Exploring the potential of thermal infrared remote sensing to improve a snowpack model through
an observing system simulation experiment
Alonso-Gonzalez et al.

Second round of Review by Bertrand Cluzet

I am pleased to say that the authors have addressed my major concerns. The Observing System Simulation Experiments setting is now more rigorously set, by using a different (and more complex) version of the model to generate the synthetic truth than the one used in the assimilation ensemble. As expected, this degrades a bit the estimated performance of the assimilation algorithms (the degradation is even very strong for the PF on Fig. 7), but increases the reliability of the statements of this paper.

The inclusion of the synthetic assimilation of Snow Cover Fraction (SCF) is also a nice addition, as it allows to estimate the relative information content of Land Surface Temperature (LST) observations with respect to SCF. As the community has a good understanding of SCF assimilation and information content, presenting LST as a similar, but potentially even more informative variable, is probably the most important statement of this paper.

Please find below some minor comments, the most important ones revolving around the definition of what is the "reference" and how should its performance be measured and asking for a better structure of the data and methods section. I am pleased to provide the authors with an annotated version of the manuscript, containing some comments to address, and some suggestions.

I have no doubt that the authors can address these comments easily and am looking forward to seeing this paper published!

Minor comments:
- The parameters chosen for the log-normal distribution are such (mean= 1.15, median = 1.57) that the prior is likely not centred on the openloop run (which is displayed in the figures, and serves as a reference for the scores), but is already pulled a bit towards the synthetic observations. In other words: the openloop is not representative of the prior, and this could be misleading. From a theoretical standpoint, in order to infer the added information of the data assimilation algorithms, the scores should be computed with respect to the prior mean or median's performance, and not the openloop (see also Aalstad et al., 2018). I would ask the authors to add the score of the prior to all the evaluation plots. Please also add the prior ensemble quartiles (25-75 quantiles + median) to Fig.2 so that the reader can visualize how does the prior look like (as is done in Aalstad et al, 2018, e.g. Fig. 2.). I acknowledge that this point was raised by the last comment of reviewer 2, and rejected by the authors, but I still think that this would be very informative for the reader.

- What is the observation error prescribed in the PBS/PF? I could not find this important piece of information in Sec. 2.

- Please substantially improve the writing of the data and methods section, by splitting it into subsections, and adding several important pieces of information (nRMSE is not introduced, observation errors are not specified, model parameterization are only vaguely described)

- MuSA version: the authors say in their response that "there is no specific tag for the version used here, so we just included the version number. The modifications mentioned here are

in the FSM code, but is also included in the MuSA version". Sorry to insist, but having a clear tag or commit reference is essential to ensure science reproducibility. I found the following tag on the github repo: https://github.com/ealonsogzl/MuSA/tree/v1.0. If this corresponds to the exact version of the code that was used for the paper, please put it into a Code and Data availability section, otherwise just provide a new tag.

-   Briefly compare FSCA-DA results with the literature. I think that dividing RMSE by a factor of 2-3 as shown in Fig. 5 is in the ballpark of the literature with real SCF data assimilation (see e.g. Aalstad et al., 2018). This could help convincing the reader that this OSSE setup provides a realistic estimate of the information content of LST and SCF.

Technical comments:
-   Fig. 2: the spread of the different "replicates" is a nice addition. Perhaps consider making it more clear to the reader that here, the replicates differ only from observation error noises (and not from e.g. different input perturbation scenarios), either in the text, or the figure legend.

-   Try to include the keywords 'assimilation' 'data assimilation' in the title.

References:

1. Aalstad, K., Westermann, S., Schuler, T. V., Boike, J. & Bertino, L. Ensemble-based assimilation of fractional snow-covered area satellite retrievals to estimate the snow distribution at Arctic sites. *The Cryosphere* **12**, 247–270 (2018).