_{\text{eff}}^*$)

12: $\quad\quad \mathbf{R} \leftarrow \frac{1}{\alpha_2}\mathbf{R}$

13: $\quad\quad w_2^i \leftarrow$ **weights**(~~$\boldsymbol{x}^i, \boldsymbol{y}, \mathbf{R}$~~ $\widehat{\boldsymbol{x}}^i, \boldsymbol{y}, \mathbf{R}$) (Eq. 2)

14: $\quad\quad N_{\text{eff}_2} \leftarrow$ **eff_weights**(w_2^i) (Eq. 3)

15: $\quad\quad$ **if** $|N_{\text{eff}_2} - N_{\text{eff}}^*| < \epsilon$ **then**

16: $\quad\quad\quad$ ~~$cond \leftarrow False$~~ $cond \leftarrow$ **False**

17: $\quad\quad\quad \alpha \leftarrow \alpha_2$

18: $\quad\quad\quad w^i \leftarrow w_2^i$

19: $\quad\quad$ **else**

20: $\quad\quad\quad \alpha \leftarrow \alpha_1$

21: $\quad\quad\quad \alpha_1 \leftarrow \alpha_2$

22: $\quad\quad\quad N_{\text{eff}} \leftarrow N_{\text{eff}_1}$

23: $\quad\quad\quad N_{\text{eff}_1} \leftarrow N_{\text{eff}_2}$

24: $\quad\quad$ **end if**

25: $\quad\quad i \leftarrow i + 1$

26: $\quad\quad$ **if** $i = maxiter$ **then**

27: $\quad\quad\quad$ **print** "failed to converge, duplicating all particles"

28: $\quad\quad\quad$ ~~$w^i \leftarrow \frac{1}{N_e}$~~ $w^i \leftarrow \frac{1}{N_e}$

29: $\quad\quad$ **end if**

30: $\quad$ **end while**

31: **end if**

2.3.2 k-localisation

215 In the k-localisation algorithm, degeneracy is mitigated by reducing the number of observations that are simultaneously assimilated. The PF analysis is applied to each simulation point sequentially. In order to build the analysis at point n, background correlations $\mathbf{B}_v$ are computed for each variable v (e.g HS or reflectance) between n and all the observed points. In a first step, all observations from points exhibiting ~~significant~~ substantial background correlations (see below **select_k_biggest** function) are used. If the PF degenerates, the number of observations is progressively decreased until degeneracy is mitigated. As earlier,

220 degeneracy is considered mitigated when $N_{\text{eff}} \geq N_{\text{eff}}^*$. This way, we ensure that a maximal number of observations has been ingested by the PF without degenerating.

In case of degeneracy, the observation point displaying the lowest correlation is ruled out. The PF weights are computed (Eq. 2), and a new effective sample size is derived (Eq. 3). While the target sample size is not exceeded, this selection proceeds iteratively. The notation k in "k-localisation" refers to the number k of retained observations of each variable. This approach

225 is similar to EnKF localisation algorithm where the localisation *domain* is based on background correlations (Hamill et al., 2001).

The detailed k-localisation algorithm is described in Alg. 2, where:

- The **select_k_biggest** method returns for each variable, the domain d_v, of up to k observed points (named p) that are the most correlated (in absolute value) with n, and match the following criteria, which were adjusted in preliminary

230 experiments:

- in ~~$\widehat{x_v^i}$~~ x_v^i, there are at least 10% of members defined in both points~~(. As~~ reflectance is not defined when there is no snow~~)~~, spurious high correlation can be obtained when the computation of correlations is based on a very low number of pairs.

- $|\mathbf{B}_v(n,p)| > 0.3$ ~~: correlations are significant.~~. If the absolute correlation is too low, it is likely that there is a

235 poor potential for the distant observation to constrain the ensemble locally. In such a situation, it is better to reject the observation from the local analysis. Negative ensemble correlations can be physically sound, e.g. after a rain-on-snow event between the HS of two points separated by the rain-snow line. In such a situation, an HS observation on either point can hold information on precipitation rates at both locations. At the observed location, the PF will probably select the members with the most appropriate precipitation rates. This sample is likely to

240 perform well at both locations, so it can be used to constrain the unobserved location.

- d is the collection of the domains d_v

- **extract_points** extracts d from ~~y, x^i~~ $y, \widehat{x}^i$ and $\mathbf{R}$.

Algorithm 2 k-localisation algorithm

Input: ~~$\boldsymbol{x}^i$~~ $\widehat{\boldsymbol{x}}^i$, $\boldsymbol{y}$, $\mathbf{B}$, $\mathbf{R}$, N_{eff}^*

Output: ~~$(\boldsymbol{w}_n^i)_{0<n\leq N_e}$~~ $(w_v^i)_{0<n\leq N_{\text{pts}}}$

1: **for** $n=1$ to ~~N_{pts}~~ N_{pts} **do**
2:     ~~$k \leftarrow k_{max}$~~ $k \leftarrow k_{\max}$ {try to ingest all available observations.}
3:     ~~$cond \leftarrow True$~~ $cond \leftarrow$ **True**
4:     **while** $cond$ and $k>0$ **do**
5:         **for** $v=1$ to ~~N_v~~ N_v **do**
6:             ~~$\boldsymbol{d_v} \leftarrow$~~ $\boldsymbol{d}_v \leftarrow$ **select_k_biggest**($n,k,$~~$\mathbf{B_v},\boldsymbol{x}_v^i$~~$\mathbf{B}_v, \widehat{\boldsymbol{x}}_v^i$, $\boldsymbol{y}$)
7:         **end for**
8:         ~~$\boldsymbol{y_k},\boldsymbol{x}_k^i,\mathbf{R_k} \leftarrow$~~ $\boldsymbol{y}_k, \widehat{\boldsymbol{x}}_k^i, \mathbf{R}_k \leftarrow$ **extract_points**($\boldsymbol{y}$, ~~$\boldsymbol{x}^i$~~$\widehat{\boldsymbol{x}}^i$, $\mathbf{R}$, $\boldsymbol{d}$)
9:         $w_n^i \leftarrow$ **weights**(~~$\boldsymbol{x}_k^i,\boldsymbol{y_k},\mathbf{R_k}$~~$\widehat{\boldsymbol{x}}_k^i, \boldsymbol{y}_k, \mathbf{R}_k$) (Eq. 2)
10:         $N_{\text{eff}} \leftarrow$ **eff_weights**(w_n^i) (Eq. 3)
11:         **if** $N_{\text{eff}} \geq N_{\text{eff}}^*$ **then**
12:             ~~$cond \leftarrow False$~~ $cond \leftarrow$ **False**
13:         **end if**
14:         $k \leftarrow k-1$
15:     **end while**
16:     **if** $k=1$ **then**
17:         $w_n^i \leftarrow$ **inflation**(~~$\boldsymbol{x}_k^i,\boldsymbol{y_k},\mathbf{R_k}$~~$\widehat{\boldsymbol{x}}_k^i, \boldsymbol{y}_k, \mathbf{R}_k$, N_{eff}^*)
18:     **end if**
19: **end for**

2.3.3 Particle Filter and reflectance observations

Assimilating reflectance with the PF requires some adaptations. ~~Snow reflectance is a bounded variable (0-1) and is not defined~~ In Crocus, TARTES optical scheme (see Sec. 2.2.1) only provides snow reflectance, not all-surface reflectance: no value for the surface reflectance is issued in the absence of snow. ~~For this reason,~~ Conversely, the weights of the particles are not defined in Eq. 2 if the members are snow-free. These issues were roughly accommodated by setting the reflectances of snow-free members and observations ~~were set~~ to 0.2 (~~snow-free ground reflectance value in Crocus~~the value of bare soil broadband albedo in ISBA model) in the PF Eq. 2 (Sec. 2.2.3).

3 Evaluation strategy

Our strategy is to assess the performance of the analysis by means of twin experiments, i.e. using synthetic observations (e.g. Reichle and Koster, 2003). The assimilation run is compared to an identical run without assimilation (~~openloop~~ open-loop run). Synthetic observations are extracted from a model run and assimilated without adding any noise. These observations

10

allow to mimic real observations with a perfect knowledge of the true state. Analysis and openloop open-loop experiments can

therefore be compared with this true state anywhere, for any variable. It allows in In a first step, this allows us to get rid of the error and bias issues inherent to real observations (Cluzet et al., 2020), a reason why we did not add any noise to the synthetic observations as commonly done in twin experiments (Lahoz and Menard, 2010). This way, we can focus on the two following questions (see Sec. 1):

– Is CrocO PF able to efficiently spread the information from sparse observations into space without degenerating?

– Is spatial information content of reflectance a valuable source of information for snowpack models?

In order to disentangle these questions, we run baseline experiments assimilating synthetic observations of HS which is strongly linked with SWE (Margulis et al., 2019). These experiments are used to evaluate the PF algorithms efficiency, and as a baseline for synthetic reflectance assimilation experiments evaluating the information content of reflectance.

Three different algorithms are evaluated: the *global* algorithm (with inflation), the *rlocal* algorithm (with inflation) and the k-localized algorithm *klocal*.

3.1 Experiments

3.1.1 Twin experiments setup

In our twin experiment setup, an openloop open-loop ensemble is used as a reference and to generate synthetic observations. Openloop open-loop simulations are carried out with CrocO for 4 consecutive winters (2013-2017) in the Grandes-Rousses (see Sec. 2.1), with 160 members. For each year, the integral average of SWE over time and space is computed from each member, and members corresponding to the 20th, 40th, 60th and 80th percentiles of the ensemble are extracted to be used as synthetic observations (denoted *year_ppercentile* e.g. 2014_p80). This method enables us to evaluate the efficiency of data assimilation experiments under contrasted snow condition scenarios. Before any assimilation experiment, the openloop open-loop member ($F_i - M_i$ couple in Fig. 2) used as true state is withdrawn and replaced by a new random member.

The spatial coverage of synthetic observations was reduced, mimicking a typical reflectance mask. Synthetic observations were only available above an assumed constant tree line at 1800 m (see Fig. 1), and not available in steep slopes (over 20º) and in northern aspects (shadows, considering a daily satellite pass around 10-11:00 UTC.), for the whole snow season. As a result, in this case, only 35 (over 187) topographic classes are observed. Observation date dates were chosen corresponding to clear-sky days with a MODIS overpass, resulting in an approximately weekly frequency (e. g. Revuelto et al. (2018); Cluzet et al. (2020)). (e.g. Revuelto et al., 2018; Cluzet et al., 2020).

Considering reflectance, a Reflectance is sensitive to the surface SSA and LAP (see Sec. 1). A minimal set of two different bands is used, corresponding to MODIS sensor band 4 (555 nm, sensitive to SSA and LAP) and 5 (1240 nm) (Charrois et al., 2016), usually only sensitive to SSA) (e.g. Fig. 2. of Cluzet et al., 2020). Observation error variances is are set to $1.0 \times 10^{-2} \mathrm{m}^2$ for HS and 5.6×10^{-4} and 2.0×10^{-3} for band 4 and band 5 reflectance respectively (Wright et al., 2014). These values are only

initial values for the inflation in the *global* and *rlocal* algorithms. Since the *klocal* algorithm ~~uses only~~ only used inflation if k drops to 1 (see Sec. 2), observation error variances are multiplied by a factor of 5 to enable the *klocal* algorithm to ingest observations from several points.

In order to study the ability of the *global*, *klocal* and *rlocal* algorithms to spread the information in space, a first set of experiments is conducted assimilating HS with 40 members (see setup in Tab. 1). In order to evaluate the algorithms ability to assimilate reflectance (Band 4 and Band 5) a second set of experiments is conducted, other things being equal (Tab. 2). The ensemble size is increased from 40 to 160 in a third set of experiments assimilating reflectance, in order to analyse the influence of a larger ensemble on the algorithms performance (Tab. 3). Note in Tab. 1-3 that N_{eff}^* is adjusted to the ensemble size, in order to preserve ~~$N_e/N_{\text{eff}}^* \approx 5 - 7$~~ $N_\text{e}/N_{\text{eff}}^* \approx 5 - 7$ following Larue et al. (2018).

3.2 Evaluation Scores

The performance of the assimilation and ~~openloop~~ open-loop run is evaluated against the synthetic truth using several scores. The Absolute Error of the ensemble mean (~~AE~~AEM) and ensemble spread σ are two common metrics of ensemble modelling. Given an ensemble $E_{m,c,t}$ of ~~N_e~~ N_e members m in topographic class c at time t and the corresponding ~~observations $o_{c,t}$~~ truth $\mathcal{T}_{c,t}$, the ensemble mean is described by Eq. 4:

$$\overline{E}_{c,t} = \frac{1}{\underline{N_e}} \frac{1}{N_\text{e}} \sum_{m=1}^{\underline{N_e}\,N_\text{e}} E_{m,c,t}$$ (4)

From which we can compute the absolute error ~~AE~~ AEM (Eq. 5) and the spread (or dispersion) σ (Eq. 6):

$$\underline{AE}\,\underline{AEM}_{c,t} = |\overline{E}_{c,t} - \underline{o}\mathcal{T}_{c,t}| \quad \forall(c,t) \in [1, N_{\underline{pts}\,\text{pts}}] \times [1, N_t]$$ (5)

$$\sigma_{c,t} = \sqrt{\frac{1}{N_e} \sum_{m=1}^{N_e} (E_{m,c,t} - \overline{E}_{c,t})^2}\sqrt{\frac{1}{N_\text{e}} \sum_{m=1}^{N_\text{e}} (E_{m,c,t} - \overline{E}_{c,t})^2} \quad \forall(c,t) \in [1, N_{\underline{pts}\,\text{pts}}] \times [1, N_t]$$ (6)

Where N_t is the number of evaluation time steps.

The Continuous Ranked Probability Score (CRPS, (Eq. 7) Matheson and Winkler, 1976) evaluates the reliability and resolution of an ensemble based on a verification dataset. An ensemble is reliable when events are forecast with the right probability, and has a good resolution when it is able to discriminate distinct observed events. For a reliable system, the resolution is equivalent to the sharpness, which is the spread of the produced forecasts.

If we denote $F_{c,t}$ the Cumulative Distribution Function (CDF) and ~~$O_{c,t}$ the corresponding observation~~ $\mathcal{T}_{c,t}$ the corresponding truth CDF (Heaviside function centred on the truth value), the CRPS is computed at (c,t) following:

$$\text{CRPS}_{c,t} = \int_{\mathbb{R}} (F_{c,t}(x) - \underline{O}\mathcal{T}_{c,t}(x))^2 dx \quad \forall(c,t) \in [1, N_{\underline{pts}\,\text{pts}}] \times [1, N_t]$$ (7)

315 In this work, CRPS$_{c,t}$ value is averaged over time alone or time and space depending on the desired level of aggregation. The CRPS can be decomposed in two terms following ~~Hersbach (2000) (Eq. 8)~~Candille et al. (2015):

$$\text{CRPS} = \text{Reli} + \text{Resol} \tag{8}$$

Where Reli quantifies the reliability of the ensemble. The associated skill scores (CRPSS and ReliS) can be used to compare the performance of an ensemble E to a reference R, here, the ~~openloop~~ open-loop run:

320 $$\text{CRPSS}(E) = 1 - \frac{\text{CRPS}(E)}{\text{CRPS}(R)} \tag{9}$$

A skill score of 1 denotes a perfect score, 0 a neutral performance and $-\infty$ is the worst achievable skill score.

4 Results

4.1 Preliminary Results

4.1.1 Impact of the inflation

325 The inflation algorithm was introduced by Larue et al. (2018) in point scale simulations but to the best of our knowledge, never applied in a spatialised context. Here we evaluate its impact on the *global* algorithm by switching it on/off. As an example, Fig. 3 shows the impact of the inflation on SWE when assimilating the HS of 2015_p80 (as defined in Sec. 3.1.1) member with the *global* algorithm, in a topographic class which is not observed (1800_N_40, as defined in Sec. 2.1). This choice of member and topographic class is ~~well~~ representative of the impact of the inflation on the *global* algorithm.

330 In this case, both inflation ($N^*_{\text{eff}} = 7$) and no inflation ($N^*_{\text{eff}} = 1$) lead to a significant reduction of the ensemble spread compared with the ~~openloop~~ open-loop (Fig 3~~a-b~~b). From January 2015 until the peak of SWE in mid-April 2015, (Fig. 1c) the simulation with inflation has significantly lower errors than without inflation and the ~~openloop~~ open-loop (10-20 $\text{kg\,m}^{-2}$ vs. 60-80 $\text{kg\,m}^{-2}$ and 30-50 $\text{kg\,m}^{-2}$ respectively), leading to a better agreement with the synthetic truth in the melting season (Fig. 3a). During the melting season (mid-April 2015 onwards), the ~~RMSE~~ AEM of the assimilation algorithms is reaching a peak, coinciding with an absence of observations. In comparison, the ~~openloop RMSE~~ open-loop AEM is smaller in the first part of the melting season, but the spread is three times larger, making it almost uninformative. For several analyses (2014, November 21th, and 2014, December 30th for example) the ensemble spread without inflation drops to 0 while its ~~RMSE~~ AEM strongly increases compared to the ~~openloop~~open-loop, suggesting that it is prone to degeneracy.

340 #### 4.1.2 Correlation patterns

The *klocal* algorithm relies on background correlation patterns to define localisation *domains*. To illustrate the potential of using such information in the PF, Fig. 4 shows the correlation patterns of the 40 members ~~openloop in a non observed~~ open-loop in a unobserved topographic class (1800_N_40, red dot) in the mid-winter ~~, several months after the snow season onset~~ (2015,

February 20th)~~for the different assimilation variables . These variables~~, several months after the snow season onset. The assimilation variables exhibit strong but ~~contrasted~~ contrasting correlation patterns. Band 4 (Fig. 4a) correlations are generally high (0.6-1) and uniform. Many of the observed classes (black dots) are strongly correlated with the considered class~~as~~. Similar results are obtained for HS (Fig. 4c). Band 5 (Fig. 4b) exhibits ~~significant~~ substantial correlations, in particular across slopes. However, they are more restricted to the northern aspects, only a few observed classes in the Eastern aspects being ~~significantly~~ substantially correlated with the considered class. ~~Finally, note that~~ Note that negative correlations are evidenced with some lower altitude South-oriented topographic classes (e.g. 1500_S_40 on Fig. 4b). Finally, these patterns vary with time but remain ~~significant~~ substantial along the whole season (not shown), and ~~that~~ increasing the ensemble size up to 160 leads to identical patterns (not shown).

4.2 Results of the experiments

4.2.1 Assimilation of the Height of Snow~~Depth~~

In a first step, assimilation of HS from the different synthetic ~~members was launched~~ observation scenarios was conducted to serve as a reference for reflectance assimilation. Fig. 5 shows the CRPSS (Eq. 9, aggregated over time only) of the HS assimilation with the three PF algorithms considering the synthetic member 2013_q20 as reference. Results for this specific synthetic member were chosen here as a representative example of the algorithms performance.

The *rlocal* performance compared with the ~~openloop~~ open-loop is high (0.7-1), but limited to the observed classes (black dots) since there is no spatial propagation in this algorithm. *global* and *klocal* algorithms have similar, overall good performance, managing to strongly reduce modelling uncertainties except at very low altitudes (600-900 m), (skills of -0.2) where snow does not usually last for more than a few weeks.

This behaviour may vary with the snow conditions, i.e., between the different assimilated synthetic ~~members~~ observation scenarios and from one year to another. In order to generalize this result, Fig. 6 shows the CRPS and Reli (aggregated over time and space) of the different algorithms for the 16 synthetic ~~members~~ observation scenarios and differentiated between observed and ~~non-observed~~ unobserved classes. CRPS and Reliability are considerably reduced compared with the ~~openloop~~ open-loop (by a factor of 2-3 and 4-5, respectively) for all the algorithms in the observed classes. This suggests that the PF manages to reduce the spread of the ensemble while reducing its errors. In the ~~non observed~~ unobserved classes, the gain is almost as good (CRPSS of 0.6) except for the *rlocal* algorithm, which is identical to the ~~openloop~~ open-loop as expected. No significant difference of skill is obtained between *global* and *klocal* algorithms.

4.2.2 Assimilation of Reflectance

Optical reflectance is a promising assimilation variable due to its extended availability in satellite observations, but assimilation of ~~reflectance~~ raw reflectance products is not expected to constrain bulk variables like SWE or HS as ~~well~~ much as HS

14

assimilation. ~~Here~~In order to assess this difference, we conduct ~~reflectance assimilation~~ assimilation of reflectance only, in the same setup as in Sec 4.2.1, all other things ~~equal, to assess this difference~~being equal.

Fig. 7 shows the performance of the reflectance assimilation for the 16 synthetic ~~members~~observation scenarios with 40 members (filled boxes). The different algorithms only lead to moderate improvements in CRPS (median CRPSS of 0.-0.2, median ReliS of 0.2-0.4). Moreover, the *global* and *klocal* algorithms frequently degrade the performance, suggesting that this configuration is not robust.

Suspecting that 40 members is insufficient to ~~well~~properly represent the multivariate probability density function of reflectance and other model variables, the ensemble size was increased to 160 (hatched boxes), leading to ~~significant~~marked improvements in the performance and robustness of the algorithms (median CRPSS of 0.2, median Reli of 0.4-0.6). Reliability of the *global* algorithm is significantly improved ~~with respect~~compared to the *klocal* algorithm.

Fig. 8 shows the spatial performance of the different algorithms for member 2016_p60. Spatial patterns similar to the HS ~~assim~~assimilation are found. *rlocal* performance is limited to the observed classes, while *global* and *klocal* manage to improve the simulations across aspects and slopes. However, ~~Skill~~skill scores are lower than for HS (0.2-0.5), and the performance of all algorithms is poor in the classes that are the farther away from the observations, i.e. at lower elevations ~~, even~~(600-900 m) and in some of the ~~observed classes.~~high altitude steep Northern classes (e.g. 2100_N_40 on Figs. 8b-c). Finally note that slight degradations of performance can sometimes be evidenced even in the observed classes for all the algorithms (e.g. in flat conditions at 3300 m on Fig. 8a for the *rlocal*, not evidenced by this example for the other algorithms).

5 Discussion

In this section, we discuss the performance of CrocO PF algorithms using the assimilation of HS, and consider the potential of the assimilation of reflectance in view of assimilating real data.

5.1 Tackling Particle Filter degeneracy

Because they assimilate several observations at the same time, *global* and *klocal* approaches could be prone to PF degeneracy. However, they almost never degrade the performances when assimilating HS in a variety of years and synthetic ~~members~~observation scenarios percentiles (Fig. 6). This suggests that either inflating the observation errors (as demonstrated by Larue et al. (2018), a result we have generalized in space) or exploiting background correlations to reduce the number of assimilated observations, are two efficient approaches to tackle degeneracy.

In several cases though, a strong degradation of score occurs when assimilating reflectance (Fig. 7), which could either be attributed to an algorithmic failure in the PF, or an intrinsic lack of informativeness of reflectance in some situations. Based on the good behaviour of the algorithm with HS, and because by construction, *global* and *klocal* algorithms cannot lead to a degenerate PF sample we consider this comes from the reflectance itself (this point will be further discussed in the following sections).

Beyond tackling degeneracy, *global* and *klocal* algorithms also beat the *rlocal* approach on Reli and CRPS scores (Figs. 7 and 8). This suggests that assimilating multiple observations increases the quality of the PF analysis, even locally. More precisely, most of the improvement is due to the Reli term of the CRPS. This property is ~~key~~ crucial for ensemble modelling, because it ensures that events are forecasted with a right ~~probability~~frequency. However, this is not sufficient, e.g. the climatology has a perfect reliability but is not informative at all. Successful assimilation manages to improve general metrics such as the CRPS while improving the reliability. On this aspect, the *global* and *klocal* algorithms have a satisfying performance.

5.2   Propagating the observations information

Having sparse observations is one of the most challenging tasks for data assimilation systems of snowpack observations ~~(Magnusson et al., 2014; ?)~~(Magnusson et al., 2014; Largeron et al., 2020). In our partially observed, ~~conceptual~~synthetic setup, the *global* and *klocal* PF variants developed here efficiently propagate the observations information to the ~~non-observed~~ unobserved classes, with generally a better performance than the ~~openloop~~ open-loop and the *rlocal* approach in the ~~non-observed~~ unobserved classes when assimilating HS (Fig. 5).

The algorithms performance is particularly good across aspects and slopes with only a few steep, northern aspect slopes exhibiting neutral to poor performances (Figs. 5 and 8). This suggests that southern aspect and flat classes are informative ~~on~~ for the majority of the simulation domain. Conversely, considering that there are strong background correlations between the western and eastern sides of the domain, we can speculate that observing either side could yield overall good results.

On these figures, propagation of the information is limited towards lower elevation (600-1200 m). At such ~~elevation~~elevations, the snow cover is usually intermittent and a good discrimination of the precipitation phase is crucial. The PF does this indirectly through HS and reflectance observations, because rain causes a decrease of HS through compaction and melting while ~~reflectance~~ Band 4 and Band 5 reflectances also decreases because of ~~because of~~quick isothermal metamorphism ~~.~~(i.e. the surface SSA decreases). However, in our setup, the lowest observed elevation is 1800 m, therefore indirect observation of the rain-snow line positioning under this level is not possible, potentially explaining the moderate performance of the PF there. In that case, assimilation of Snow Cover Fraction might be the best solution: since the snowpack is intermittent there, the informativeness of this variable is maximal (Aalstad et al., 2018).

Global and *klocal* algorithms exhibit strong performances when assimilating HS (Fig. 5)~~, and moderate performances for reflectance~~. HS is ~~well~~ closely linked with the SWE (by the bulk density) and the interest of this variable for data assimilation is clear (Margulis et al., 2019). Here, it should be kept in mind that HS assimilation is used as a baseline experiment to evaluate the algorithms and put reflectance assimilation into perspective. The prescribed HS observation errors ($\sigma_0 = 0.1$ m) are not necessarily realistic. They should be adapted to the nature of the HS sensor. For example, space-borne HS observation errors are typically larger (e.g. Eberhard et al., 2020; Deschamps-Berger et al., 2020). The assimilation of such observations would probably yield lower improvements.

Though the performance is lower for Reflectance ~~than in our HS experiments~~, it remains considerable and in line with previous results on point simulations (Charrois et al., 2016), with an average score improvement of 20-40%. ~~An outstanding result here, is that our study~~ This study quite surprisingly suggests that reflectance information can be spread from southern slopes to the northern ones, although in many situations, the snowpack evolves in different ways ~~between these~~ for these two aspects. For example, in sunny conditions, melt and wet metamorphism will cause a drop in reflectance in southern slopes, while reflectance will not evolve much in northern slopes. ~~Therefore, a reflectance observation in a southern slope is~~ Such a phenomenon could explain why low background correlations between southern and northern aspects are exhibited in Band 5 (Fig. 4), which is the most sensitive to surface metamorphism through SSA. This example shows that Band 5 reflectance observations in southern slopes are not necessarily informative on ~~reflectance value~~ Band 5 reflectance values in the northern aspect per se ~~. It is informative however, in our ensemble data assimilation framework. Indirectly, in this case this observation could~~ on every date. On average, however, the positive impact of reflectance observations suggests that they enable the PF to reject ~~all~~ the ensemble members ~~that did not have an appropriate meteorological forcing~~ with inadequate meteorological forcings (snowfall or cloud cover would lead to wrong reflectance values), or multiphysical parametrisations ~~, thus enabling to correct~~ (influencing e.g. the surface metamorphism), thus correcting the ensemble in the whole domain. These insights are consistent with the study of Winstral et al. (2019), where in situ observations are used to correct meteorological forcing parameters across large simulation domains.

Regarding the observations, our study has some methodological limits, however. Observation errors are very roughly prescribed, and the assimilated observations are not corrupted as usually done in synthetic experiments (e.g. Durand and Margulis, 2006). These choices were motivated by the fact that very little is known about the spatial correlation of reflectance observation errors in the semi-distributed setting (e.g. Cluzet et al., 2020). In a recently submitted paper, the impact of random and systematic errors of reflectance observations on point-scale assimilation experiments is thoroughly investigated (Revuelto et al., in prep). Efforts to better characterize the spatial structure of these observation errors should be conducted in future work.

5.3 Towards the assimilation of real observations of reflectance

Reflectance is an appealing variable for snowpack modelling because of its sensitivity to snowpack surface properties (Dozier et al., 2009) and the abundance of ~~moderate-resolution~~ moderate to high resolution space-borne sensors (MODIS, Sentinel2-3, VIIRS, Landsat...) providing us with a handful of observations to assimilate, contrary to HS. The potential for assimilation of SCF, which is retrieved from reflectances, is clear (Margulis et al., 2016; Aalstad et al., 2018; Alonso-Gónzalez et al., 2020). This study demonstrates the potential of the PF to spread information and assimilate ~~reflectance~~ raw reflectances with a positive impact (Sec. 5.2). Yet, assimilating real observations of reflectance is another challenge, for two reasons.

First, space-borne reflectance observations are ~~usually~~ generally noisy and biased (e.g. Cluzet et al., 2020). Satellite retrievals could be improved in the future (Kokhanovsky et al., 2019; Lamare et al., 2020), and Cluzet et al. (2020) showed that assimilating ratios of reflectance could be a workaround to tackle this issue. ~~However, the~~ In the near-infrared, the signal-to-noise ratio of

reflectances observations might be sufficient to constrain the surface microphysical properties (Durand and Margulis, 2007; Mary et al., 201~~~~, whereas the required accuracy for ~~reflectance retrieval~~ visible reflectance retrievals to remain informative on the snowpack ~~properties~~ light absorbing particles content is high (Warren, 2013), and it is yet to prove whether either approach can achieve this requirement.

480 Second, in this twin experiment framework, spatial patterns of the synthetic observations are likely compatible with the ensemble since they come from the same modelling system. This may not be the case in reality, therefore making it more difficult to assimilate, and we refer to this issue as model or ensemble realism.

We must assess the strengths and weaknesses of the *global* and *klocal* approaches facing those two issues. The *global* algorithm assumes that a global optimum can be found across the whole domain, e.g. the information from the different observations is

485 consistent and can be ingested in one block by the PF. With this strategy, the degeneracy due to the size of the observation vector is efficiently mitigated by the inflation algorithm as discussed in Sec. 5.1. The *klocal* approach considers that only a fraction of the observation information is relevant to constrain the model state at a given location. This algorithm tries to ingest as much information as possible while rejecting observations coming from too statistically different snowpack conditions. As a consequence, because we do not account for the real spatial patterns of observation errors, and because we work in a twin

490 experiment setup, a global optimum on the whole domain can exist and can be found by the *global* algorithm. This might be a reason why it beats the *klocal* approach (Figs. 6 and 7). In the real world, from the model point of view, there might be contradictory ~~informations~~ information among the observations that would be difficult to disentangle with a *global* strategy. The *klocal* algorithm could be more suited to this situation, because it is looking for local optima, based on the assumption that background ~~correlation patterns are realistic~~ correlations are a realistic representation of modelling errors.

495 These background correlation structures could be overestimated by the ensemble, and tests with real observations are necessary. Strong Band 4 correlations (Fig. 4a) might be due to the spatially homogeneous perturbations of LAP fluxes used to force the simulations (see Sec. 2.2.2), a key driver of this variable, and because the same snow model configuration is applied for a given member across the simulation domain. Several studies suggest that LAP fluxes vary with elevation and other topographic parameters (de Magalhães et al., 2019; Sabatier et al., 2020), but to date no reliable model ~~for that~~ of such processes exists in

500 complex terrain. In such a context, assuming uniform LAP forcing seems a reasonable compromise. Strong and almost uniform ~~correlations in HS~~ HS correlations (Fig. 4b) might be caused by the spatial homogeneity of precipitation perturbations and because we do not account for ~~snow transport by the wind~~ e.g. wind drift, intra massif heterogeneity of meteorological conditions and gravitational redistribution of snow (Wayand et al., 2018). Despite this semi-distributed framework ~~suffers~~ suffering from obvious limitations, ~~NWP models still suffer for large errors in mountainous areas, hampering~~ the potential for high-resolution

505 snowpack modelling ~~(Vionnet et al., 2016; Fiddes et al., 2019)~~ (Vionnet et al., 2020; Fiddes et al., 2019; Marsh et al., 2020) is hampered by large errors of the NWP models in mountainous areas (e.g. Nousu et al., 2019).

In the future, improving the ~~realism~~ ability of ensemble correlations to represent modelling errors could make the spreading of information an even more challenging task with the *klocal* algorithm. But there should remain significant potential for in-

510 formation propagation, as ~~results at a larger scale suggest~~ suggested by results at larger scales (Magnusson et al., 2014; Cantet

18

et al., 2019). The potential de-correlation of topographic classes would also impact the *global* algorithm. In a ~~non-observed~~ unobserved class, constraining the state of the snowpack with information from ~~area~~ areas that are not linked with it would likely degrade the forecasting skill, as suggested by the poor performance of the algorithms at low altitudes (Figs. 5 and 8). On the contrary, applying CrocO ~~into~~ over larger domains (e.g. distributed simulations or a collection of semi-distributed *mas-*

515 *sifs*), would probably see the *klocal* algorithm ~~take the best over the~~ outperform the *global*. The increased domain size would make it less plausible to find a global optimum over the domain, whereas spatial flexibility would be an asset of the *klocal* algorithm. Finally, in the case of modelled coupling between simulation points (e.g. snow drift), which was not the case here, the spatial discontinuities of the *klocal* analyses (see Sec. 1) might be a drawback compared to the global approach. Spatial discontinuities may reveal impractical for the interpretation of individual simulations outputs by snow forecasters too. The

520 *klocal* approach is likely to reduce these discontinuities compared to the *rlocal*, because similar locations will receive similar analyses (i.e. based on similar sets of observations). This issue could be partly mitigated by e.g. *state-block-domain* approaches (Farchi and Bocquet, 2018).

5.4 Outlook for ensemble modelling and data assimilation

525 In the snowpack modelling community, ensemble modelling appears as a powerful tool to represent modelling uncertainties (Vernay et al., 2015; Richter et al., 2020) and for data assimilation (Essery et al., 2013; Lafaysse et al., 2017; Piazzi et al., 2018; Aalstad et al., 2018). This study offers a novel approach to extract valuable information on the snowpack spatial behaviour from spatial correlation patterns of the ensemble. These patterns could be used to diagnose links between locations, transfer information between areas, or assess the representativeness of point simulations. More broadly, ensemble background

530 correlations have been exploited for long in the NWP and oceanographic communities to refine modelling errors representation which led to significant improvements in the DA systems (Evensen, 2003; Buehner, 2005).

Ensembles might open a way for the assimilation of point scale observations, or sparse remotely-sensed observations into spatialised simulations of the snowpack as suggested by Winstral et al. (2019) and the present work. For instance, there are numerous snow gauges and snow pit observations in the ski resorts of the French Alps. These data could be assimilated to

535 correct the ensemble in spatialised simulations (Winstral et al., 2019). The spatial pattern of assimilated observations in the experiments of Sec. 4 do not correspond to the real-life spatial coverage of this kind of observations. To give an insight of their potential, we also applied our methodology to assimilate only five synthetic HS observations with the *global* PF in the 1200 m to 2400 m flat classes. The results are shown in Fig. 9. The assimilation improves the performance in all aspects and slopes. Naturally, this suffers from the same limitation as discussed in Sec. 5.3, not to mention the limited spatial representativeness

540 of in situ observations but it shows some potential for this idea.

In that way, a more rational use of the available observations could be implemented towards a new ensemble data assimilation system. In the present CrocO system, SAFRAN reanalysis are only assimilating weather station information (precipitation phase, temperature, wind), and makes no use of the numerous snow observations available. Here, snow observations are assimilated by the PF, but are not used to correct meteorological forcings (only snow variables, see Fig. 2). In ~~the way of~~ a new

19

ensemble data assimilation system, within CrocO, the SAFRAN meteorological analysis could be bypassed, the PF operating directly both on the meteorological and snowpack variables in a more comprehensive and coupled strategy.

6   Conclusions

In this study, we introduced CrocO, a new ensemble data assimilation system able to reduce the errors of a spatialised snowpack model in locations that are not observed. The ensemble is built by a combination of meteorological and multi-physical ensembles to represent modelling uncertainties. A Particle Filter assimilates observations of HS and Reflectance. We developed two variants of the PF using inflation or k-localisation, in order to spread the information from partial observations of the system, without degeneracy of the PF. ~~We~~ In the framework of synthetic experiments, we have shown in particular that:

1. ~~these~~ These variants are able to ingest numerous observations without degeneracy;

2. ~~an~~ An efficient spreading of the observations information towards the ~~non observed~~ unobserved areas is achieved with the *global* and *klocal* approaches;

3. ~~reflectance~~ Reflectance assimilation leads to an overall 20% improvement in CRPS and 60% in reliability.

We suggest that this approach could be used in any spatialised framework to assimilate sparse observations from e.g. networks of in-situ snowpack observations. Beyond the snowpack modelling community, the inflation and k-localisation strategies could help address the problem of partially observed systems. This work is also a first step towards the operational assimilation of reflectance in a semi-distributed context. To reach that goal, biases of reflectance retrievals should be studied, and observation ~~errors~~ error structures duly quantified. Snow ~~Cover Fraction~~ cover fraction would be a good companion variable to ~~assimilate at lower elevations~~ jointly assimilate with reflectances, requiring the use of an appropriate observation operator. Extending the simulation domain to several massifs would allow the exchange of information between neighbouring massifs with the *klocal* algorithm.

Code availability. The Crocus snowpack model (including all physical options of the ESCROC system) and the Particle Filter algorithm are developed inside the opensource SURFEX project. The source files of SURFEX code are provided at https://doi.org/10.5281/zenodo.3774861 to guarantee the permanent reproductibility of results. However, we recommend potential future users and developers to access to the code from its git repository (git.umr-cnrm.fr/git/Surfex_Git2.git) to benefit from all tools of code management (history management, bug fixes, documentation, interface for technical support, etc.). This needs a quick registration, the procedure is described at https://opensource. cnrm-game-meteo.fr/projects/snowtools/wiki/Procedure_for_new_users. The version used in this work is tagged as CrocO_v1.0.

A python software called CrocO_toolbox was specifically developed, in order to pre-post process and launch CrocO experiments. It is available on Github (https://github.com/bertrandcz/CrocO, release v1.0) along with a documentation.

575 The article version of CrocO_toolbox is archived at: https://doi.org/10.5281/zenodo.3784980. This software strongly relies on two external python projects ensuring the files management between the different steps of a simulation and the interface with Meteo-France HPC system (including parallelization and data storage): snowtools and vortex. Their sources are available at https://doi.org/10.5281/zenodo.3774861 (same archive as SURFEX) to guarantee the permanent reproducibility of results. However, as for the SURFEX project and for the same reasons, it is recommended to access snowtools code from its git repository (git.umr-cnrm.fr/git/snowtools_git.git). The version used in this

580 work is also tagged as CrocO_v1.0. The vortex project gathers all environment-specific codes of Météo-France modelling systems relative to its HPC computing system. For this project, only the sources which are specific to this article simulations are provided. The common objects inheritance is based on Vortex version 1.6.1. The version used in this work is also tagged as CrocO_v1.0 in the vortex git repository.

Because these software could not be applied outside Météo-France HPC environment, CrocO python software offers the possibility to run
585 CrocO simulations locally. This functionality was not used here due to the high numerical cost of our simulations, which required the use of Météo-France HPC environment.

Data availability. Input and output data necessary to reproduce the manuscript simulations and figures are provided at https://doi.org/10.5281/zenodo.3775007. This archive includes : SAFRAN reanalyses, (also available at https://doi.org/10.25326/37), MOCAGE forcings, namelists, configuration files and spinup files necessary to reproduce the simulations. Raw model outputs can be provided on re-
590 quest but since they amount up to 500+ Gigabytes, only post-processed simulations outputs are provided in this archive, along with scores and scripts to reproduce the manuscript figures.

Appendix A: Stochastic perturbations of the forcings

The stochastic perturbation procedure of the forcings is introduced in Sec. 2.2.2 and is identical to Charrois et al. (2016) for the meteorological parameters and Cluzet et al. (2020) for the light absorbing particles (LAP) fluxes. For a given date and forcing
595 variable, perturbation values are the same for all the points in space (no spatial auto-correlation is considered), as SAFRAN semi-distributed massifs have a limited spatial extent (about $1000 \, \text{km}^2$). Precipitation, incoming radiations, wind speed and air temperature from SAFRAN are perturbed with temporally autocorrelated stochastic parameters. The precipitation, incoming shortwave radiation, and wind speed are perturbed with a multiplicative noise. Longwave radiation and air temperature are perturbed with an additive noise.
600 For meteorological variables, the perturbation vector V is built as follows:

$$V(t) = \phi V(t-1) + \varepsilon(t) \tag{A1}$$

Where $\phi = e^{-dt/\tau}$, with dt the forcing timestep, τ the decorrelation time (in h) and ε a normal law of mean 0 and variance $\sigma^2(1-\phi^2)$. Parameter values for each variable are described in Tab. A1. The significantly high auto-correlation time of precipitation $1500 \, \text{h}$ was tuned to roughly adjust the ensemble spread to the observed intra-massif variability of yearly-cumulated
605 precipitation. Note that the precipitation phase is adjusted with the perturbed air temperature to ensure a physical consistency.

21

Further details on the procedure can be found in Charrois et al. (2016).

Regarding LAP fluxes, dry and wet black carbon and mineral dust deposition fluxes from MOCAGE are perturbed with a random factor which keeps constant throughout the year. Each member has a single multiplicative factor following a log-normal law of mean μ and variance σ (see Tab. A2). The mean of black carbon random perturbations was adjusted based on comparisons between simulations and field observations at col du Lautaret, a mountain pass within the considered SAFRAN massif.

Appendix B: Complements on the implementation

B1 Technical implementation and code performance

CrocO is implemented within Météo-France HPC (High Performance Computing) environment, enabling to fully parallelize the ensemble (one core per member), and bridge the gap with operational applications (Lafaysse et al., 2013; Morin et al., 2020). This implementation is strongly parallel. As an example, the execution time of a one-year assimilation run of 187 model points with 160 members on 4 nodes of 40 cores each lasts for only two hours. The PF is a lightweight algorithm, most of the computational burden owing to the propagation of the ensemble and input/output. Note also that no significant difference in execution time can be noted between the different PF algorithms.

B2 PF sample reordering

As mentioned in Sec. 2.3, a reordering step was implemented after the PF resampling from Kitagawa (1996), for practical reasons.

- (3) from s, build $\tilde{s}$ such that all elements of the unique values of s lie in the position given by their value. Example with 16 particles:

$$s = [1,1,2,3,3,3,8,8,9,9,9,9,9,16,16,16] \Rightarrow \tilde{s} = [1,2,3,1,3,3,8,8,9,9,9,9,16,16,16]$$

Indeed, I/O represents a bottleneck in the PF. When building the analysis $\widehat{\mathbf{X}^a}$, the background $\widehat{\mathbf{X}^b}$ is already loaded in memory. Since $\widehat{\mathbf{X}^a}$ is just a reordering of $\widehat{\mathbf{X}^b}$ columns based on s, a reordering of s avoids to build a copy of $\widehat{\mathbf{X}^b}$. This way, $\widehat{\mathbf{X}^a}$ is built by an online modification of $\widehat{\mathbf{X}^b}$ using two pointers. Reordering is a growing consideration in the PF community (Farchi and Bocquet, 2018).

References

Aalstad, K., Westermann, S., Schuler, T. V., Boike, J., and Bertino, L.: Ensemble-based assimilation of fractional snow-covered area satellite retrievals to estimate the snow distribution at Arctic sites, The Cryosphere, 12, https://doi.org/10.5194/tc-12-247-2018, https://doi.org/10.5194%2Ftc-12-247-2018, 2018.

Aalstad, K., Westermann, S., and Bertino, L.: Evaluating satellite retrieved fractional snow-covered area at a high-Arctic site using terrestrial photography, Remote sensing of Environment, 239, 111 618, 2020.

Albergel, C., Munier, S., Leroux, D. J., Dewaele, H., Fairbairn, D., Barbu, A. L., Gelati, E., Dorigo, W., Faroux, S., Meurey, C., et al.: Sequential assimilation of satellite-derived vegetation and soil moisture products using SURFEX_v8. 0: LDAS-Monde assessment over the Euro-Mediterranean area, Geoscientific Model Development, 10, 3889–3912, 2017.

Alonso-Gónzalez, E., Gutmann, E., Aalstad, K., Fayad, A., and Gascoin, S.: Snowpack dynamics in the Lebanese mountains from quasi-dynamically downscaled ERA5 reanalysis updated by assimilating remotely-sensed fractional snow-covered area, Hydrology and Earth System Sciences Discussions, pp. 1–31, 2020.

Baba, M., Gascoin, S., and Hanich, L.: Assimilation of Sentinel-2 Data into a Snowpack Model in the High Atlas of Morocco, Remote Sensing, 10, 1982, https://doi.org/10.3390/rs10121982, https://doi.org/10.3390%2Frs10121982, 2018.

Bengtsson, T., Bickel, P., Li, B., et al.: Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems, in: Probability and statistics: Essays in honor of David A. Freedman, pp. 316–334, Institute of Mathematical Statistics, 2008.

Buehner, M.: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting, Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 131, 1013–1043, 2005.

Candille, G., Brankart, J.-M., and Brasseur, P.: Assessment of an ensemble system that assimilates Jason-1/Envisat altimeter data in a probabilistic model of the North Atlantic ocean circulation., Ocean Science, 11, 425–438, 2015.

Cantet, P., Boucher, M., Lachance-Coutier, S., Turcotte, R., and Fortin, V.: Using a particle filter to estimate the spatial distribution of the snowpack water equivalent, Journal of Hydrometeorology, 20, 577–594, 2019.

Carmagnola, C. M., Morin, S., Lafaysse, M., Domine, F., Lesaffre, B., Lejeune, Y., Picard, G., and Arnaud, L.: Implementation and evaluation of prognostic representations of the optical diameter of snow in the SURFEX/ISBA-Crocus detailed snowpack model, The Cryosphere, 8, 417–437, https://doi.org/10.5194/tc-8-417-2014, 2014.

Charrois, L., Cosme, E., Dumont, M., Lafaysse, M., Morin, S., Libois, Q., and Picard, G.: On the assimilation of optical reflectances and snow depth observations into a detailed snowpack model, The Cryosphere, 10, 1021–1038, https://doi.org/10.5194/tc-10-1021-2016, 2016.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., et al.: A unified approach for process-based hydrologic modeling: 1. Modeling concept, Water Resources Research, 51, 2498–2514, 2015.

Cluzet, B., Revuelto, J., Lafaysse, M., Tuzet, F., Cosme, E., Picard, G., Arnaud, L., and Dumont, M.: Towards the assimilation of satellite reflectance into semi-distributed ensemble snowpack simulations, Cold Regions Science and Technology, 170, 102 918, 2020.

Davaze, L., Rabatel, A., Arnaud, Y., Sirguey, P., Six, D., Letreguilly, A., and Dumont, M.: Monitoring glacier albedo as a proxy to derive summer and annual surface mass balances from optical remote-sensing data, The Cryosphere, 12, 271–286, https://doi.org/10.5194/tc-12-271-2018, https://www.the-cryosphere.net/12/271/2018/, 2018.

670   De Lannoy, G. J., Reichle, R. H., Arsenault, K. R., Houser, P. R., Kumar, S., Verhoest, N. E., and Pauwels, V. R.: Multiscale assimilation of Advanced Microwave Scanning Radiometer–EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado, Water Resources Research, 48, 2012.

de Magalhães, N., Evangelista, H., Condom, T., Rabatel, A., and Ginot, P.: Amazonian Biomass Burning Enhances Tropical Andean Glaciers Melting, Scientific reports, 9, 1–12, 2019.

675   Dechant, C. and Moradkhani, H.: Radiance data assimilation for operational snow and streamflow forecasting, Adv. Water Resour., 34, 351–364, https://doi.org/10.1016/j.advwatres.2010.12.009, 2011.

Deschamps-Berger, C., Gascoin, S., Berthier, E., Deems, J., Gutmann, E., Dehecq, A., Shean, D., and Dumont, M.: Snow depth mapping from stereo satellite imagery in mountainous terrain: evaluation using airborne laser-scanning data, The Cryosphere, 14, 2925–2940, https://doi.org/10.5194/tc-14-2925-2020, https://tc.copernicus.org/articles/14/2925/2020/, 2020.

680   Doucet, A., De Freitas, N., and Gordon, N.: An introduction to sequential Monte Carlo methods, in: Sequential Monte Carlo methods in practice, pp. 472–474, Springer, 2001.

Dozier, J., Green, R. O., Nolin, A. W., and Painter, T. H.: Interpretation of snow properties from imaging spectrometry, Remote Sensing of Environment, 113, S25–S37, 2009.

Durand, M. and Margulis, S. A.: Feasibility test of multifrequency radiometric data assimilation to estimate snow water equivalent, Journal
685   of Hydrometeorology, 7, 443–457, 2006.

Durand, M. and Margulis, S. A.: Correcting first-order errors in snow water equivalent estimates using a multifrequency, multiscale radiometric data assimilation scheme, Journal of Geophysical Research: Atmospheres, 112, 2007.

Durand, Y., Brun, E., Mérindol, L., Guyomarc'h, G., Lesaffre, B., and Martin, E.: A meteorological estimation of relevant parameters for snow models, Ann. Glaciol., 18, 65–71, 1993.

690   Eberhard, L. A., Sirguey, P., Miller, A., Marty, M., Schindler, K., Stoffel, A., and Bühler, Y.: Intercomparison of photogrammetric platforms for spatially continuous snow depth mapping, The Cryosphere Discussions, pp. 1–40, 2020.

Essery, R., Morin, S., Lejeune, Y., and Bauduin-Ménard, C.: A comparison of 1701 snow models using observations from an alpine site, Adv. Water Res., 55, 131–148, https://doi.org/10.1016/j.advwatres.2012.07.013, 2013.

Evensen, G.: The ensemble Kalman filter: Theoretical formulation and practical implementation, Ocean dynamics, 53, 343–367, 2003.

695   Farchi, A. and Bocquet, M.: Comparison of local particle filters and new implementations, Nonlinear Processes in Geophysics, 25, 765–807, 2018.

Fiddes, J., Aalstad, K., and Westermann, S.: Hyper-resolution ensemble-based snow reanalysis in mountain regions using clustering, Hydrology and Earth System Sciences, 23, 4717–4736, 2019.

Gordon, N. J., Salmond, D. J., and Smith, A. F.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation, in: IEE Proceedings
700   F (Radar and Signal Processing), vol. 140, pp. 107–113, IET, 1993.

Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, Monthly Weather Review, 129, 2776–2790, 2001.

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

705   Houtekamer, P. L. and Mitchell, H. L.: Data assimilation using an ensemble Kalman filter technique, Monthly Weather Review, 126, 796–811, 1998.

Josse, B., Simon, P., and Peuch, V.-H.: Radon global simulations with the multiscale chemistry and transport model MOCAGE, Tellus B: Chemical and Physical Meteorology, 56, 339–356, 2004.

Kitagawa, G.: Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, Journal of computational and graphical statistics, 5, 1–25, 1996.

Kokhanovsky, A., Lamare, M., Danne, O., Brockmann, C., Dumont, M., Picard, G., Arnaud, L., Favier, V., Jourdain, B., Le Meur, E., et al.: Retrieval of snow properties from the Sentinel-3 Ocean and Land Colour Instrument, Remote Sensing, 11, 2280, 2019.

Krinner, G., Derksen, C., Essery, R., Flanner, M., Hagemann, S., Clark, M., Hall, A., Rott, H., Brutel-Vuilmet, C., Kim, H., et al.: ESM-SnowMIP: assessing snow models and quantifying snow-related climate feedbacks, Geoscientific Model Development, 11, 5027–5049, 2018.

Lafaysse, M., Morin, S., Coléou, C., Vernay, M., Serça, D., Besson, F., Willemet, J.-M., Giraud, G., and Durand, Y.: Toward a new chain of models for avalanche hazard forecasting in French mountain ranges, including low altitude mountains, in: Proceedings of the International Snow Science Workshop - Grenoble and Chamonix, pp. 162–166, 2013.

Lafaysse, M., Cluzet, B., Dumont, M., Lejeune, Y., Vionnet, V., and Morin, S.: A multiphysical ensemble system of numerical snow modelling, The Cryosphere, 11, 1173–1198, https://doi.org/10.5194/tc-11-1173-2017, https://www.the-cryosphere.net/11/1173/2017/, 2017.

Lahoz, B. K. W. and Menard, R.: Data assimilation, Springer, 2010.

Lamare, M., Dumont, M., Picard, G., Larue, F., Tuzet, F., Delcourt, C., and Arnaud, L.: Simulating Optical Top-Of-Atmosphere Radiance Satellite Images over Snow-Covered Rugged Terrain, The Cryosphere Discussions, pp. 1–46, 2020.

Largeron, C., Dumont, M., Morin, S., Boone, A., Lafaysse, M., Metref, S., Cosme, E., Jonas, T., Winstral, A., and Margulis, S. A.: Towards snow cover estimation in mountainous areas using modern data assimilation methods: A review, Frontiers in Earth Science, 8, 325, 2020.

Larue, F., Royer, A., Sève, D. D., Roy, A., and Cosme, E.: Assimilation of passive microwave AMSR-2 satellite observations in a snowpack evolution model over northeastern Canada, Hydrology and Earth System Sciences, 22, 5711–5734, 2018.

Libois, Q., Picard, G., France, J., Arnaud, L., Dumont, M., Carmagnola, C. M., and King, M.: Influence of grain shape on light penetration in snow, The Cryosphere, 7, 1803–1818, https://doi.org/10.5194/tc-7-1803-2013, 2013.

Libois, Q., Picard, G., Arnaud, L., Dumont, M., Lafaysse, M., Morin, S., and Lefebvre, E.: Summertime evolution of snow specific surface area close to the surface on the Antarctic Plateau, The Cryosphere, 9, 2383–2398, https://doi.org/10.5194/tc-9-2383-2015, 2015.

Liu, J. S. and Chen, R.: Blind deconvolution via sequential imputations, Journal of the american statistical association, 90, 567–576, 1995.

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, 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, 1125–1143, 2017.

Mankin, J. S., Viviroli, D., Singh, D., Hoekstra, A. Y., and Diffenbaugh, N. S.: The potential for snow to supply human water demand in the present and future, Environmental Research Letters, 10, 114 016, 2015.

Margulis, A., Cortés, G., Girotto, M., and Durand, M.: A Landsat-Era Sierra Nevada Snow Reanalysis (1985–2015), Journal of Hydrometeorology, 17, 1203–1221, https://doi.org/0.1175/JHM-D-15-0177.1, 2016.

Margulis, S. A., Fang, Y., Li, D., Lettenmaier, D. P., and Andreadis, K.: The Utility of Infrequent Snow Depth Images for Deriving Continuous Space-Time Estimates of Seasonal Snow Water Equivalent, Geophysical Research Letters, 46, 5331–5340, 2019.

Marsh, C. B., Pomeroy, J. W., and Wheater, H. S.: The Canadian Hydrological Model (CHM) v1. 0: a multi-scale, multi-extent, variable-complexity hydrological model-design and overview., Geoscientific Model Development, 13, 225–225, 2020.

745 Mary, A., Dumont, M., Dedieu, J.-P., Durand, Y., Sirguey, P., Milhem, H., Mestre, O., Negi, H. S., Kokhanovsky, A. A., Lafaysse, M., and Morin, S.: Intercomparison of retrieval algorithms for the specific surface area of snow from near-infrared satellite data in mountainous terrain, and comparison with the output of a semi-distributed snowpack model, The Cryosphere, 7, 741–761, https://doi.org/10.5194/tc-7-741-2013, 2013.

Masson, T., Dumont, M., Mura, M. D., Sirguey, P., Gascoin, S., Dedieu, J.-P., and Chanussot, J.: An Assessment of Existing Methodologies
750 to Retrieve Snow Cover Fraction from MODIS Data, Remote Sensing, 10, 619, 2018.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gibelin, A.-L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J.-F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V., and Voldoire, A.: The
755 SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of Earth surface variables and fluxes, Geoscientific Model Development, 6, 929–960, https://doi.org/10.5194/gmd-6-929-2013, 2013.

Matheson, J. E. and Winkler, R. L.: Scoring rules for continuous probability distributions, Management science, 22, 1087–1096, 1976.

Morin, S., Horton, S., Techel, F., Bavay, M., Coléou, C., Fierz, C., Gobiet, A., Hagenmuller, P., Lafaysse, M., Ližar, M., Mitterer, C., Monti, F., Müller, K., Olefs, M., Snook, J. S., van Herwijnen, A., and Vionnet, V.: Application of physical snowpack models in support of
760 operational avalanche hazard forecasting: A status report on current implementations and prospects for the future, Cold Regions Science and Technology, 170, 102 910, https://doi.org/https://doi.org/10.1016/j.coldregions.2019.102910, http://www.sciencedirect.com/science/article/pii/S0165232X19302071, 2020.

Nousu, J.-P., Lafaysse, M., Vernay, M., Bellier, J., Evin, G., and Joly, B.: Statistical post-processing of ensemble forecasts of the height of new snow, Nonlinear Processes in Geophysics, 26, 339–357, 2019.

765 Penny, S. G. and Miyoshi, T.: A local particle filter for high-dimensional geophysical systems, Nonlinear Processes in Geophysics, 23, 391, 2016.

Piazzi, G., Thirel, G., Campo, L., and Gabellani, S.: A particle filter scheme for multivariate data assimilation into a point-scale snowpack model in an Alpine environment, The Cryosphere, 12, 2287–2306, 2018.

Pomeroy, J. W., Fang, X., and Marks, D. G.: The cold rain-on-snow event of June 2013 in the Canadian Rockies—Characteristics and
770 diagnosis, Hydrological Processes, 30, 2899–2914, 2016.

Poterjoy, J.: A localized particle filter for high-dimensional nonlinear systems, Monthly Weather Review, 144, 59–76, 2016.

Poterjoy, J. and Anderson, J. L.: Efficient assimilation of simulated observations in a high-dimensional geophysical system using a localized particle filter, Monthly Weather Review, 144, 2007–2020, 2016.

Poterjoy, J., Wicker, L., and Buehner, M.: Progress toward the application of a localized particle filter for numerical weather prediction,
775 Monthly Weather Review, 147, 1107–1126, 2019.

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, Hydrol. Earth Syst. Sci., 19, 3153–3179, https://doi.org/10.5194/hess-19-3153-2015, 2015.

Reichle, R. H. and Koster, R. D.: Assessing the impact of horizontal error correlations in background fields on soil moisture estimation, Journal of Hydrometeorology, 4, 1229–1242, 2003.

780 Revuelto, J., Lecourt, G., Lafaysse, M., Zin, I., Charrois, L., Vionnet, V., Dumont, M., Rabatel, A., Six, D., Condom, T., et al.: Multi-Criteria Evaluation of Snowpack Simulations in Complex Alpine Terrain Using Satellite and In Situ Observations, Remote Sensing, 10, 1171, 2018.

Revuelto, J., Cluzet, B., Duran, N., Fructus, M., Lafaysse, M., Cosme, E., and Dumont, M.: Assimilation of surface reflectance in snow simulations; impact on bulk snow variables, in prep.

785 Richter, B., van Herwijnen, A., Rotach, M. W., and Schweizer, J.: Sensitivity of modeled snow stability data to meteorological input uncertainty, Natural Hazards and Earth System Sciences Discussions, 2020, 1–25, https://doi.org/10.5194/nhess-2019-433, https://www.nat-hazards-earth-syst-sci-discuss.net/nhess-2019-433/, 2020.

Sabatier, T., Largeron, Y., Paci, A., Lac, C., Rodier, Q., Canut, G., and Masson, V.: Semi-idealized simulations of wintertime flows and pollutant transport in an alpine valley: Passive tracer tracking (Part II), Quarterly Journal of the Royal Meteorological Society, n/a,

790 https://doi.org/10.1002/qj.3710, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3710, 2020.

Shaw, T. E., Gascoin, S., Mendoza, P. A., Pellicciotti, F., and McPhee, J.: Snow depth patterns in a high mountain Andean catchment from satellite optical tri-stereoscopic remote sensing, Water Resources Research, 2019.

Sirguey, P., Mathieu, R., and Arnaud, Y.: Subpixel monitoring of the seasonal snow cover with MODIS at 250m spatial resolution in the Southern Alps of New Zealand: methodology and accuracy assessment, Remote Sens. Environ., 113, 160–181,

795 https://doi.org/10.1016/j.rse.2008.09.008, 2009.

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

Snyder, C., Bengtsson, T., Bickel, P., and Anderson, J.: Obstacles to high-dimensional particle filtering, Monthly Weather Review, 136, 4629–4640, 2008.

Stigter, E. E., Wanders, N., Saloranta, T. M., Shea, J. M., Bierkens, M. F., and Immerzeel, W. W.: Assimilation of snow cover and snow depth

800 into a snow model to estimate snow water equivalent and snowmelt runoff in a Himalayan catchment, The Cryosphere, 11, 1647–1664, 2017.

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, 5825–5850, 2013.

Tuzet, F., Dumont, M., Lafaysse, M., Picard, G., Laurent, A., Voisin, D., Lejeune, Y., Charrois, L., Nabat, P., and Morin, S.: A multilayer

805 physically based snowpack model simulating direct and indirect radiative impacts of light-absorbing impurities in snow, The Cryosphere, 11, 2633, 2017.

Van Leeuwen, P. J.: Particle filtering in geophysical systems, Monthly Weather Review, 137, 4089–4114, 2009.

Van Leeuwen, P. J., Künsch, H. R., Nerger, L., Potthast, R., and Reich, S.: Particle filters for high-dimensional geoscience applications: A review, Quarterly Journal of the Royal Meteorological Society, 145, 2335–2365, 2019.

810 Vernay, M., Lafaysse, M., Merindol, L., Giraud, G., and Morin, S.: Ensemble Forecasting of snowpack conditions and avalanche hazard, Cold. Reg. Sci. Technol., 120, 251–262, https://doi.org/10.1016/j.coldregions.2015.04.010, 2015.

Vetterling, W. T., Teukolsky, S. A., Press, W. H., and Flannery, B. P.: Numerical recipes in Fortran the art of scientific computing, Cambridge University Press, 1992.

Veyssière, G., Karbou, F., Morin, S., Lafaysse, M., and Vionnet, V.: Evaluation of Sub-Kilometric Numerical Simulations of C-Band Radar

815 Backscatter over the French Alps against Sentinel-1 Observations, Remote Sensing, 11, 8, 2019.

Vionnet, V., Brun, E., Morin, S., Boone, A., Martin, E., Faroux, S., Le-Moigne, P., and Willemet, J.-M.: The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2, Geosci. Model. Dev., 5, 773–791, https://doi.org/10.5194/gmd-5-773-2012, 2012.

Vionnet, V., Marsh, C. B., Menounos, B., Gascoin, S., Wayand, N. E., Shea, J., Mukherjee, K., and Pomeroy, J. W.: Multi-scale snowdrift-permitting modelling of mountain snowpack, The Cryosphere Discussions, pp. 1–43, 2020.

28

820 Vionnet, V., Dombrowski-Etchevers, I., Lafaysse, M., Quéno, L., Seity, Y., and Bazile, E.: Numerical weather forecasts at kilometer scale in the French Alps : evaluation and applications for snowpack modelling, J. Hydrometeor., https://doi.org/http://dx.doi.org/10.1175/JHM-D-15-0241.1, 2016.

Warren, S. G.: Can black carbon in snow be detected by remote sensing?, J. Geophys. Res., 118, 779–786, https://doi.org/10.1029/2012JD018476, 2013.

825 Wayand, N. E., Marsh, C. B., Shea, J. M., and Pomeroy, J. W.: Globally scalable alpine snow metrics, Remote Sensing of Environment, 213, 61–72, 2018.

Winstral, A., Magnusson, J., Schirmer, M., and Jonas, T.: The Bias-Detecting Ensemble: A New and Efficient Technique for Dynamically Incorporating Observations Into Physics-Based, Multilayer Snow Models, Water Resources Research, 55, 613–631, 2019.

Wright, P., Bergin, M., Dibb, J., Lefer, B., Domine, F., Carman, T., Carmagnola, C., Dumont, M., Courville, Z., Schaaf, C., et al.: Comparing
830 MODIS daily snow albedo to spectral albedo field measurements in Central Greenland, Remote sensing of environment, 140, 118–129, 2014.

Würzer, S., Jonas, T., Wever, N., and Lehning, M.: Influence of initial snowpack properties on runoff formation during rain-on-snow events, Journal of hydrometeorology, 17, 1801–1815, 2016.
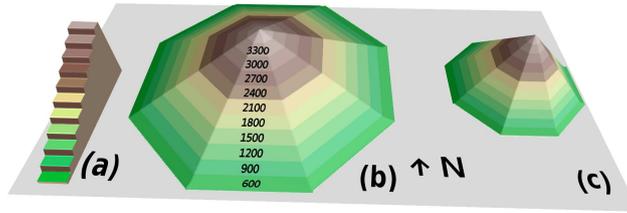
Figure 1. 3D schematic view of the semi-distributed geometry, where the numbers represent the ~~altitudes of the~~ elevation bands ~~altitudes~~ (~~m~~in m). From left to right, the three different mountains represent the flat, 20° and 40° ~~degrees~~ slopes.
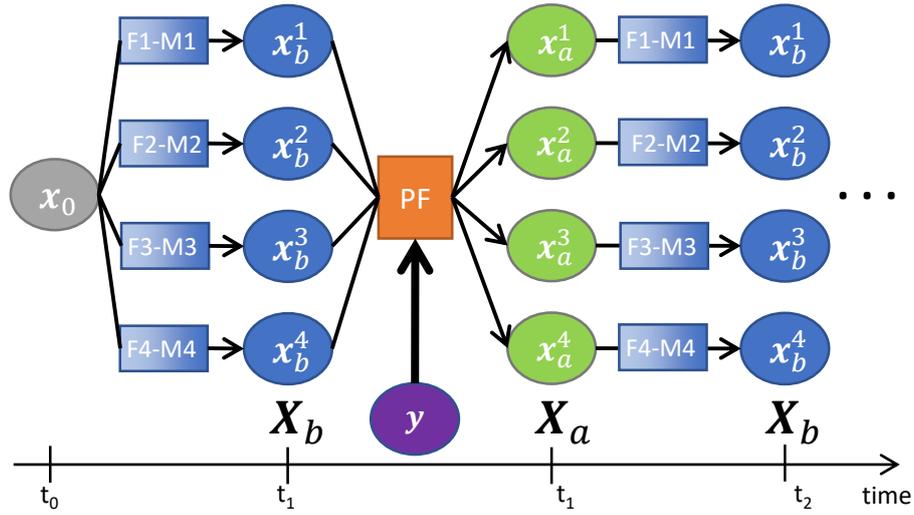
Figure 2. Workflow of CrocO ensemble data assimilation system with 4 members. $\widehat{x_0}$: initial state at time t_0, Fi: forcing, Mi: ESCROC member, $\widehat{X_b}$: background state, $\widehat{x_b^i}$: background particles, $\widehat{X_a}$: analysis, $\widehat{x_a^i}$: analysis particles, y: observation, t_1 and t_2: observation dates.

PF Algorithm	N_e	inflation	N_{eff}^*	HS σ_o^2 (m^2)
rlocal	40	on	7	1.0×10^{-2}
global	40	on	7	1.0×10^{-2}
klocal	40	on (if k=1)	7	5.0×10^{-2}

Table 1. ~~setup of~~ Setup for the ~~Snow depth~~ height of snow assimilation experiment.

PF Algorithm	N_e	inflation	N_{eff}^*	B4 σ_o^2	B5 σ_o^2
rlocal	40	on	7	5.6×10^{-4}	2.0×10^{-3}
global	40	on	7	5.6×10^{-4}	2.0×10^{-3}
klocal	40	on (if k=1)	7	2.8×10^{-3}	1.0×10^{-2}

Table 2. setup ~~of~~ for the ~~second~~ first reflectance assimilation experiment.

PF Algorithm	N_e	inflation	N_{eff}^*	B4 σ_o^2	B5 σ_o^2
rlocal	160	on	25	5.6×10^{-4}	2.0×10^{-3}
global	160	on	25	5.6×10^{-4}	2.0×10^{-3}
klocal	160	on (if k=1)	25	2.8×10^{-3}	1.0×10^{-2}

Table 3. ~~setup of~~ Setup for the second reflectance assimilation experiment.
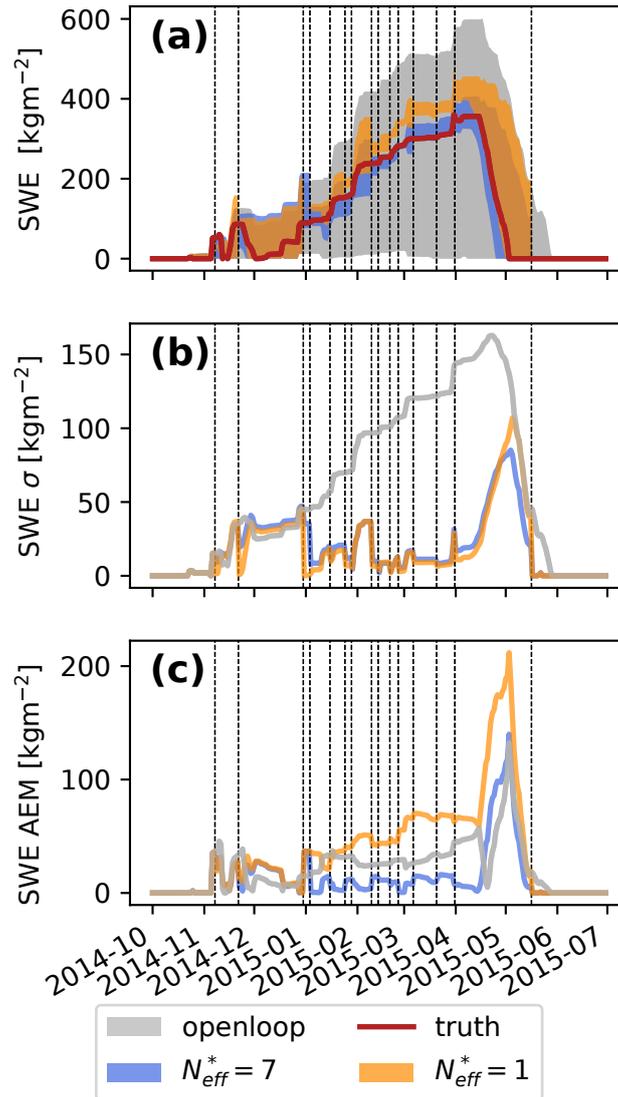
Figure 3. Impact of the inflation ($N^*_{\text{eeff}} = 7$ $N^*_{\text{eff}} = 7$) versus no inflation ($N^*_{\text{eff}} = 1$) in the 1800_N_40 topographic class (not observed), when assimilating HS of 2015_q80 with the *global* PF. (a) SWE minimum-maximum envelopes as a function of time, (b) spread and (c) RMSE AEM. Dashed lines represent the assimilation dates.
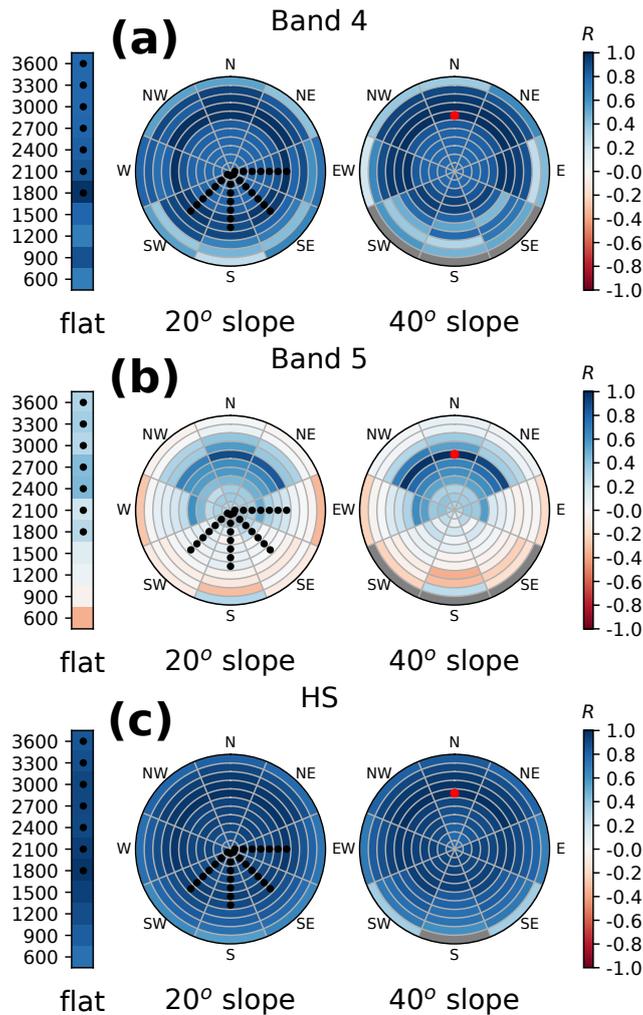
Figure 4. 2015, February 20th ~~openloop~~ open-loop (40 members) Pearson correlations between the domain points and the 1800_N_40 topographic class (red dot), in Band 4 (a), Band 5 (b) and HS (c). Left bars show the flat topographic classes in the associated elevation bands, while pie plots show the 20° and 40° slope topographic classes, as depicted in Fig. 1. Black dots ~~the~~ denote the observed classes.
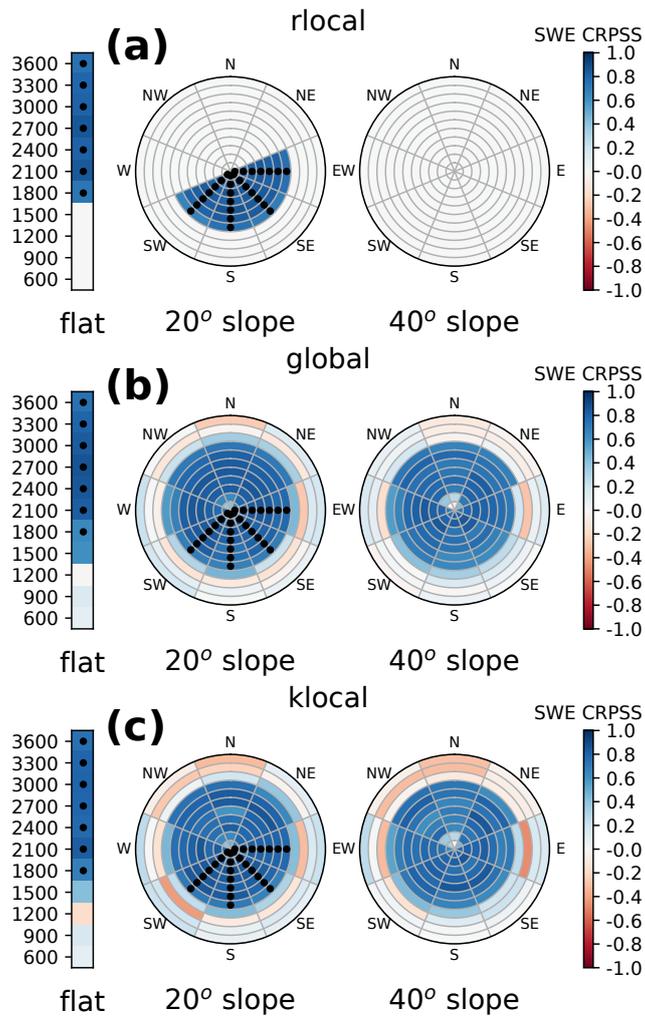
Figure 5. CRPS ~~Skill Score~~ skill score of SWE for the *rlocal* (a), *global* (b) and *klocal* (c) algorithms assimilating the HS of 2013_p20 synthetic ~~member~~observation scenario. The score is computed ~~on~~ for the whole snow season for each topographic class. Black dots denote the observed classes.
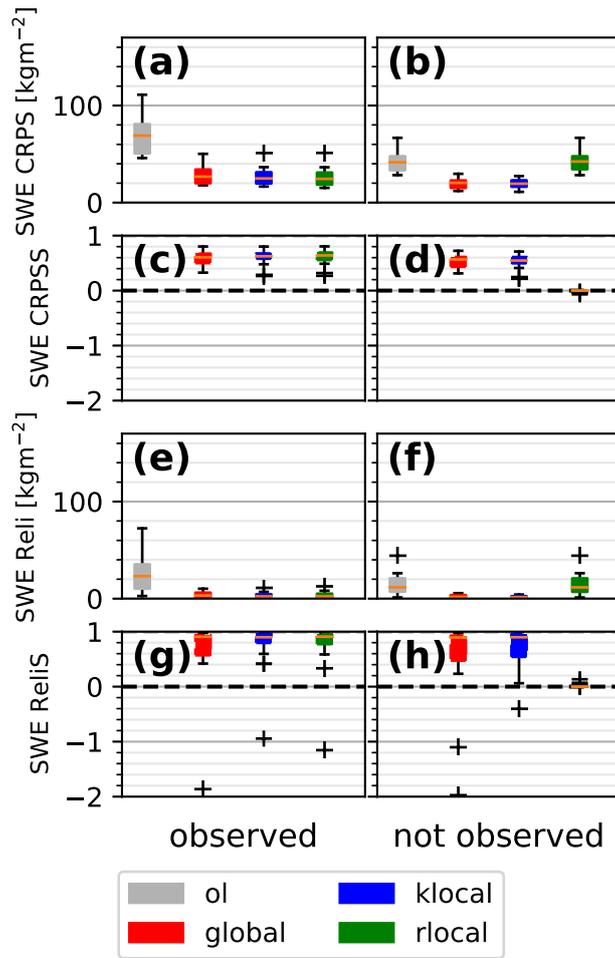
Figure 6. Boxplots of SWE CRPS (a,b) and Reli (e,f) for the different algorithms for the 16 different synthetic ~~members~~observation scenarios, separated between observed (left column) and not observed (right panel) classes. Panels (c,d) and (g,h) show the associated ~~Skill Scores~~skill scores.
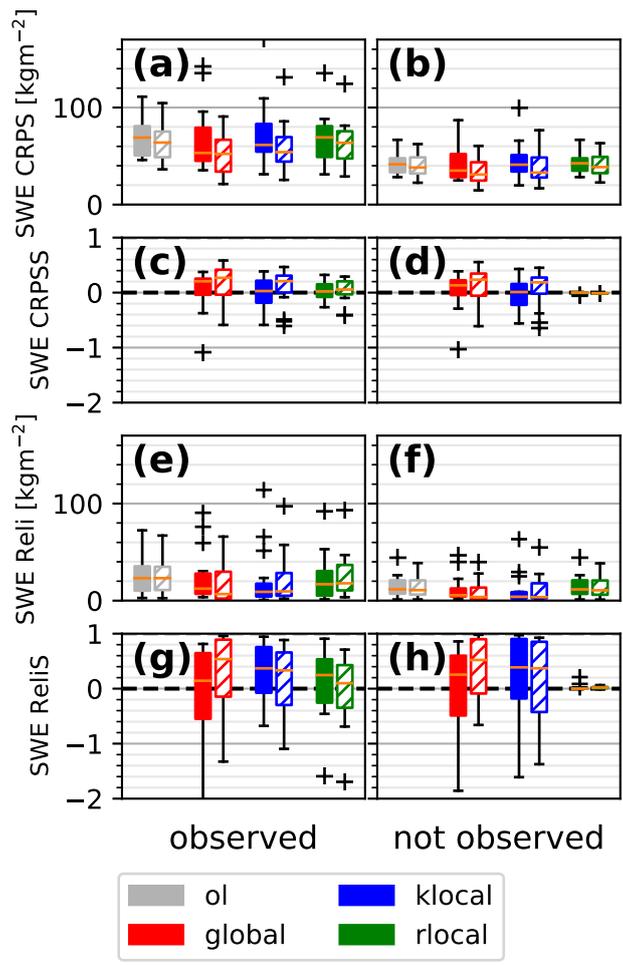
Figure 7. Same as Fig. 6 for reflectance with 40 members (filled) and 160 members (hatched).
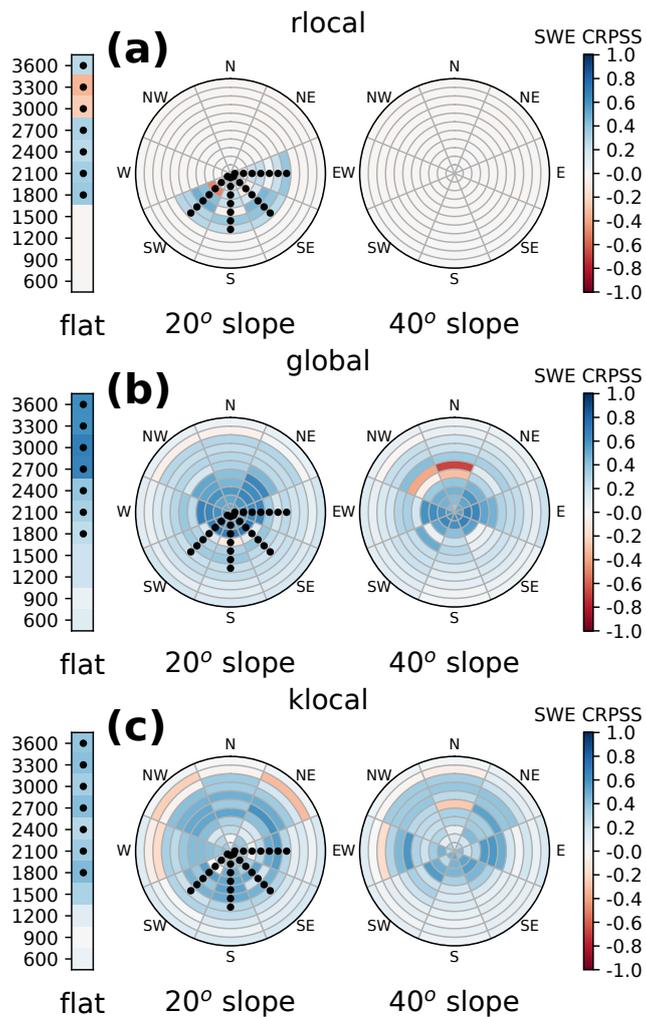
Figure 8. Same as Fig. 5 for the assimilation of the reflectance of 2016_p60 ~~member~~synthetic observation scenario.
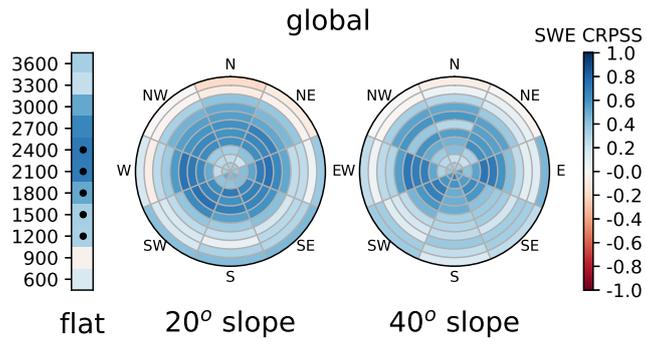
Figure 9. Same as Fig. 5 for the assimilation of HS of 2016_p60 ~~member~~ synthtic observation scenario in the 1200-2400 m flat classes.

Variable	Perturbation	σ	τ (h)
Precipitation $(\text{kgm}^{-2}\text{h}^{-1})$	Multiplicative	0.7	1500
Shortwave radiation (Wm^{-2})	Multiplicative	0.7	3
Wind speed (unitms^{-1})	Multiplicative	0.6	100
Longwave radiation (Wm^{-2})	Additive	$24.5\,\text{Wm}^{-2}$	30
Air Temperature (K)	Additive	1.08 K	15

Table A1. Perturbation parameters for the meteorological variables.

Variable	μ	σ
BC (wet and dry) $(\text{kgm}^{-2}\text{h}^{-1})$	-2	1
Dust (wet and dry) $(\text{kgm}^{-2}\text{h}^{-1})$	0	1

Table A2. Perturbation parameters for the LAP fluxes.