Review of "*The influences of incorporating dynamical external forcing in WRF v3.8.1 on regional climate simulation in China*"

By: Jinming Feng, Meng Luo, Jun Wang, Yuan Qiu, Qizhong Wu, and Ke Wang

This study investigates the influence of varying external climate forcings on regional climate simulations over China. The authors modified parts of the WRF v3.8.1 code to include varying external forcing for greenhouse gases, anthropogenic aerosols, and volcanic aerosols, instead of using the climatological means that are usually prescribed. The authors then compared simulations with climatological means for all forcing, varying all external forcing or varying one forcing component and keeping climatological means for the others. They also investigated the influence of nudging and the choice of parameterization schemes on these simulations. The authors recommended varying all external forcing components for regional climate model simulations as well as preforming long-term preliminary tests to determine whether to consider nudging or not.

In my opinion, the presented results lack robustness, owing to the limited number of simulations performed. Also, for a Development and Technical paper, it lacks enough technical details, especially on the specific changes made to the WRF code, the specific parameters/variables modified in the code, details about the nudging, etc. These needs to be addressed thoroughly before the paper is further considered for GMD.

Main comments:

My main concern, which is a fairly serious one, is the lack of robustness in the results presented. This stems from inconsistency in the experimental designs. There are 7 P1 simulations (the control and 6 experiments). But there are just 2 for P2 simulations. Even when we look at the P1 simulations only, there are many inconsistencies in the experimental designs. For example, there is a P1-GHG (with nudging) and a P1-GHG-NNG (without nudging). Why did the authors not use the same convention to design the other experiments? Consequently, the impact of nudging (for the P1 simulations) is effectively investigated only for P1-GHG but not the others. However, the influence of nudging could be different for different model configurations and choice of parameterization schemes (e.g., Song et al. (2011), Wootten et al. (2016)). Therefore, the results are thus not robust enough to be generalized the way they are presented in the paper.

In my opinion, further simulations are needed to investigate the robustness of the results presented. Otherwise, there should enough compelling reasons in the text to address the inconsistency in the experimental design, although this may not be enough to address the lack of robustness of the results.

Additionally, the lack of ensemble simulations for each configuration makes it impossible to investigate any influence of the model's internal climate variability. Separating the internal variability from the forced external signals could be essential to could also be key to understand the role of anthropogenic climate change in the reported trends of temperature and precipitation (Frankcombe et al. 2015).

Specific comments:

- The authors should mention the values used for the different nudging parameters.

- L69: change resulted to result.

- L70: remove quiet.

- L81: change resulted to result.

- L82: change region to regions.

- L84: "pay cautious" to "be cautious".

- L87: remove were

- L91: change are to is

- L91-93: Following from my main comment above: The authors said that "China is one of the most densely populated areas in the world, and its monsoon climate are strongly modulated by internal climate variabilities ….". If this is the case, why was there not any investigation with different ensemble members of the same configuration to investigate the impact of the internal variation on the results?

- L103: change focus to focusing

- L113-117: Here the authors start by saying "The present version of WRF has only considered ...." and the end by saying "...physics schemes in WRF v3.8.1". This makes one think WRF v3.8.1 is the current version of WRF, which is obviously not the case (I believe there is a WRF v4.5 currently). I think this needs to be clarified. It may also be necessary to mention why WRF v.3.8.1 is used rather than the current version. And if there are any differences, the authors should briefly describe those as well.

- L122-124: What is the criteria for considering water-friendly aerosols and ice-friendly aerosols as anthropogenic and natural aerosols respectively? If there any studies that has shown this, please cite them.

- L133: Please give examples of studies that have used those two datasets for research and model evaluation.

- L159-161: This is only true for the southeastern part of the domain. I think it should be mentioned.

- L180: external

- L184: weaken

- L185-187: This is very well known already. Different parameterization schemes can considerably alter simulation results.

- L189-193: Here the authors suggest that "spectral nudging can improve temperature trend simulation…". In my opinion, this is not robust since the authors did not investigate it for the other configurations.

- L196-198: This will be clearer/robust if there is a P2-GHG and a P2-GHG-NNG as well.

- L208-209: I am not sure if this is entirely true. In my opinion, this is true when considering only the sign of the trend but not the magnitude. I think the forcing configurations can have a considerable localized influence on the magnitude of the trend. If you zoom into the southern areas of the box shown on the P1 figures, you can notice some changes in the magnitudes of the trend. This should be clarified.

- L232-234: A robust conclusion cannot be made based on just the 2 choices used here. I suggest the authors revise this so that it does not seem like this result is generalizable.

- L246: change "So, the large…" to "So, a large…"

- L254: cautions to caution

- L250: (63.4 mm/10a) Is this value reported in Table 2? Because I cannot find it. I see -20.7 for P2-CTRL. Please check and clarify.

- L256-257: The authors should make it clear which data source is being spoken about here. Is it the observations? If yes, which one? They should also ensure that the values reported in the text correspond with those in the Table.

- L279: region to regions

- L294: remove in

- L330: Did the authors mean Clausius-Clapeyron?

- L339: area to areas

- L345: contributed to contribute

- L362-364: I think this sentence should be revised, and possibly broken down into two, as it is a bit difficult to understand.

- L364: remove obviously

- L370-371: The reason why there needs to be more ensemble members to evaluate internal variability. This should be addressed somehow in the text at least.

- L377: change overestimate to overestimation

- L386: remove seems

- L396: change records to record

- L412: change resulted to result

References

Frankcombe, L. M., England, M. H., Mann, M. E., & Steinman, B. A. (2015). Separating internal variability from the externally forced climate response. *Journal of Climate*, *28*(20), 8184-8202. https://doi.org/10.1175/JCLI-D-15-0069.1

Song, S., Tang, J. & Chen, X. Impacts of spectral nudging on the sensitivity of a regional climate model to convective parameterizations in East Asia. *Acta Meteorol Sin* **25**, 63–77 (2011). https://doi.org/10.1007/s13351-011-0005-z

Wootten, A., Bowden, J. H., Boyles, R., & Terando, A. (2016). The sensitivity of WRF downscaled precipitation in Puerto Rico to cumulus parameterization and interior grid nudging. *Journal of Applied Meteorology and Climatology*, *55*(10), 2263-2281. https://doi.org/10.1175/JAMC-D-16-0121.1