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-2mˆ2), 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, m^2/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”.

Corrected

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.

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.