Response to RC4 from Referee # 2

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.

RC4, Anonymous Referee #2

This paper concerns a study of the performance of forecasts of air pollution in China, with focus on a large number of sites. The description of the methods and results is comprehensive. The results suggest that the method has potential to forecast air quality conditions in China and, likely, elsewhere. The paper should thus be of interest to the air quality scientific community.

However, as it stands the paper is not suitable for publication in ACP. There are two reasons for this (I note another reviewer identifies these reasons too): (i) The writing of the paper needs improving, the English needs to be checked; (ii) I cannot see much detail of how the study links to health concerns, even though health is in the title of the paper. There is discussion about the application to health issues in the conclusions, but this is cursory and has to come earlier in the paper. The authors should address these two points before publication of the paper in ACP. Furthermore, the authors should address a number of specific issues (not exhaustive), mainly concerning clarification of the text, examples of which I detail below.

Response: Thanks for the comments and suggestion.

(i) We are sorry about the gramma errors as we are eager to introduce the study. We have read the manuscript carefully and made correction to the typos and mistakes in the revised manuscript. We also revised a few long sentences and used short sentences in the revised manuscript.

(ii) This study is part of a project to investigate the long-term health impacts of the severe outdoor air pollution in China. This is the first part of the series study aiming to provide more accurate air pollution exposure assessment for the health analysis. The predicted air pollution fields then will be used in a number of epidemiology studies. The first such analysis used the annual $PM_{2.5}$ ensemble predictions to investigate the premature mortality attributable to various sources of $PM_{2.5}$ in China and the responses of premature mortality to the $PM_{2.5}$ reduction objectives in different regions of China. The paper has been accepted for publication in *Environmental Science & Technology* (Hu et al., 2017). A few studies are undergoing to analyze the correlations between air pollutants and certain health outcomes in China using the ensemble predictions of gaseous pollutants, PM mass and compositions.

A few epidemiology groups also expressed their interest of using the ensemble predictions of $PM_{2.5}$ and O_3 for some short-term health effect studies in China. Thus, we are confident to the applications of the products from this study.

We added a brief discussion on the current and future applications of our dataset for health effect studies in China at the end of Section 3.3. We also carefully addressed the specific issues the reviewers listed below to improve the manuscript.

Specific comments

L. 135: Indicate here what you will discuss in each section of the paper. Response: We added a brief description of the structure of the paper in the end of the Introduction section.

L. 349: It would be helpful to remind the reader of the location of the stations, instead of just using the acronym. Response: We moved Figure S1 from the supplemental materials to the main context, so that the readers can understand the locations of the regions.

Table 3: Should the weights add up to 1? They do not for, e.g., for the annual case.

Response: From the mathematical point of view, it is not necessary to constrain the weighting factors or the sum of weighting factors. We choose to limit each weighting factor in the range of [0,1] to ensure that ensemble predictions maintain positivity and do not grow to large unrealistic values in the entire domain. Enforcing a unit constrain on the sum of the weighting factors further limits the overall ensemble prediction to be within the range of individual simulations. However, this could unnecessary limit the capability of the ensemble for regions where higher ensemble values can lead to smaller overall error. Thus, such a constraint was not applied in this study.

Table 4: Could authors condense the information? For example, at how many stations is the ensemble prediction better or worse?

Response: We moved Table 4 to the supplemental materials as Table S3. We have briefly summarized the information in the manuscript in section 3.3: "The results show that the ensemble predictions are better than those with EDGAR, MEIC, REAS2 and SOE at 36, 37, 32 and 40 cities for $PM_{2.5}$, and 39, 39, 43, and 38 cities for O_3 -1h, respectively. The ensemble predictions are better than ≥ 2 of the individual predictions at 45 and 41 cities for $PM_{2.5}$ and O_3 -1h, respectively."

Figure 2: The authors need to explain more in the caption what the lines represent. For example, there are two solid and dotted lines in the panels – do they represent a standard deviation about a mean?

Response: The solid and dotted lines in the panels are the model performance criteria and goals, as indicated in the key caption. We added the definitions of "criteria" and "goal" in the manuscript and the figure caption of Figure 3 in the revised manuscript (Figure 2 in the original manuscript):

"The model performance goals represent the level of accuracy that is considered to be close to the best a model can be expected to achieve, and the model performance criteria represent the level of accuracy that is considered to be acceptable for modeling applications."

Figure 3: Indicate in the caption what the horizontal and vertical panels represent. Same for figures 4, 5 and 7. Response: We added descriptions about the horizontal and vertical panels in the figure captions for Figures 4, 5, and 6 in the revised manuscript (Figures 3, 4, and 5 in the original manuscript). Figure 7 in the original manuscript (Figure 8 in the revised manuscript) illustrates the concentrations of PM_{2.5} and its components. Each panel is labeled with the species name, so no explanation was added for it.

Reference:

Hu, J., Huang, L., Chen, M., Liao, H., Zhang, H., Wang, S., Zhang, Q., and Ying, Q.: Premature Mortality Attributable to Particulate Matter in China: Source Contributions and Responses to Reductions, Environ Sci Technol, 10.1021/acs.est.7b03193, 2017.