Response letter to Referee Reviewer 2

REVIEWER COMMENTS

Referee Reviewer 2

This article deals with a hot topic that may suit HESS. This study examines the sensitivity and robustness of the PGW method over NEUS by conducting multiple PGW experiments. In addition, several PGW experiments are conducted to answer the three key questions related to the application of the PGW method. The results may help further understand the impact of different PGW simulations on climate projections. Overall, I think it is a pretty good job. However, some scientific or presentation issues need to be carefully addressed. Therefore, the reviewer recommends that this manuscript should be accepted after minor revision.

Dear reviewer, thank you very much for your appreciation and suggestions. Certainly, these suggestions are very helpful in improving the scientific rigor and presentation of this work. We modified the paper correspondingly and please check the point-to-point responses below.

Minor revision:

It is recommended to use consecutive line numbers.

The line numbering is provided by the HESS LaTeX template. In the posted PDF line numbers appear to be consecutively numbered.

It is suggested that some necessary statistical parameters should be provided to quantify the difference in precipitation and temperature simulation performance of different schemes.

Thank you for your suggestions. We employed the K-S Test, which aims to determine if two data have the same distribution, and Student's t-test, which aims to examine if two data have the same mean value among all PGW simulations during the total simulation period. We chose the PGW_T_gp as the baseline and compared it with other PGW simulations on the regional mean precipitation and 2-meter temperature. Results show that the p-values from both the K-S Test and t-test are much larger than 0.05 in all cases, which indicates we cannot reject the null hypothesis that the regional mean precipitation and 2-meter temperature of PGW_T_gp and other PGW simulations have the same distribution and mean values at the 95% confidence level.

Further we conducted similar statistical tests on the precipitation and 2-meter temperature at each gridpoint. The results show that, except PGW_T_ZG_gp, the p-values of both K-S Test and t-test between all other PGW simulations and PGW_T_gp are nearly zero (much less than 0.05), indicating evidence against the null hypothesis that the precipitation and 2-meter temperature at each gridpoint of PGW_T_gp and other PGW simulations (except PGW_T_ZG_gp) have the same distribution and mean values at the 95% confidence level. However, the p-values of these two tests on precipitation and 2-meter temperature at each gridpoint between PGW_T_gp and

PGW_T_ZG_gp are still much larger than 0.05 (0.9997 and 0.7403 for K-S Test and t-test), indicating that, even at gridpoint scale, PGW_T_ZG_gp's precipitation and 2-meter temperature still have the same distribution and mean values as PGW_T_gp.

This lends evidence to our claim that modifying ZG (geopotential height) in PGW simulations makes little difference on the final simulation and is not necessary. Also, the statistical tests further confirm our conclusion that different PGW methods will displace the weather events (the precipitation and 2-meter temperature at each gridpoint) but have a few impacts on the regional mean meteorological fields (the regional mean precipitation and 2-meter temperature).

The figures legend/caption is not self-explanation, which should be improved. In addition, many subgraphs in Figures 2 and 3 are very similar, making their differences challenging to identify. It is suggested to adopt the form of Figs.8-12 or add the statistical parameters mentioned in question 2.

Thank you very much for pointing it out. We will improve the legends and captions to make them clearer. In Figures 2 and 3, we will plot the differences between our simulation and other datasets. Attached is the updated the Figure 3.

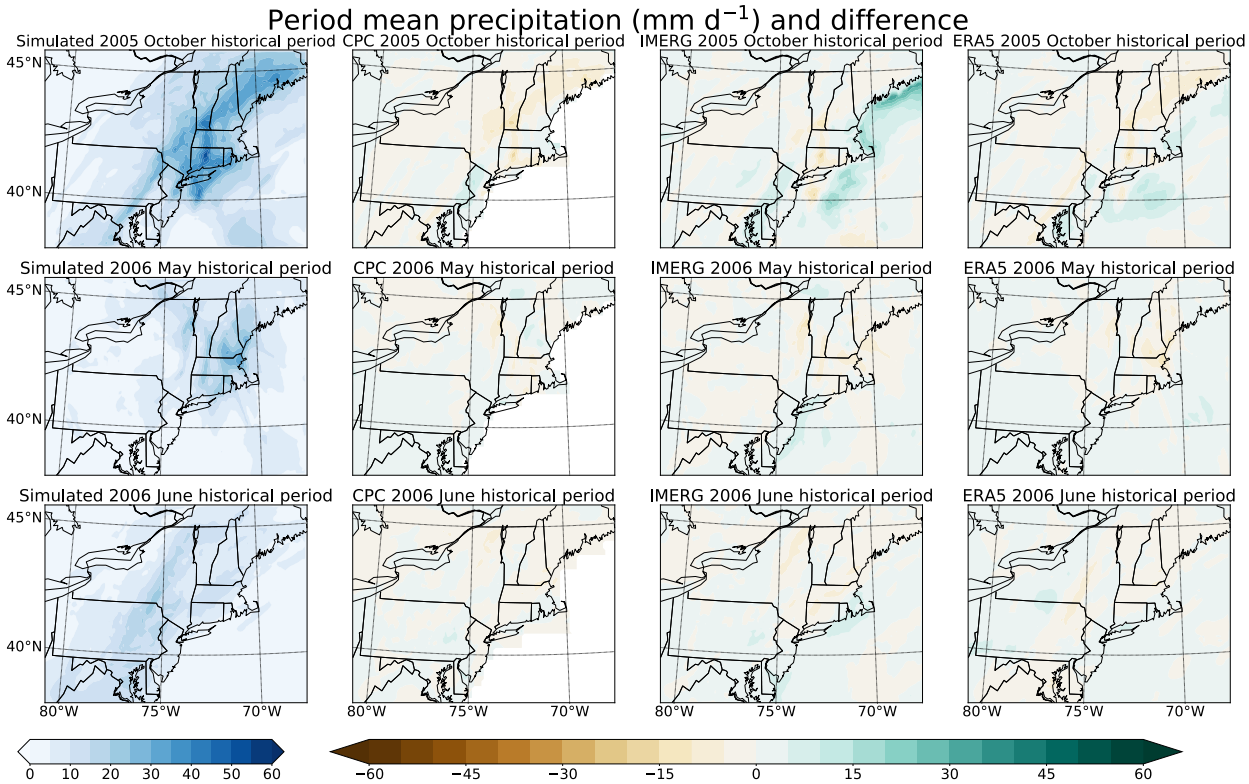


Figure R1 Period mean precipitation (mm/d) and difference (the difference is calculated as the difference between each reanalysis/observational data and our simulation)

Figure 1 recommends that the macro location comes from the continent to geologically locate. The locations of the three regions should be marked in Figure 1

Sorry, we're not sure if we understand this suggestion clearly. The three storms pass through our domain, along with their associated precipitation, as shown in the references in Table 1. In Figure 5 and Figure 8 (in the main paper), we can observe the distribution of precipitation during three storms and speculate about their centers; however, because they are noisy and spread out, it's difficult to directly specify the locations of these three storms and associated flooding on the domain figure.

Please clarify the "returned" flood period (2055 April to 2056 July). Is this the 50-year return period for floods considered? Why not consider using other periods?

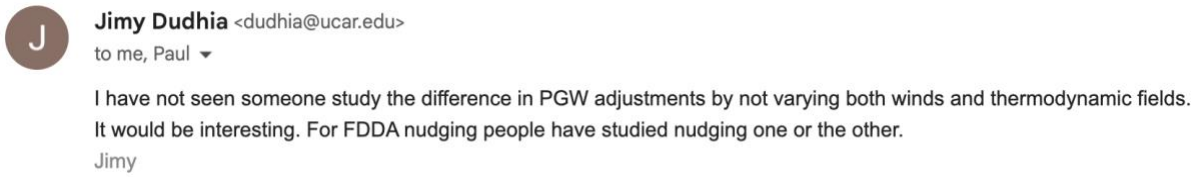
The "returned" flood period (2055 April to 2056 July) in this paper means the scenario when the same larger-scale circulation and dynamic fields of the historical 2005 April to 2006 July flood period reoccurred under the RCP8.5 warming scenario in the middle of this century. The historical 2005 April to 2006 July flood period includes three major extreme precipitation / flood events over the NEUS. By simulating this consecutive period, we can reduce the spin-up period needed. We choose the middle of this century (2050s) as the simulated future period because it's widely used and studied.

Please check whether Figure 13 is incorrect. Also, please adjust the color bars in Figures 11, 3, and S1, as some of them are not valid.

Thanks a lot for pointing this out. Figure 13 should be correct and we changed its first column's color bar to improve its appearance; however, the caption of Figure 13 has one problem – it should be "over the inner domain" instead of "over the whole domain." We also adjusted the color bars in Figures 3, 11 and S1 as you suggest.

The authors should rearrange the structures of the manuscript. The discussion is missing, maybe the result should change to result and discussion. You must buy the results from other similar studies in the discussions section. It is worth completing comparisons or differences with similar studies in other regions of the country or the world with related studies.

Thank you for the suggestion. We will restructure the paper and add a discussion section. However, because our PGW method sensitivity study is novel, it is difficult to contextualize it in the literature. In fact, we haven't found any similar papers (we even consulted with Dr. Jimy Dudhia – one of the authors of WRF and he told us that this study is very novel). But we will try to elaborate the discussion with enough reference papers.



The authors should further clarify the shortcomings and limitations of the study.
Please check the format of the references to meet the journal's requirements.

Agreed. We have identified the several shortcomings or limitations of the study and will discuss these in the text: for example, our simulation has a coarse resolution and uses cumulus scheme which reduce the model performance; the number of flood events studied is limited.