The Cryosphere Discuss.,
https://doi.org/10.5194/tc-2017-109-AC1, 2017

The Cryosphere
Discussions
Open Access

Interactive comment on "Ensemble-based assimilation of fractional snow covered area satellite retrievals to estimate snow distribution at a high Arctic site" *by* Kristoffer Aalstad et al.

Kristoffer Aalstad et al.

kristoffer.aalstad@geo.uio.no

Received and published: 24 October 2017

Reply to Reviewer 1

We are grateful to the reviewer for the thoughtful comments and suggestions to our manuscript. We have compiled a revised version and in the following provide a point-by-point reply to all issues raised. The reviewer's comments appear in bold font and our replies in normal font. Excerpts from and changes to the manuscript are quoted in italics. Page and line numbers refer to positions in the original manuscript.

Printer-friendly version

Discussion paper

Ensemble-based assimilation of fractional snow covered area satellite retrievals to estimate snow distribution at a high Arctic site

The Cryosphere, Aalstad et al., 2017

The paper presents a case study in which MODIS (and Sentinel-2) fractional snow cover area retrievals are assimilated into a simple snow model to infer peak snow water equivalent and subgrid snow estimates. The ensemble smoother with multiple data assimilation technique is used. The paper adds interesting new insights to the discipline of snow data assimilation and could be considered for publication, after addressing the topics below:

1) - There is a weird mix of poor sentences in a very weak English language (too much to list, 1 random example: p.8, L5: "In addition, not accounted for. . . contribute to. . .") and parts with excellent English. My hope is that the excellent English parts are inserted by one of the co-authors and not copied from elsewhere. Please revise the entire paper for its language. Related to this, I would also suggest to revise the title to something like: "Ensemble-base assimilation of satellite-based fractional snow cover area to estimate the snow distribution at Arctic sites" (Arctic is already high-latitude; the sites do not have a particularly high elevation - I have no idea what "high" was referring to)

Following the reviewer's suggestion the entire manuscript has been revised for its language.

According to the reviewer's suggestion, the title has been revised to

“Ensemble-based assimilation of fractional snow covered area satellite retrievals to estimate the snow distribution at Arctic sites”

- Text: nice literature review, well done. The other text is a bit long in general, and has quite some repetitions in the discussion section in particular – please condense where possible. E.g. referring twice to the adaptive version of the ES-MDA, referring twice to the spatially distributed modeling, etc.

The potential for spatially distributed modeling and assimilation is now only mentioned in the outlook (Section 5.4). Furthermore, the adaptive ES-MDA is now first mentioned in Section 5.2 (Evaluation of data assimilation schemes) and once more briefly in the outlook. In addition, we have removed the statements concerning the novelty and basis of our work early in Section 3.3 as this was already mentioned in the introduction. Some necessary repetition between the outlook (Section 5.5) and other parts of the discussion remains, but this is kept as brief as possible.

- what are “patterned gorund features”? A Nordic term?

Thanks for spotting a typo. The term has been changed to

“patterned ground features”

which refers to a phenomenon in periglacial regions where patterns, such as sorted circles, form in the ground material.

2) MODIS, p.13, L.14: how exactly do you average the pixels for each study area? What to you mean by "all the pixels"? Would you include pixels with any (low) cloud fraction? (cloud fraction is given as additional information as a confidence measure for the fSCA retrieval estimates) What is the cloud cover limit? Mention explicitly that you are averaging the satellite data to the 1 km (I suspect), both for MODIS and Sentinel-2. Yet, it is mentioned that the Sentinel-2 data are averaged to the footprint of the snow surveys. . . so perhaps the latter is not right. If the resolution of the Sentinel-2 data, MODIS data and model are different, please explain how you reconcile the space-mismatch. Also mention the 1-km spatial resolution upfront in the modeling for clarity.

For each study site and day we simply take the mean fSCA over all the corresponding MODIS pixels shown in Figure 1. This average is only taken if none of the pixels are flagged as cloudy by the MODIS cloud mask. The MOD10A1 and MYD10A1 version 6 products only accept pixels flagged as cloud free (i.e. either "confident clear", "probably clear" or "uncertain clear") by the MODIS cloud mask (see Riggs and Hall, 2016). Cloud fraction is not given as a confidence measure in these products (Riggs and Hall, 2016). If cloud free pixels are available from both Terra and Aqua, then Terra pixels are chosen. To clarify how clouds are dealt with we have added the following sentence to Section 3.2.1:

'We average over all the pixels for each day and study site (see Figure 1). This average is only taken if cloud free (as determined by the MODIS cloud mask) retrievals are available for each of these pixels."

As stated in the Section 3.2.2, the Sentinel-2 fSCA retrievals are mapped to the approximate footprint of (i.e. the area encompassed by) the snow surveys. The areal extent of the Sentinel-2 retrievals is close to those given in Table 1. As such, the following sentence has been added to Section 3.2.2:

"Therefore, the areal extent of the Sentinel-2 fSCA retrievals closely matches the areas of the corresponding study sites given in Table 1."

Consequently, there is a space-mismatch between the MODIS and Sentinel-2 fSCA retrievals. The latter have lower representativeness error as they provide a better match to the area covered by the snow surveys. The point of the exercise in Section 4.3 is to check whether or not this mismatch has a considerable effect on the assimilation results. Still, the space-mismatch in the MODIS pixels is not huge (Figure 1). Moreover, the space-mismatch is reconciled by the difference in the observation error variance (RMSE^2) between the MODIS and Sentinel-2 retrievals, determined based on the field measurements of fSCA, where the Sentinel-2 error is considerably lower.

As for the resolution of the modeling, the following has been added to the end of the modeling section (Section 3.1.2) for clarification:

"The model resolution is defined by the footprint of (area encompassed by) the snow surveys for each site (see Table 1 and Figure 1)."

In addition, the following has been added to the end of the forcing section (Section 3.1.3):

"While the resolution of the downscaled forcing data do not exactly match the model resolution (i.e. the footprint of the snow surveys, Section 3.1.2), the mismatch is small considering the gentle topography of the study sites (Section 2.1)."

Printer-friendly version

Discussion paper

3) Ensemble data assimilation:

P15, L4 explicitly name the "parameters" as "perturbation parameters" to avoid confusion with model parameters.

Done.

P16, L8-9: "remaining parameters", "prior parameter ensemble": again refine the word choice here or define "state", "parameter" and "perturbations" more clearly or efficiently upfront, because this sentence is referring to two types of parameters and confusing.

Since peak mean SWE is also treated as a state variable in the SSM, the sentence has been reformulated to

"We emphasize that through the perturbation parameters we effectively perturb the melt rate, precipitation rate and coefficient of variation. By performing a subsequent ensemble integration of the SSM we also get an ensemble of state variables that are consistent with the prior perturbation parameter ensemble."

Where we have also explicitly referred to the parameters that are perturbed as perturbation parameters. In addition, we have changed *"prior ensemble of parameters"* in Table 4 to *"prior ensemble of perturbation parameters"*.

P16, L18: identify here (again) what is in the state and parameter vector. I suspect that there are no "snow model parameters" included anywhere, but it would be good to explicitly mention this. Clarify upfront that perturbation parameters are updated and not state variables or model parameters, as is often the case in other hydrology/cryosphere research.

The section starting from the middle of P16 L16 running to P16 L21 has been modified to

"Let N_e, N_o, N_a, N_s, N_p and N_t denote the number of ensemble members, observations, assimilation cycles, state variables, perturbed parameters and time steps during an annual (September-August) model integration. $\mathbf{X}$ is the $(N_s \times N_t) \times N_e$ matrix containing the ensemble of states ($fSCA_{n,j}$, $D_{m,n,j}$, $\overline{D}_{n,j}$ and $\mu_{n,j}$) and $\boldsymbol{\Theta}$ is the $N_p \times N_e$ matrix containing the ensemble of perturbation parameters listed in Table 4. The $N_o \times 1$ observation vector $\mathbf{y}$ contains all the fSCA satellite retrievals during the ablation season (Section 3.2), $\mathbf{Y}$ is the $N_o \times N_e$ matrix containing the ensemble of perturbed fSCA satellite retrievals and $\widehat{\mathbf{Y}}$ is the $N_o \times N_e$ matrix containing the ensemble of predicted fSCA observations. Additionally, $\mathbf{H}$ is the observation operator (mapping from the state space to the observation space) and $\mathbf{R}$ is the $N_o \times N_o$ observation error covariance matrix which is a diagonal matrix containing the observation error variances (Section 3.2)."

To emphasize that the model constants (model parameters) in Table 2 are fixed and that only the perturbation parameters are updated in the analysis we have changed the sentence on P18 L9 to

"We emphasize that the analysis step (21) only updates the perturbation parameters

and a consistent ensemble of states is found from the subsequent ensemble model integration. The model constants listed in Table 2 remain unchanged by the analysis and the integration.”.

In addition, all further mentions of *”parameters”* have been corrected to *”perturbation parameters”*.

Fig.3: replace block with “Update states” ⟶ “Propagate state” (I believe that the updating is done on the parameters in the ES-MDA Analysis step)

Done. For consistency, in this Figure the block *”Generate Prior Parameter Ensemble‘”* was changed to *”Generate prior perturbation parameter ensemble”*.

4) Have you evaluated the updated parameters (result of Eq. 21) themselves? Do the bias estimates make sense? E.g. can you compare the re-analysis forcings, the bias-corrected ones and in situ observed meteorological data?Âă

Snowfall observations are not available for our field sites (Boike et al., 2017) and are very difficult to conduct in the Arctic due to undercatch (e.g. Førland and Hanssen-Bauer, 2000). For the snowmelt, a field based estimate was available from Westermann et al. (2009). We have added the following to the end of Section 4.1:

“The posterior bias parameters can be directly evaluated by comparing the bias corrected forcing to field measurements. Due to a lack of snowfall observations (see Boike et al., 2017), an evaluation of the precipitation bias parameter is not possible. However, the melt bias parameter can be evaluated by comparing the snowmelt flux

(which is directly proportional to the perturbed melt depth) to field-based values. For June 2008, Westermann et al. (2009) estimate an average snowmelt flux of 27 Wm^{-2}, which compares well to the ES-MDA posterior median estimate (averaged for the same period) of 29 Wm-2, while the prior median estimate is too low with 19 Wm-2."

5) P.20, L17 and Fig 4: peak measurements. . . who or what knows that this measurement is taken at the peak? How can we know for sure that it is a peak measurement, if there is only one data point? Why is that peak measurement always located on May 1st in Fig 4? The peak measurement must be at a different time every year. . .

Only a single survey is available for each site for a given snow season. The times of these surveys (April/May), coincide closely with the peak SWE. We have added a quantification of the associated error using snow depth measurements from a sonic ranger at the Bayelva climate station to Section 2.2 (P8 L7):

"Although the snow surveys coincide closely with peak SWE, some accumulation (ablation) may occur after (before) the surveys. To assess the magnitude of this error source, we used snow depth measurements at the Bayelva station (Boike et al., 2017) to compare the snow depth at the survey dates to the maximum snow depth for each snow season. We found an average relative difference of 8% (maximum: 17%, minimum: 0.3%)."

Figure 4 has also been changed. It now shows the measured peak mean SWE μ (as determined by the snow surveys) over the whole snow season as a horizontal dotted black line. The caption of Figure 4 was changed to

"Time evolution of the prior (red) and ES-MDA ($N_a = 4$, $N_e = 10^2$) posterior (blue) fSCA (first column) and mean SWE ($\overline{D}$; second column); shading: 90^{th} percentile range; solid lines: ensemble median; yellow dots: assimilated MODIS and Sentinel-2 fSCA retrievals; dotted black line: independently observed peak mean SWE (μ) from snow surveys (Section 2.2); x-axis: months. These results are from a single run."

6) Fig 5: these distributions seem not to be cross-masked, i.e. why are the prior and posterior estimates not cross-masked to the times and locations of the in situ observations?

We are not sure what the reviewer means by cross-masked. As discussed above, the times of the snow surveys coincide closely with peak SWE. In Figure 5 we are comparing the prior and posterior estimates of the peak subgrid SWE distributions with the corresponding empirical distribution measured in the field through the snow surveys for each study site and snow season. To clarify this, the caption in Figure 5 has been changed to

"Prior (red), ES-MDA ($N_a = 4$, $N_e = 10^2$) posterior (blue) and the corresponding independently observed (from snow surveys; dashed black) peak subgrid SWE distributions; shaded areas: 90^{th} percentile range; solid lines: ensemble median; markers: mean value. These results are from a single run."

7) Table 4, Figure 7: clarify in the caption which validation data are used, where and during which period.

The caption of Figure 8 (previously Figure 7) was changed to

”Fractional improvement in RMSE through the analysis step (1 being perfect and 0 no effect) as a function of the number of ensemble members for the fSCA, peak mean SWE μ and coefficient of variation χ; top left: particle batch smoother, PBS; top right; ensemble smoother, ES; bottom left; ensemble smoother with multiple data assimilation, ES-MDA; bottom right: FI as a function of assimilation cycles in the ES-MDA. The FI for $N_e \leq 100$ are averaged over 100 independent ensemble model integrations. Errors were computed based on comparisons to all the corresponding field measurements presented in Section 2.2.”

The caption of Table 5 (previously Table 4) was changed to

“Summary of evaluation metrics, i.e. bias, RMSE and square correlation coefficient (R^2), for the fSCA, peak SWE (μ) and peak subgrid coefficient of variation (χ). These metrics are based on comparisons to all the field measurements presented in Section 2.2 with the number of observations for the comparisons in brackets next to the corresponding symbols. All the metrics are averaged over 100 independent runs each with 100 ensemble members. The ES-MDA was run with $N_a = 4$ assimilation cycles.”

For consistency, the caption in Table 6 (previously Table 5) was changed to

”Summary of evaluation metrics, i.e. fractional improvement in bias and RMSE as well as prior and posterior square correlation coefficient (R^2), using the ES-MDA ($N_e = 10^2$, $N_a = 4$) for peak mean SWE (μ) and coefficient of variation (χ) when assimilating only MODIS as well as assimilating both MODIS and Sentinel-2 observations. These metrics are based on a comparison to all the snow surveys conducted in 2016 (see Table 1) and are averaged over 100 independent runs each with 100 ensemble members.”

Printer-friendly version

Discussion paper

8) P.23, L6: in situ fSCA retrievals – are these the camera-based data? Please clarify. It would also help to use the same term to refer to validation data throughout the paper – i.e. in situ vs ground-based vs ground truth vs "Field measurements" (section 2.2)

Yes, these are the fSCA field measurements based on data from the automatic camera system, UAV and GPS-based surveys discussed in Section 2.2. To clarify and avoid confusion with the satellite retrievals, the corresponding section of P23 L6 was modified to

"field measurements of fSCA (Section 2.2.)"

For consistency, all mentions of *"in situ"* (with one exception) or *"ground based"* or *"ground truth"* (with one exception) have been changed to *"field measurements"*.

9) P24, 15: "lowest improvements": how confident are you with this statement? With only 8 samples, it is hard to get statistical significance of any sort.

To qualify this statement, the sentence on P24 L15 was changed to

"For all schemes the available validation data suggests that the greatest improvements are achieved for fSCA, followed by peak mean SWE, while by far the lowest improvements are found for the peak subgrid coefficient of variation."

Printer-friendly version

Discussion paper

10) Table 5: not sure if this exercise has any value with only 3 observations as validation. . . what is the confidence interval on these metrics?

As the Sentinel-2 mission is relatively new (the first satellite, Sentinel-2A, was launched in June 2015), only 3 snow surveys (from 2016) are available for comparison at our study sites since the start of the mission. As such we have added the following cautionary statement to Section 4.3 (P26 L15):

"We emphasize that this evaluation is based on the only 3 available field measurements of μ and χ in 2016 (from the snow surveys), so that these preliminary results need to be consolidated by future studies with more validation data."

11) Section 4.3: "Effects of observation error and assimilation frequency" ?

The heading has been changed accordingly. The same change was made to the heading of Section 5.3 (previously Section 5.4).

12) Conclusions, P31: "For peak mean SWE. . . lower than in previously. . .": this is an apples to oranges comparison to published work over different regions and different periods and thus invalid; please remove.

Done. The end of the sentence has been removed so that it now reads

"For the peak mean SWE, the ES-MDA features an average RMSE of 0.09 m w.e. compared to field measurements"

Thank you once again for all the helpful comments and suggestions,
On behalf of all the co-authors,
Kristoffer Aalstad

References

Boike, J., Juszak, I., Lange, S., Chadburn, S., Burke, E., Overduin, P., Roth, K., Ippisch, O., Bornemann, N., Stern, L., Gouttevin, I., Hauber, E., and Westermann, S. (2017). A 20-year record (1998-2017) of permafrost, active layer, and meteorological conditions at a High Arctic permafrost research site (Bayelva, Spitsbergen): an opportunity to validate remote sensing data and land surface, snow and permafrost models. *Earth System Science Data Discussions*, pages 1–86. 10.5194/essd-2017-100.

Førland, J. E. and Hanssen-Bauer, I. (2000). Increased Precipitation in the Norwegian Arctic: True or False? *Climatic Change*, 46:485–509. 10.1023/A:1005613304674.

Riggs, G. A. and Hall, D. K. (2016). MODIS Snow Products Collection 6 User Guide. Version 1.0.

Westermann, S., Lüers, J., Langer, M., Piel, K., and Boike, J. (2009). The annual surface energy budget of a high-arctic permafrost site on Svalbard, Norway. *The Cryosphere*, 3:245–263. 10.5194/tc-3-245-2009.

Interactive comment on The Cryosphere Discuss., https://doi.org/10.5194/tc-2017-109, 2017.