Response to Referee # 3

Dear reviewer,

Thank you for the comments to help improve the quality of the paper. We have revised the manuscript to address your comments and a detailed response to each comment is provided in this file. The comments are in regular font and the responses are in red.

RC1, Anonymous Referee #3

Reviewer suggestion: Accept

The authors presented a novel technique to improve air quality predictions that can be utilized for health effect studies in China. A WRF/CMAQ modeling pair has been employed to simulate air quality for 2013 using four different anthropogenic emissions inventories that are publically available. The emissions inventories included the Multi-resolution Emission Inventory for China (MEIC), the Emission Inventory for China by the School of Environment of Tsinghua University (SOE), the Emissions Database for Global Atmospheric Research (EDGAR) and the Regional Emission inventory in Asia version 2 (REAS2). The entire year was simulated independently using four different anthropogenic emissions inventories along with the same biogenic emissions processed by MEGAN at a 36-km horizontal resolution to encompass entire China, South Asia and parts of East Asian countries. The model performance was evaluated using observed data available from 422 sites across 60 cities throughout China. In general, the WRF/CMAQ pair predicted ozone (O3) and particulate matter (PM2.5) concentrations within the standards for model performance criteria, however, significant difference also exist depending on location, and time of the year. In order to calculate ensemble concentrations, predicted pollutant concentrations from each set of model run were linearly combined in such a way to minimize the sum of the squared errors between ensemble concentrations and observations from all 60 cities of interest in the study. The statistics such as the mean fractional bias (MFB) and mean fractional errors (MFE) of the predicted ensemble annual PM2.5 seem to improve in all 60 cities compared to statistics for individual emissions inventories. Similarly, performance statistics for ensemble concentrations also improved for hourly, daily, and annual concentrations of O3 for all 60 cities in China.

I believe the authors have correctly identified one of the major problems with accurate air quality predictions, i.e. lack of accurate emissions inventories particularly available to public in China. In order to overcome such issue the authors have come up with an ensemble approach. Although ensemble averaging of predicted concentrations is a common approach in literature, the approach described in the current study is unique. I believe the manuscript has enough scientific merits to be accepted and published in the ACPD. I, however, have the following suggestions that the authors may consider:

1. There is no need for such details in the title. The authors may consider changing the title to "Ensemble prediction of air quality using the WRF/CMAQ modeling system for health effects studies in China".
Response: We accepted the reviewer's suggestion and changed the title to "Ensemble prediction of air quality using the WRF/CMAQ modeling system for health effects studies in China"

2. There are minor grammatical mistakes that authors can fix in the later versions of the manuscript. This is just a suggestion in advance.
Response: We went through the manuscript carefully for several times and corrected the typos, mistakes, and grammar errors.

3. Table S1: It is better to define seasons rather than providing the names of the months for each emissions inventory.
Response: We provided season names instead of months in the revised file.

4. Ammonia emissions seem to have the highest unexplained variability, how would it change the prediction of

PM2.5 mass concentrations, if the standard deviation were likely to be lower?

Response: Ammonia emissions have high variability among different inventories, reflecting high uncertainties in the current estimation for ammonia emission in China. Source apportionment studies have shown that ammonia emissions account for over 10% of $PM_{2.5}$ total mass in China (Shi et al., 2017). The difference in modeled $PM_{2.5}$ concentrations using different inventories is partially due to the ammonia emissions. In addition, since it also affects the formation of secondary nitrate and sulfate, large variations in NH_3 emission could potentially have large impacts on $PM_{2.5}$. More consistent $PM_{2.5}$ predictions would be expected if the difference of ammonia emissions in different inventories was smaller. No changes were made to the manuscript.

5. What is the property of the weighting factor in the ensemble concentration calculation? How is it affected by the sample size?

Response: The weight factor w is set to in the range of [0, 1] with w=0 represents no influence of the individual simulation on the ensemble prediction, and w=1 indicates that concentrations of the individual simulation are fully accounted in the ensemble prediction. The weighting factors for the ensemble predictions are based on minimizing the overall difference in the ensemble predictions with observations. By increasing the sample size (i.e. number of cities with observations) and their spatial coverages, the capability of each individual simulation to reproduce the magnitude and spatial variation of the $PM_{2.5}$ concentrations can be more accurately represented, and thus the weighting factors can better represent the actual strength of each inventory in predicting regional concentrations.

Above discussion was added in Section 2.5.