Response to Reviewers' Comments on the 3rd Version of Manuscript of

Assimilation of GPM-retrieved Ocean Surface Meteorology Data for Two Snowstorm Events during ICE-POP 2018

We would like to thank the reviewer and the editor for their thorough analysis of our manuscript and for suggesting changes that will help our paper to have a better quality. We have reproduced the reviewer's comments below — our responses appear in blue. The page and line numbers refer to the trackchanged manuscript.

Comments from Reviewer #1:

Re-Review of
Assimilation of GPM-retrieved Ocean Surface Meteorology Data for Two Snowstorm Events during ICE-POP 2018
by
Li et al.

General comments
This is the third version of the manuscript. The authors have changed the focus of the manuscript and further removed some unnecessary discussions. The manuscript is now more focused and refined. I appreciate the efforts of the authors. Although I personally feel like the content is a little too thin (more in-depth diagnostics are recommended), I believe the manuscript is now closer to publication with a few more comments.

Comments:
1. L309-334: I agree with the authors that "Since this paper focuses on case studies for assimilating a new dataset with high resolution regional simulation, it is always a good idea to check the background, the observation innovation (O – B) and the analysis increment (A – B) and see whether or not the data assimilation is effective (similar plots can be found in numerous past studies such as Xiao et al. 2007, Bi et al. 2011, Chen et al. 2021, etc).". However, as far as I can tell from the three citations and lots of other studies, a reasonable check for the DA effectiveness usually starts with O-B vs O-A (e.g. Fig. 6 in Xiao et al. 2007; Fig. 1& 4 in Bi et al. 2011; Fig. 4 in Chen et al. 2021), not O-B vs A-B. One simplified example is A=3, O=2, B=1, O-B=1 & A-B=2. You'll see both red in Figure 6, yet you cannot say the overcorrection is an effective DA. It is not wrong to use A-B, just less straightforward. Also, why does A-B look much smoother than O-B? Are you plotting one in observation space, and the other one in model space? If so, please use a consistent approach.

Thank you so much for the comment. We agree with the reviewer that either O-A (as shown in the literatures mentioned above in the reviewer's comment) or A-B (e.g., Benjamin et al. 2004; Bölöni and Horvath 2010; Piccolo and Cullen 2016, to name a few) can be used to verify the data assimilation effect. We prefer to use A-B simple because A-B (the so-called analysis increment) provides the readers a direct vision on where and by how much the data assimilation corrects the model guess fields. In the example that the reviewer provided, A=3, O=2, B=1. Since the reviewer mentioned the color red, we think the reviewer was talking about the temperature field. Using these numbers, we will have O-B=1 and O-A=-1, meaning the analysis

overpredicts temperature. At the same time, with O-B=1 and A-B=2, we can reach the same conclusion that the analysis has an overprediction. This can be seen in Fig. 6. In Fig. 6b, we will see the color for O-B is golden (the golden color is for the value ≥ 1.0 and < 2.0). In Fig. 6c, we will see the color for O-A is orange (the orange color is for the value ≥ 2.0 and < 3.0). This tells us that the analysis has overpredicted temperature.
No, we did not smooth any of the fields and both plots are in the observation space. During 3dvar, the impact of the each obs data was distributed over an influence area which includes the observation location and other grid points (in both horizontal and vertical direction), that's why we see smoother fields in analysis.

Reference:

Benjamin, S. G., Dévényi, D., Weygandt, S. S., Brundage, K. J., Brown, J. M., Grell, G. A., Kim, D., Schwartz, B. E., Smirnova, T. G., Smith, T. L., & Manikin, G. S. (2004). An Hourly Assimilation–Forecast Cycle: The RUC, *Monthly Weather Review*, *132*(2), 495-518. https://journals.ametsoc.org/view/journals/mwre/132/2/1520-0493_2004_132_0495_ahactr_2.0.co_2.xml

Bölöni, G., and Horvath, K. (2010). Diagnosis and tuning of background error statistics in a variational data assimilation system. IDŐJÁRÁS Quarterly Journal of the Hungarian Meteorological Service, 114, 1 – 19.

Piccolo, C., and Cullen, M. (2016). Ensemble Data Assimilation Using a Unified Representation of Model Error, *Monthly Weather Review*, *144*(1), 213-224. Retrieved Jun 6, 2022, from https://journals.ametsoc.org/view/journals/mwre/144/1/mwr-d-15-0270.1.xml

2. Fig. 7 & L335-375: I'm still not fond of this comparison. The evolution of the difference between CTL & DA can tell us very limited information. It can only show us that the model is being stabilized after a while, but it cannot tell us if the changes are good or bad. If the authors cannot find appropriate observations, ECMWF analysis or FNL re-analysis can be other potential sources for verifications and are commonly used in early studies.

Thank you so much for the suggestion, but we don't think the global analysis or reanalysis like ECMWF or FNL would be good candidates for verification for this particular study. First, these analyses have a much coarser resolution (30 km resolution for ERA5 or 1 deg to 0.25 deg resolution FNL analysis) when compared to our model simulations (9 km + 3 km + 1 km). With the resolution of 30 km or .25 deg, these analyses would not be able to describe the local features over the complex terrain in Korean Peninsula, but these are important features to look at for the cases (e.g., Figs 8-11, 13-14) that we examined in this paper. Second, many literatures indicated that the global analyses and reanalyses have problems in providing accurate description for winter storms (e.g., Hamill et al. 2013 for case studies and Feser et al. 2021 for statistical analysis). Before the control run, we have looked at the FNL analysis and also conducted test runs WRF forecast using the FNL analysis as the initial condition and the result supported the statement. This served the motivation for our data assimilation experiments which incorporated additional (prepbufr and the satellite-retrieved) datasets.

Reference:

Feser, F., Krueger, O., Woth, K., & van Garderen, L. (2021). North Atlantic Winter Storm Activity in Modern Reanalyses and Pressure-Based Observations, *Journal of Climate*, *34*(7), 2411-2428. https://journals.ametsoc.org/view/journals/clim/34/7/JCLI-D-20-0529.1.xml

Hamill, T. M., Yang, F., Cardinali, C., & Majumdar, S. J. (2013). Impact of Targeted Winter Storm Reconnaissance Dropwindsonde Data on Midlatitude Numerical Weather Predictions, *Monthly Weather Review*, *141*(6), 2058-2065. https://journals.ametsoc.org/view/journals/mwre/141/6/mwr-d-12-00309.1.xml

3. Although the authors have improved the manuscript, there are still some irrelevant number listings. For example,
(a) L297: What is the point of the median values of the observations? They are time-/case-sensitive values. What can the readers learn from them?

The median values of the observations have been removed from the content. Please see Page 10 Line 296-298.

(b) L300-308: Although the skewness can be calculated in a variety of ways, I don't believe only the mean and standard deviation can be enough (at least a median value of those departures is needed depending on the calculation, but then the earlier median values can be misleading). Again, what is the point of those values here? It might be better to use skewness values here instead, or you can just use the figure to illustrate the points.

We have revised the manuscript. The information about the mean and standard deviation was removed. The sknewness values for the deference fields were added. Please see Page 10 Line 302-309.