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

By: Xuanli Li, Jason B. Roberts, Jayanthi Srikishen, Jonathan L. Case, Walter A. Petersen, GyuWon Lee, and Christopher R. Hain

Submitted to: *Geoscientific Model Development*

We would like to thank the reviewers 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 line number in this response refers to the manuscript without track changes.

Reviewer #1 comments:

General comments

This manuscript intends to investigate the impact of the assimilation of the Global Precipitation Measurement (GPM) retrieved ocean surface observations on two winter storms during 2018 using GSI and WRF model. The discussions primarily focus on the analysis impacts in the observation space, and the precipitation predictions for each case. This study has potentials in improving our understanding of how to optimally utilize the GPM retrievals. But the unclear motivation and lack of in-depth discussions make the paper reading frustrating. I would not recommend publication of this work until substantial revisions have been made. The detailed comments are listed below.

We have enhanced introduction section on the motivation (please see the last paragraph in Page 4 Line 113-134) and the data section (please see Line 153-166). We have made substantial changes throughout the result section and the conclusion section. We have also added one more figure to discuss how the data assimilation modified the moisture convergence and moisture transport, hence influenced the precipitation production (please see Line 396-408 in Page 13).

For the motivation of this research, as we stated in the introduction section, is that there has been very little effort to assimilate these types of observations in modelling systems. It has not been an emphasis in the satellite-based flux community methods focused on estimating the near-surface meteorology have been concentrated. So, this is a unique opportunity to test the assimilation of such a dataset. For this particular paper, the focus is on establishing the ability to model the systems with WRF, implementing a strategy to assimilate the widespread surface observations (e.g. the technical implementation, data thinning, etc.), and doing a characterization. We think that demonstrating the ability of the data to impact the model fields and understanding how those impacts were made are prerequisites to understanding whether or not to pursue more sophisticate process-based diagnostics (this will be the focus of a follow-on research).

Major comments/concerns:

One major concern I have about this manuscript is the unclear goal. If this study is focusing on the assimilation of GPM retrievals as suggested by the title, then the introduction should provide more information about the GPM retrievals, such as how it was (or was not) previously applied/researched in the field of snowstorm predictions (or other similar fields), and what's the novelty of this current study (not just the technical specifics in section 2.1). Also, the results section should contain more discussions on how the corrections from this dataset are physically benefiting the storm predictions, not just showing there is a difference and the model is improved. This is a case study after all. Please consider re-organize the manuscript/title to better represent your point and outstand your novelty.

We have enhanced the introduction section by adding the following discussion on the GPM retrieved data, as well as the motivation and novelty of the current research: "In the satellite-based surface flux community (e.g. see Curry et al., 2004), significant efforts have been undertaken to estimate the near-surface meteorology from passive microwave observations to support the development of turbulent flux estimates from space. In particular, efforts have been made to estimate 2m air temperature and humidity (e.g., Jackson et al. 2006, Roberts et al. 2010, and Tomita et al., 2018) to complement long-standing wind speed estimates from microwave observation. However, the aforementioned efforts have almost explicitly focused on large-scale production of the fluxes for climatological analyses with long latencies. However, the surface retrieval products essentially provide similar measurements to those of buoys and generally with accurate performance. There is a long heritage of assimilating ocean surface buoy measurements within a data assimilation framework, but there has been little effort focused on assimilating the surface retrievals. This is part due to a lack of a real-time availability of these estimates and partly due to the focus on radiance-based assimilation system. The latter are not particularly tuned for leveraging lower-layer information in microwave observations as the stand-alone efforts originating from the satellite-derived flux community. The ICE-POP 2018 campaign provided a unique opportunity with near-real time passive microwave estimates of surface meteorology and a heavily observed regional environment to test the potential impact of assimilating wide-spread observations of near-surface meteorology. In this research, we explored the assimilation of this dataset using case studies with two snowstorm events occurred during the ICE-POP 2018 period. The objectives of the current research are to characterize the forecast ability of snowstorm events over complex terrain with the WRF model and further to develop and evaluate an approach to assimilate the passive microwave derived surface meteorology. Our focus herein emphasizes the large-scale impacts of assimilation of the surface meteorology on the corresponding model fields and downstream forecast accuracy."

A new figure (Figure. 10, and discussions in Line 396-408 in Page 13) is also added to discuss how the data assimilation influenced the low-level moisture convergence and moisture transport and hence benefit the precipitation forecast of the snowstorm. More details on the physical *processes (e.g. ocean evaporation, water and energy budget analyses, etc.) will be investigated in follow-on researches.*

The results section spends large portions of contents on how the background is modified towards the observations for each snowstorm event. But the discussions on either A-B or O-B can just tell us that the DA is working properly since the verification is dependent. Any successful DA should make the analysis closer to observations than the background, this is from the mathematical nature of the minimization of the cost function. As indicated between L126-132, there should be plenty of other observations (e.g. D3R) available in addition to the retrievals. Why not use those independent observations to verify your analysis? It will objectively measure if the DA is really improving the analysis.

It is true that successful DA is meant to make the analysis closer to the observation than the background, but the data impact can vary, and sometimes it is case dependent. Therefore, it is important to know how close the analysis is to the observation. For this purpose, we compared A-B and O-B in Fig. 6 and Fig. 12 to examine how much influence the data assimilation brought to the model fields, temperature, specific humidity, and wind speed, and how close the analysis fields are to the observations. It is shown in Figs. 6 and 12 that the data impact does follow the pattern in O-B, but it varies in different events, even with the same DA procedure and parameters.

Yes, there are plenty of data collected from the ICE-POP 2018, as we mentioned in the last paragraph on Page 4. Some of the data are more readily to be used for model verification than the others. In the past research, we have examined the hourly precipitation rate retrieved from multiple ground-based radars (including D3R). We've also compared it with IMERG precipitation and found inconsistency between them due to the complicated nature of solid precipitation particles and the operational mode of the radars. Therefore, in this study, we choose to use the South Korean Surface Analysis dataset for model verification. This dataset is in-situ data based on observations from the dense Automatic Weather Station network and have a better accuracy in snowfall measurement. This 1-km resolution dataset is particularly produced for ICE-POP 2018. In the previous manuscript, we forgot to stress this point in the Data and Methods section. We've updated this information in our new manuscript (Line 154-160).

Minor but still important comments:

L220: The cost function is not just to measure the difference between the model and observations (otherwise it should be just yobs-Hxb). It is the sum of weighted (B and R) differences between the analysis estimation xa and model background xb (Jb), and between xa and observations y (Jo). The goal of DA is to find an optimal estimation xa that minimizes this J.

Thank you for correcting us, we have revised this sentence in Line 252. Equation 1 itself explains the cost function very clear that it is the differences between the analysis and model background weighted by background error plus the differences between the observation and analysis weighted by observation error.

L225: xguess is xb, please use just one subscription for consistency. This y term is also commonly known as innovation in recent DA literatures.

Xguess is now updated as Xb and the y term is now defined as the observation innovation (please see Line 255).

L240: This comment may be trivial, but can you provide more details on how you perform the cycling? There are two start mode in WRF, restart and cold start. In the restart mode, WRF is able to continue the model integration without interruption by using both the analysis at the current time step and the tendencies from the previous time step. However, our current DA methods usually only updates the analysis, not the tendencies. I'm wondering if you saw any discontinuity issue if you are using this mode. On the other hand, if you are using cold start mode (restart=.false. by default, which only uses the analysis at the current step), it will interrupt previous integrations and perform differently with your control experiments even without DA. You have to let the control exp stop at the same time for consistency.

The cycled data assimilation is conducted in the following steps, taking the March case as an example:

For the experiment DA_Mar: The WRF model simulation started at 00 UTC 7 March. At 06 UTC 7 March, the 6-h WRF forecast was used as the background field and prepbufr data was assimilated. The boundary condition file was updated and the data assimilation analysis was used as the input file for the new WRF simulation. The WRF model was then ran for 3 hours. At 09 UTC, the 3-h WRF forecast file was used as the background field and the GPM retrieved meteorology data was assimilated. The data assimilation analysis was used as the input file for another new round of WRF simulation. The above-mentioned steps were repeated for each data assimilation cycle.

The GSI data assimilation only updates the model control variables, not the tendency terms. Therefore, the tendency terms from the previous time step may not be used after data assimilation. We checked the model fields, wind, pressure, etc, and did not find any apparent discontinuity issues in the model runs. The assimilation of the GPM-retrieved data might have interrupted the integration, but the same interruptions happened in the CTRL_MAR when the prepbufr data was assimilated every 6-h. Therefore, we think DA_MAR vs. CTRL_MAR, and DA_FEB vs. CTRL_FEB are fair comparisons.

L261: Current Fig. 5 may not be the best way to show those info. I don't think it's possible to retrieve the percentile info directly from the figure.

This sentence has been revised as: "The 25th (and 75th) percentile values of the observed surface temperature, specific humidity and wind speed are 4.96 (and 16.71) °C, 4.06 (and 8.70 g kg-1), and 6.04 (and 10.74 m s-1), respectively, which indicates that a larger part of the observational data located at the south part of Sea of Japan and western North Pacific Ocean than the north part." (Please see Line 299-302)

L261-271: What do these statistics mean? Some of the o-b standard deviations are almost comparable or smaller than the corresponding observation errors (e.g. specific humidity). Are you suggesting your model background is very good already?

The purpose of the statistics is to provide the overall distribution of the observational data and the distribution of the difference between the model background and the observations in temperature, specific humidity and wind speed fields. No, since the observational data was retrieved from satellite observations, we would like to use the observational errors recommended by the GSI system.

L291-293: Maybe just show the RMSD of O-A and O-B to better present the point if you are not going to physically explain how those improvements can benefit the storm analysis or predictions?

The RMSD has been calculated and discussed in Line 338-340 on Page 11.

L299-301: I'm confused by the motivation of the discussion here. Differences should be there as long as you changed the initial analysis. Of cause the differences will spread out as the model tries to balance the changes. But without observations in those areas, how do we know if this "spread of information" is correct or not?

Following Reviewer #2's suggestion, we have removed this figure (previous Fig. 7). But we pointed it out because "the spread of information" and accumulation of the impact were caused by the cycled data assimilation, and the spread of information was not evenly distributed over the domain, even though we don't have observational data (out of the ICE POP 2018 domain) to evaluate the correctness of the information.

L308-315: What is your localization length scale? Are these reducing differences in the area of your observations and increasing differences in the remote area reasonable? What's the physical explanations or guesses?

For these two events, we have created our own background error statistics files because an effective data assimilation could not be made with the default regional background error file in GSI. The regional background error matrix (including horizontal length-scales, vertical length-scales, regression coefficients, etc. of the control variables) was created with 1-month 24-h and

12-h forecasts with the same WRF domains as the case studies using the WRFDA gen_be package. This point has been added in the updated manuscript (Line 260-264 Page 9). Sensitivity tests were conducted with different horizontal scale parameters for reasonable result from the assimilation of the GPM retrieved surface data. The current values of horizontal scale are set as 0.375, 0.75, and 0.75 in the GSI namelist file. This set gives a reasonable distribution of data increment for surface temperature, relative humidity, and wind speed as seen in Figs. 6 and 12.

The change in differences should be the effect of model dynamic adjustment during WRF integration, not from the data assimilation.

L324-329: How should we interpret this result from Fig. 8? Is it reasonable for a continuously cycled DA experiment to become more alike the NoDA experiment after several cycles? It appears to me that the impact of DA is fading after cycling. Is this true? Also, the increasing RMSD of wind speed at almost all levels from 21 UTC to 22 UTC (Fig. 8d) seems to be inconsistent with the reducing total RMSD in Fig. 8a. Why?

Figure. 7 (previous Fig. 8) shows the RMSD in surface temperature, specific humidity, and wind speed between DA_Mar and CTRL_Mar from 21 UTC 7 March to 03 UTC 8 March. The GPM retrieved surface data was assimilated at 21 UTC 7 March, followed by 6-h WRF integration. Fig. 8 indicates that the significant impact was made in the analysis field (21 UTC), the data impact declined by 17%, 15%, and 4% on average in the first hour of model integration on surface temperature, specific humidity, and wind speed, respectively. In addition, the changes can be seen not only in surface and low levels, but also in mid to high levels. But as shown in Fig. 8a, a large part of the data impact (0.78°C in temperature difference, 0.29 g/kg in specific humidity difference, and 1.59 m/s in wind speed difference) still retained in the 6-h forecast (03 UTC). Therefore, Fig. 8 does not support the idea that the continuously cycled DA experiment becomes more alike the NODA experiment.

Thank you for catching the mistake in the wind speed plot. The legend in the wind speed plot was incorrect. The gray line should be WSPD1 and the blue line should be WSPD. We have corrected it in the updated figure.

Fig. 9 -14: Can you use something like ETS score to quantitively verify the precipitation predictions? Also, in your discussions with the precipitation patterns, can you physically relate them to your previous analysis impacts? E.g. how does a warmer temperature in the analysis within certain area result in more precipitation predictions.

We added the ETS information for Fig. 9 (please see Line 388-395 for the discussion). ETS would not be necessary for Figs. 11 and 14 since the PDF is already provided to evaluate the forecast accuracy. We have also added Fig. 10 to explain how data assimilation modified the physics fields and hence influence precipitation (please see Line 396 – 408 on Page 13). More *investigation on the specific physically fields and processes that data assimilation has influenced will be explained in follow-on researches.*

L371-380 and others: Why does the accuracy of this IMERG matter? How is this related to the focus of this paper? It reads to me that the discussions on the IMERG is out of no where. If you think this is a important part of your project and is somehow related to the focus of this paper, please add more information in the intro.

Because the ICE-POP 2018 is also a part of the GPM ground validation effort in terms of solid precipitation, we think it might make some interesting points on the ability of IMERG data for snowstorms like the ones discussed in this paper. It provides a great opportunity to estimate the accuracy of GPM data for winter precipitation over complex terrain. We have added this point in the discussion section Line 435-438, and also in the conclusion section Line 541-544.

L407-410: Why does the specific humidity increase while the innovations are mostly negative?

Specific humidity varies with temperature, pressure and water vapor mixing ratio, and specific humidity is not a control variable in GSI. In GSI, the moisture control variable is pseudo relative humidity. For a more direct illustration of the data assimilation effect, we have replaced the specific humidity in Figs. 6 and 12 with relative humidity. From the new figures, you will find the A-B in relative humidity follows the general pattern of O-B.

L444-445: What makes the Feb case less significant than the Mar case? Any hypothesis? It appears to me that the O-B differences in Feb case is larger than the Mar case (Fig. 6 v.s. Fig. 12). Could it be related to my previous comment on the specific humidity?

Yes, the larger O-B may be a contributor to the less significant impact of the GPM retrieved data for the Feb case (we have added this point in the conclusion section, please see Line 532-534 on Page 17). Also, the different synoptic configurations and the undersampling of observational data are also important factors for that.