Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-130-RC2, 2020

Geoscientific Model Development Discussions

Interactive comment on "CrocO_v1.0 : a Particle Filter to assimilate snowpack observations in a spatialised framework" *by* Bertrand Cluzet et al.

Kristoffer Aalstad (Referee)

kristoffer.aalstad@geo.uio.no

Received and published: 26 August 2020

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 ob-

servations 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.

Specific comments

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

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

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

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

L7 Consider changing *"non observed"* to *"unobserved"*.

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

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).

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

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

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.

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 Synder 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. 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.

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.

L69 Change *"It makes"* to *"This makes"*.

L79 Change *"operationally used"* to *"used operationally"*.

L81 Change *"enables to"* to *"enables us to"*.

L91 Change *"reflectance"* to *"reflectance observations"*.

L95 Change *"Following"* to *"Subsequently"*.

L101 Change *"the model into"* to *"the model for"*.

L102 Change *"enables to"* to *"enables us to"*.

L112 Change *"enabling to represent the snowpack coupling"* to *"coupling the snowpack with"*.

L117 Change *"This way,"* to *"As such,"*.

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 they are not matrices or vectors. That means (for example) using $\widehat{\mathbf{x}}_b^i$ rather than $\widehat{\mathbf{x}_\mathbf{b}^\mathbf{i}}$ and $\widehat{\mathbf{X}}_b$ rather than $\widehat{\mathbf{X}_\mathbf{b}}$. 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 $\mathbf{x}$) and instead use a hat for the predicted observations ($\widehat{\mathbf{x}}$ or better yet $\widehat{\mathbf{y}}$).

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

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

L165 I didn't see N_{eff} defined or even mentioned in Doucet et al. (2001), but maybe I missed it.

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

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

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

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.

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 $\mathbf{R}$? Also, perhaps use

another term then *"significant"* which unfortunately still has strong statistical (null hypothesis significance testing) connotations.

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

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.

L207 Change *"It allows"* to *"This allows us to"*.

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.

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? 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?

L230 Change *"date"* to *"dates"*.

L233 Change *"is set"* to *"are set"*.

L235 Change *"uses only"* to *"only uses"*.

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 $\mathcal{T}_{c,t}$ ($\mathcal{T}$ 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 $SWE^{\star}_{c,t}$ for the true SWE in a given scenario?

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

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?

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

L292 Change *"contrasted"* to *"contrasting"*.

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

L296 Again consider using another word than significant. 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)?

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.

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. HS observations are also often not representative of the model scale.

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.

L324 Change *"well represent"* to *"properly represent"*. Also on the next line use (e.g.) *"marked"* instead of *"significant"*.

L327 Change *"with respect to"* to *"compared to"*.

L330 Why is *"Skill"* capitalized?

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).

L356 Change *"conceptual"* to *"synthetic"*.

L360 Change *"on the"* to *"for the"*.

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.

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

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.

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

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

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"*.

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

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

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

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).

L408 Change *"informations"* to *"information"*.

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.

L414 Change *"reliable model for that"* to *"reliable LAP model"*.

L418 Change *"suffers from obvious"* to *"suffering from obvious"* and *"suffer for large"* to *"suffer from large"*.

L421 As before, in Bayesian probability theory how can an ensemble correlation be real?

L424 Change *"area"* to *"areas"*.

L426 Change *"into larger"* to *"for larger"*.

L427 Change *"take the best"* to *"outperform"*.

L451 Change *"in the way of a new"* to just *"in a new"*.

L456 Change *"spatialized"* to *"semi-distributed"*. Also mention somewhere in the conclusion that this is a synthetic experiment.

L460 Capitalize the leading words in this enumeration.

L469 Change *"errors"* to *error"*.

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.

L490 Change *"softwares"* to *"software"*.

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

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.

Table 1: Change N^*_{eeff} to N^*_{eff}. In the caption, change *"setup of"* to *"Setup for"* and change *"snow depth"* to *"height of snow"* to be consistent with the rest of the manuscript. The same applies to the title of subsection **4.2.1**.

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.

Table 3: Same problems as with the other Tables.

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

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

Fig. 4: In the caption, change *"on the whole"* to *"for the whole"*.

Fig. 6: In the caption, consider changing *"synthetic members"* to *"synthetic scenarios"* (since these are not ensemble members). Also, why is *"Skill"* capitalized?

Fig. 8&9: In the caption, consider changing *"member"* to *"scenario"* to avoid confusing the concept of your truth scenarios and the ensemble.

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.

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.

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.

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

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

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.

Synder 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.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-130, 2020.