Review of

Ensemble-based snow depth data assimilation for a multi-layer snow scheme over the European Arctic

by Åsmund Bakketun, Jostein Blyverket, and Malte Müller

General comments

This manuscript presents a regional reanalysis system for snow water equivalent. It is named CARRA-Land-Pv1 and is based on a multi-layer snow model (ISBA, with 12 vertical layers) and a Local Ensemble Transform Kalman Filter to assimilate snow depth in the model. This new reanalysis is evaluated against a reference analysis, CARRA, which also used ISBA but as part of a larger model that also includes all the other components of a numerical weather prediction model. Therefore, CARRA and the new CARRA-Land-Pv1 proposed by the authors share the same snow model and the same meteorological forcings (generated by CARRA) but they differ in 1) the assimilation method itself and 2) the observation sets that are assimilated (partially, as some observation stations are used by both systems). The two systems are also compared against observations from six completely independent snow pillow stations that are not assimilated in CARRA nor CARRA-Land-Pv1. The authors show that, in general, their proposed new reanalysis outperforms CARRA.

I believe the topic is of great interest to the readers of The Cryosphere. However, there are a few methodological aspects that need to be clarified, and I also found minor linguistic issues that need to be corrected. Specifically, details regarding the perturbation of meteorological forcings need to be added, and the LENKF method needs to be explained better. Finally, I could not find the resolution of the grid anywhere. I apologise if I have missed it, but if it is not mentioned, it should.

Specific comments

1. Introduction: too much emphasis on mountains
2. Line 118: The terms in equation (1) are not adequately defined, as it is not indicated what « x » stands for. You should take what is written at line 159 (« where x represents the ensemble control vector (…) a and b indicate analysis and background (…) » and place it on line 118 instead.
3. Lines 143-148: You mention on several occasions that the state vector has high dimensionality. It would be good to provide the reader with numbers regarding what size is small, what size is average, and what size is large. In addition, I think a more detailed explanation of the concept of localization is needed. In particular, how do you ensure that neighboring points remain correlated (preserving the spatial structure of snow depth and SWE) if data assimilation is performed independently point by point?
4. Figure 1: I cannot understand this figure. I have read the text several times, I still don't understand the remapping. I understand that the purpose is to better represent the spatial uncertainty of precipitation, but I don't understand the method and Figure 1. Is it possible to modify Figure 1 to make it clearer? Maybe provide an example with numbers and/or real grid cells? The Appendix also did not help.
5. Section 2.4: Please provide a table that indicates the range of perturbations used for each variable, as well as any relevant equations.
6. Line 310: You find the ensemble to be only marginally better than the average of all its members. However, I'm wondering about the potential users of that new reanalyse. Do they want only deterministic estimates of snow depth, or is there value for them in having access to the full ensemble and information about the uncertainty? I think this should be discussed.
7. Line 316: I'm not sure « climatology » is the right word for geographically close stations. I understand that they might have differences in their observation records, but can we really talk about different climatology?

Linguistic comments

8. Abstract, lines 2-4: « merge information from the two sources » is not clear, as the « two sources » could be understood as either « prediction systems » and « historical reanalysis » or « observations » and « physical laws in models ». The reader has to logically deduce that you are referring to the latter, but it is not clear from the way the sentence is written. I suggest reformulating.
9. Abstract line 11: Is it possible to replace « relatively large » by something more precise?
10. Line 28, remove the comma between « models » and « is »
11. Line 155: This group of equations should be numbered
12. Line 167, there is an « s » missing in « represents»
13. Lines 382-383: The sentence « In emission modelling, the use of statistical models for subgrid processes has been used to encounter similar problems » Needs to be reformulated. I suggest « In emission modelling, **statistical models for subgrid processes have been used** to **counter** similar problems
14. Line 325: the word « stations » is missing between « observation » and « available »