"The influences of incorporating dynamical external forcing in WRF v3.8.1 on regional climate simulations in China" by Feng et al.,

The authors modified the WRF v3.8.1 to include the spatial-temporal varying external forcing components to study the impact of dynamical forcing on the long-term simulation in China. Though the objective of the work is fascinating, I have some serious concerns about the WRF modifications and the experimental design.

1. One of the major points of this work is the development of WRF to include the external forcing components in WRF. But authors spent just one paragraph (Model improvement) under the 'Model and Data' section, which is also very general. They did not spend a single sentence about their improvements compared to the current version of the WRF (WRF v4.5.1). Why did the authors modify the WRF v3.8.1 is also unclear and explained in the manuscript? WRF v3.8.1 was released in Aug 2016, so why we need to improve that 7-year-old version compared to the current version is also not explained.

2. My second primary concern of this manuscript is its experimental design. The authors recommended judging the need for the nudging technique before the formal simulation, but their experimental design for nudging is surprising to me. Under the P1 subset, authors performed a control simulation, followed by all forcings and individual forcings of GHGs, Aerosol, and Volcano. All the experiments mentioned above were performed using Nudging techniques, but suddenly then, they performed only a single simulation without Nudging experiments (for GHGs). What is the scientific basis for choosing GHGs forcing without Nudging experiments? Why not choose all forcing experiments for Nudging justification? I recommend that authors perform without Nudging experiments, at least for all forcings and Aerosol forcings on top of GHG forcings. Also, for the P2 experiment, they only performed a control simulation and nudging experiment, but their major objective was to see the impact of external forcings.

3. Panel 'b' in Figures 3 – 6 shows the combined effects of GHG and Aerosol on the temperature, precipitation, and circulation trend. I want another experiment by adding GHG and Aerosol as external forcings. I believe that impact might have differed from panel 'b' in Figures 3 – 6 since they are not in a linear relationship in the actual

atmosphere. We might not speculate some conclusions by linearly adding to independent single forcing simulations specifically for precipitation.

My other points are also listed below.

4. This point is related to my first major concern. How is your modification different from the most recent WRF version (WRF 4.5.1) regarding aerosol contribution through Thompson Microphysics and RRTMG radiation scheme?

5. Authors must provide more detail about their modification and sensitivity with and without modification experiments, at least for a single sensitivity.

6. Why are authors using the ERA40 reanalysis, not the most recent ERA5, to force the WRF model? ERA5 provides much better spatial and temporal resolution compared to ERA40.

7. Similar to the above point, why authors are using CMIP5 forcings instead of CMIP6 data? These can make a significant difference in their simulations.

8. Figure S4 is very much ambiguous, and it's difficult to quantify which scheme performs better other than by eye estimation. Authors should do a more quantitative way to choose the best scheme, and authors can try the Taylor Diagram for U, Geopotential, Temperature, and Precipitation.

9. To quantify the Nudging effect robustly, authors must perform without nudging experiments for all forcings (P-MFC-NNG) and aerosol forcing (P1-AERO-NNG). This will help them to quantify the nudging impact.

10. For precipitation trend analysis, how reliable are CRU and ERA40 precipitation data?