Dear Editor and Reviewers,

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.

RC2, Anonymous Referee #4
General Comment
The authors showed an inclusive validation about the ability of CMAQ model to simulate the air pollutants (O3 and PM2.5) in China with using four different EI data in recent year (2013). They used the widely-used statistical indices for the validation and observations which covered wide areas in China. An ensemble method to obtain better prediction of air pollutants in China was proposed which is the main part of this paper. This paper is well within the scope of this journal, however, I noticed several issues in this paper which cannot be passed over to be published. I suggested that the authors should consider the following comments: two major and several specific comments.

Major Comment 1:
My biggest concern is the lack of carefulness in the manuscript. Several typos, mistakes in table and figure, and the insufficient explanations can be found which make the manuscript difficult to read and greatly damage the value of this paper. I pointed out some of those points in the specific comment below,

and I strongly suggest that the authors consider those comments and should carefully and thoroughly check the manuscript again before revised submission.

Response: Thanks for pointing out the typos and mistakes. We have checked the manuscript carefully and made correction to the typos and mistakes in the revised manuscript. The changes can be found in the manuscript with changes marked.

Major Comment 2:

The authors set a goal of this paper on proposing a method for using the model simulation to health impact study and so the authors put "for health effect study" in the title. However, it was not clear which part of the manuscript was particularly dedicated for the health effect study. I concerned if the indices of air pollutants used in the manuscript: daily, monthly, and annual means, 1hourly and 8hourly O3, are appropriate for this purpose. I think more sentences is necessary to discuss the validity of those indices to be used for the health impact research, if they want to claim it as, at least, a part of heath effect study.

Response: 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 health analysis. The predicted air pollution fields then will be used in a number of epidemiology studies. Actually, the first such analysis using the annual ensemble $PM_{2.5}$ 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 was recently 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 expressed their interest of using the ensemble predictions of $PM_{2.5}$ and O_3 from this study for short-term health effect studies in China. Therefore, we also evaluated the performance of daily and monthly ensemble predictions for both $PM_{2.5}$ and 1h- and 8h- O_3 in this manuscript so that it can provide a validation for future applications for such dataset.

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.

Specific Comments:

- Model description: There was no descriptions about the model domain. Figure S1 can be moved from the supplement to the manuscript since the abbreviation for the different regions in China were frequently used in the manuscript.

Response: We moved Figure S1 from the supplement to the manuscript.

- E1-E4: How did you treat the observation from 422 sites? Are these data once averaged out to form the city average for each of 60 cities, and then calculate the statistical indices (MNB, MNE, MFB, MFE)? Please make it clearly described in the manuscript.

Response: Yes, the city averages were firstly calculated by averaging the observations in all the sites located in that city, and then the statistical indices were calculated based on the city averages. We added above information in the revised manuscript.

- L249-251: It is better to briefly describe the reason why different statistical indices are used for O3 and PM2.5.

Response: In air quality modeling studies, it has been common to use MNB and MNE to evaluate the model performance for O_3, and use MFB and MFE to evaluate the model performance for $PM_{2.5}$. And accordingly the MNB and MNE criteria and goals have been set for O_3, and MFB and MFE criteria and goals for $PM_{2.5}$. We added above information in the revised manuscript.

- E6: A brief explanation of the method to minimize the function Q is necessary.

Response: The linear least square solver 'lsqlin' in matlab was used to minimize the function Q. This information was added in the revised manuscript.

- Table 1: Are these statistical indices calculated using annual mean? not clearly described.

Response: The original statistics in Table 1 were calculated using hourly average concentrations. We clearly added this information in the table caption.

- L286-288: The description here is inconsistent with Figure1. Is this sentence correct?

Response: The description here is about the 'overall' performance, i.e., the average indices over the entire modeling period and over the entire regions of China. Figure 1 shows the performance in different months and regions. Therefore, there seems some difference, but we double checked the numbers, they are correct.

- L295: Why were January and February omitted?

Response: The national air quality monitoring network started publishing ambient air quality observations since March 2013. Therefore, no observations were available for January and February in 2013.

- L300-301: I couldn't understand the meaning of this sentence. Are there any typo or mistake?

Response: We corrected the sentence to "O_3 predicted using MEIC, EDGAR, and REAS2 meets the performance criteria in most regions except for the YRD by MEIC and the PRD by EDGAR."

- L302-304: It is difficult to see what this sentence said from in Figure 1.

Response: We modified and expanded the sentence to be clearer: "CO and NO_2 are under-predicted in all regions, with the largest under-predictions in NW and Other. This pattern is similar among the results with all inventories. SO_2 is generally under-predicted in all regions, but over-predicted in the Sichuan Basin (SCB) by all inventories. SO_2 is also over-predicted by EDGAR in the PRD region. SO_2 in Northeast (NE) is substantially under-predicted by MEIC and REAS2. In general, model performance in the more developed regions such as YRD, NCP, and PRD are relatively better, compared to NW and Other regions."

- Figure 2: The explanation to properly see this figure is highly insufficient. What does the x-axis stand for? Is it the absolute concentration of observation or simulation? Furthermore, "goal" and "criteria" in Figure 2 should be explained somewhere in the manuscript. Otherwise the readers cannot take the messages properly from this figure.

Response: The x-axis shows the observed $PM_{2.5}$ and PM_{10} concentrations. We added definitions of "goal" and "criteria" in 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."

- L311: typo?, a period -> comma?

Response: corrected.

- Figure 3: Are these indices (O3-1h, -8h) maximum 1h- or 8h- mean concentration in a day (=daily maximum 1h or 8h-mean O3)? If so, should be more clearly stated.

Response: We added the definitions for O3-1h (daily maximum 1h O3) and O3-8h (daily maximum 8h mean O3) in the figure caption.

- L327-328: I don't think so. There were large differences between SOE and MEIC over the oceanic area east of China.

Response: The O_3 difference between SOE and MEIC is generally less than 1ppb over the oceanic area east of China, indicated by the 'green' color (the color scheme is shown in the bottom of the figure). To be more accurate, we added the "(the difference is generally less than 1ppb)" in the sentence.

- L344: typo?, South Asia -> Southeast Asia

Response: We corrected it to Southeast Asia.

- L354 & L361: What is NCY?

Response: We corrected it to NCP.

- L362 typo?, YRD -> PRD?

Response: We corrected it to PRD.

- Table2: This is too detailed information. It can be moved to supplement.

Response: We moved Table 2 to the supplemental materials as Table S2

- L410-412: Why are the values referred here as the MFB of individual simulation (-0.25 – -0.16) different from those appeared in Table 1 (-0.32 – -0.21)? If the definitions are different for both, it should be clearly written in the manuscript. I really confused here.

Response: Following the discussion of annual average concentration in Table 2, the values in L410-412 refer to the MFB and MFE calculated using the annual averages. The MFB and MFE values in Table 1 were calculated using the hourly averages. We clearly clarify the calculation of the values in the paragraph.

- L412-413 Something wrong with English.

Response: We corrected "and" to "any" in the sentence.

- L413-415: Same as the two comments above, why are the values of MNB of individual simulation (0.06 – 0.19) different from those appeared in Table 5?

Response: Again, the values in L413-415 were calculated using annual averages, while the values in Table 5 were calculated using the daily averages. We clearly clarify the calculation of the values in the paragraph.

- Table 3: The authors showed that the weighting factor of each EI can vary for different averaging time. in general, EDGAR and REAS have large weight for daily and monthly, and the other two Chinese EI were weighted large for annual time scale. I encouraged the authors to discuss more on the interpretation of it.
Response: The weighting factors in different averaging times were determined by the model performance. The model performance in different averaging times was affected by the total emission rates, temporal profiles (which assigned the annual total emission rates into different months/days). The results probably indicate the annual total emission rates of MEIC and SOE were accurate but the temporal profiles were not as good as the ones in EDGAR and REAS.

We added above discussion in the revised manuscript.

- Table 4: This is also too detailed information. If you only want to say how many cities out of 60 can improve their prediction with ensemble and do not intend to describe its regional differences, this table can be moved to supplement and it is enough to briefly describe the result in the manuscript.
Response: We moved Table 4 to the supplemental materials as Table S3 and only brief description was kept in the manuscript.

- Table 5: This table showed that the weighting factor can vary large depending on the region. Table 3 demonstrated the factor also change for different averaging time scale. And the factor may be different for the different year. The purpose of this study is proposing an ensemble method for obtaining the better air pollutants concentration data for health effect estimation, from this point of view, how do the authors think the best way to calculate the weighting factor in China? Need some more sentences on it.
Response: Even though the weighted factors vary depending on the regions, averaging times and different years, the ensemble method that we proposed in this study is to minimize the difference between predictions and observations and can be applied in different regions with different averaging time scales, and for any years. The ensemble analysis is a post-process method to improve the agreement between predictions and observations in any averaging time scales, as shown in the manuscript. The way to calculate the weighting factors depends on the objectives of specific studies. But in general, more observation data used in the calculation, more accurate the ensemble prediction would be.

We added above discussion in the revised manuscript.

RC3, Anonymous Referee #3

Reviewer suggestion: Accept after revision

General comments:
This study is somewhat comprehensive; the modeled output has been compared with observations adequately. I think there is enough scientific merit in the manuscript, and so I would recommend it be accepted after minor revision. I encourage the authors to pay attention to the following comments:

Major comments:

The manuscript needs to be carefully revised. In general, there are places where it is difficult to follow what the authors are trying to convey to the readers. It suffers from lack of flow, perhaps, because of typos, wrong expressions, and many grammatical mistakes. The authors may consider to pay more attention to the construction of sentences and read the manuscript carefully to avoid typos. The results section is well described; however, I would request the authors to use short sentences to avoid getting the readers lose their track of what was said in the beginning of a sentence.

Response: We have read the manuscript carefully and made correction to the typos and mistakes in the revised manuscript. We revised a few long sentences and used short sentences in the revised manuscript.

Specific comments:

1. P3L58: The first sentence of the manuscript is inaccurate. Correct English is “China has been suffering from ….” The authors may also want to rewrite it.

Response: In the first round of revision for ACPD, we have revised this sentence to “Large population in China has been exposed to severe air pollution….”.

2. P3L64-65: …threatens public health in this country (Which country is this?). I am guessing the authors wanted to say …threatens public health in China.

Response: We corrected the sentence to “threatens public health in China”.

3. P3L69: What are “central monitor” measurements? Please do not assume the readers will have an idea about it. It would be better if the authors explain it in a nutshell.

Response: we added an explanation of “Ambient air quality is usually measured at monitoring sites and used to represent the exposure of population in the surrounding areas of the sites”.

4. P3L77: Omit “the” in front of meteorological fields.

Response: Corrected.

5. P3L86-87: ..large uncertainties remain. Correct: …large uncertainties still remain.

Response: Corrected.

6. P3L89-90: ..,and the efficiency of emissions controls and their fractional penetrations into the industries. The authors got me lost here. Please make it clear about the intent of this sentence. It is hard to get the meaning out it.

Response: We deleted “and their fractional penetrations into the industries” from the sentence to avoid confusion.

7. P5L150: It has been showed that these…, correct present particle is “shown”.

Response: Corrected.

8. P5L171-172: I see that the description of the techniques for emission estimates are somewhat referred for reading, but I would urge the authors to describe the “technology-based uncontrolled” and “penetrations of control technologies” terms in some plain language so that the readers have some understanding of these terms without having to look into the referenced materials.

Response: “technology-based” and “uncontrolled” are two separate description words for emission factors, the former means the emission factors are different for different facilities using different technologies with same fuels. Uncontrolled means before control since the control effects were added later. “penetrations of control technologies” means the fraction of pollutant not collected, in comparison to efficiency.
We have added “and” between “technology-based” and “uncontrolled” and “fractions of pollutants not collected” after “penetrations” in parentheses.

9. P5L187-188: The S1 Table contains the emissions summary for a “typical workday” in season? What about the weekend? How does the weekend emissions vary from a weekday?
Response: This table is to show the differences between inventories. Weekly factors we used to apply the emission to workday and weekend day are certain, which gives the ratios of workday to weekend day of 1.3-1.5. Thus, all inventories will have lower values in weekend day and the relative differences among different inventories do not change. We added explanation to the caption.

10. P6L212: Reference needed for MEGANv2.1 biogenic emissions processor.
Response: A reference was added for MEGANv2.1 (Guenther et al., 2012).

11. P6L218: “In-line” is a one word. Correct it in the manuscript.
Response: Corrected.

12. P6L222: Reference for WRFv3.6.1 is needed.
Response: A reference was added for WRFv3.6.1 (Skamarock et al., 2008).

13. P6L232: Maybe it will be a good idea, if the authors summarize boundary concentrations for major species and put them in a table in the supplemental.
Response: The boundary concentrations for a given species vary in locations (latitude, longitude, and altitude). It is impossible to summarize them in a table. We clearly stated in the manuscript that the initial and boundary conditions were generated using the CMAQ default profiles.

14. P9L325: REAS2 predicted O3-hr values are lower….. Needs attention, comparative sentence missing “than”.
Response: We added “than MEIC” in the sentence.

15. P15: The authors may consider using unabbreviated forms of the performance matrices so that the readers can follow easily in the conclusion section as these are defined way earlier in the manuscript.
Response: We appreciated the reviewer’s suggestion. In our previous experience, we have been advised to keep abbreviations in the conclusion to avoid duplicate definitions. Since the performance matrices have been discussed multiple times in the results section, we feel it is appropriate to keep the abbreviations in the conclusion section. But we are willing to change to use the full names if the editor also think it is necessary.

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 $\geqslant 2$ 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.

References

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., and Wang, X.: The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions, Geosci Model Dev, 5, 1471-1492, 10.5194/gmd-5-1471-2012, 2012.

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.

Shi, Z., Li, J., Huang, L., Wang, P., Wu, L., Ying, Q., Zhang, H., Lu, L., Liu, X., Liao, H., and Hu, J.: Source apportionment of fine particulate matter in China in 2013 using a source-oriented chemical transport model,

Sci Total Environ, 601, 1476-1487, http://dx.doi.org/10.1016/j.scitotenv.2017.06.019, 2017.
Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W., and Powers, J. G.: A Description of the Advanced Research WRF Version 3, NCAR Technical Note NCAR/TN-475+STR, 2008.

Ensemble Prediction~~s~~ of Air Quality ~~Pollutants in China in 2013 for Health Effects Studies~~ Using the WRF/CMAQ Modeling System for Health Effects Studies in China~~with Four Emission Inventories~~

Jianlin Hu[1], Xun Li[1], Lin Huang[1], Qi Ying[2,1], Qiang Zhang[3], Bin Zhao[4], Shuxiao Wang[4], Hongliang Zhang[5,1*]

[1] Jiangsu Key Laboratory of Atmospheric Environment Monitoring and Pollution Control, Jiangsu Engineering Technology Research Center of Environmental Cleaning Materials, Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, School of Environmental Science and Engineering, Nanjing University of Information Science & Technology, 219 Ningliu Road, Nanjing 210044, China
[2] Zachry Department of Civil Engineering, Texas A&M University, College Station, TX 77843-3136
[3] Ministry of Education Key Laboratory for Earth System Modeling, Center for Earth System Science, Tsinghua University, Beijing, China
[4] State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing 100084, China
[5] Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA 77803

[*]Corresponding author:
Hongliang Zhang, Email: hlzhang@lsu.edu. Phone: +1-225-578-0140.

Abstract

Accurate exposure estimates are required for health effects analyses of severe air pollution in China. Chemical transport models (CTMs) are widely used ~~tools~~ to provide ~~detailed information of~~ spatial distribution, chemical composition, particle size fractions, and source origins of air pollutants. The accuracy of ~~CTMs'~~ air quality predictions in China is largely affected by the uncertainties of public available emission inventories. The Community Multi-scale Air Quality ~~model~~ (CMAQ) model with meteorological inputs from the Weather Research and Forecasting ~~model~~ (WRF) model were used in this study to simulate air ~~quality~~ pollutants in China in 2013. Four sets of simulations were conducted with four different anthropogenic emission inventories, including the Multi-resolution Emission Inventory for China (MEIC), the Emission Inventory for China by School of Environment at Tsinghua University (SOE), the Emissions Database for Global Atmospheric Research (EDGAR), and the Regional Emission inventory in Asia version 2 (REAS2). Model performance was evaluated against available observation data from 422 sites in 60 cities across China. Model predictions of O_3 and $PM_{2.5}$ ~~with the four inventories~~ generally meet the model performance criteria ~~of model performance~~, but performance difference exists in different pollutants and ~~different~~ regions among ~~the~~ inventories. Ensemble predictions were calculated by linearly combining the results from different inventories ~~under the constraint that of minimized~~ to minimize the sum of the squared errors between the ensemble results and the observations from all the cities ~~was minimized~~. The ensemble ~~annual~~ concentrations show improved agreement with observations in most cities. The mean fractional bias (MFB) and mean fractional errors (MFE) of the ensemble ~~predicted~~ annual $PM_{2.5}$ at the 60 cities are -0.11 and 0.24, respectively, which are better than the MFB (-0.25 – -0.16) and MFE (0.26 – 0.31) of individual simulations. The ensemble annual daily maximum 1-hour ~~peak~~ O_3 (O_3-1h) concentrations are also improved, with mean normalized bias (MNB) of 0.03 and mean normalized errors (MNE) of 0.14, compared to MNB of 0.06 – 0.19 and MNE of 0.16 – 0.22 of the individual predictions. The ensemble predictions agree better with observations with daily, monthly, and annual averaging times in all regions of China for both $PM_{2.5}$ and O_3-1h. The study demonstrates that ensemble predictions by combining predictions from individual emission inventories can improve the accuracy of predicted temporal and spatial distributions of air pollutants. This study is the first ensemble model study in China using multiple emission inventories and the results are publicly available for future health effects studies.

Key words: chemical transport model; emission inventory; ensemble; China; $PM_{2.5}$

1. Introduction

~~Large~~A significant portion of the population in China has been exposed to severe air pollution in recent decades as the consequence of intensive energy use without efficient control measures. Based on ambient air pollution data published by the China National Environmental Monitoring Center (CNEMC), most of the major cities are in violation of the Chinese Ambient Air Quality Standards g~~G~~rade II standard (35 μg m^{-3}) for annual average particulate matter with diameter of 2.5 μm or less ($PM_{2.5}$) (Zhang and Cao, 2015;Wang et al., 2014b), with a mean population weighted $PM_{2.5}$ concentration of over 60 μg m^{-3} during 2013-2014. Long-term exposure to such high levels of $PM_{2.5}$ greatly threatens public health in ~~this country~~China. Recent studies have suggested that approximately more than one million premature deaths can be attributed to outdoor air pollution each year in China (Lelieveld et al., 2015;Liu et al., 2016;Hu et al., 2017a).

Accurate exposure estimates are required in health effects studies. Ambient air quality is usually measured at monitoring sites ~~Central monitor measurements are usually~~ and used to represent the ~~in~~ exposure of population in the surrounding areas~~ of the sites~~~~assessment~~.~~,~~ ~~but~~ However, a routine central monitoring network in China has just been ~~built up~~established ~~from~~since 2013, and is still limited in spatial coverage and lack of detailed information of the chemical composition, PM size fractions, and source origins of air pollutants. Chemical transport models (CTMs) have been widely used in health effects studies to overcome the limitations in central monitor measurements for exposure estimates (Philip et al., 2014;Lelieveld et al., 2015;Liu et al., 2016;Laurent et al., 2016a;Laurent et al., 2016b;Ostro et al., 2015). However, the accuracy of ~~CTMs~~ the predictions from CTMs is largely affected by the accuracies of the emission inventories (Wang et al., 2010), ~~the~~ meteorological fields (Hu et al., 2010), and numerical solutions to the equations that describe various atmospheric processes (Hu et al., 2006;Yu et al., 2005). Emission inventories are indispensable tools for a wide range of environmental activities from management of chemicals to the prevention of air pollution. Several emission inventories have been created ~~for~~to cover China. Different emission inventories focus on specific geographical regions in the urban, regional (Zhao et al., 2012;Zhang et al., 2008), national or continental (Zhang et al., 2009;Kurokawa et al., 2013) scales; and/or focus on pollutants from individual (Su et al., 2011;Ou et al., 2015) and specific sectors (Zhao et al., 2008;Xu et al., 2017).

Despite the great efforts in improving the accuracy of emission inventories in China, large uncertainties still remain. Generally, the emissions of pollutants are estimated as the product of activity levels (such as industrial production or energy consumption), unabated emission factors (i.e. mass of emitted pollutant per unit activity level), and the efficiency of emission controls~~ and their fractional penetrations into the industries~~. Large uncertainties are associated with activity levels, emission source fractions, and emission factors (Akimoto et al., 2006;Lei et al., 2011a). The uncertainties are especially significant for some pollutants, such as ammonia (NH_3) and volatile organic compounds (VOCs). For example, ~~it is shown that~~ for a Pearl River Delta (PRD) inventory in 2006, SO_2 emission has low uncertainties of -16%~21% from power plant sources quantified by Monte Carlo simulations~~. However,~~, while NO_x has medium to high uncertainties of -55%~150% and VOC, CO, and PM have even higher uncertainties (Zheng et al., 2009). For an inventory for the Yangtze River Delta (YRD) region, the overall uncertainties for CO, SO_2, NOx, PM_{10}, $PM_{2.5}$, VOCs, and NH_3 emissions are ±47.1 %, ±19.1 %, ±27.7 %, ±117.4 %, ±167.6 %, ±133.4 %, and ±112.8 %, respectively (Huang et al., 2011). A comprehensive quantification study by Zhao et al. (2011) using Monte Carlo simulations showed that the uncertainties of Chinese emissions of SO_2, NOx, $PM_{2.5}$, BC, and OC in 2005 are −14%~13%, −13%~37%, −17%~54%, −25%~136%, and −40%~121%, respectively.

The uncertainties in emission inventories are carried into CTMs simulations, leading to biases in air quality predictions, which need to be carefully evaluated to identify the useful information ~~that can be used in~~ for health effects studies (Hu et al., 2016b;Hu et al., 2014c;Hu et al., 2014b;Hu et al., 2015b;Tao et al., 2014). An evaluation of one-year air pollutants predictions using the Weather Research and Forecasting (WRF) / Community Multi-scale Air Quality (CMAQ) modeling system with the Multi-resolution Emission Inventory for China (MEIC) has been reported (Hu et al., 2016a). The model predictions of O_3 and $PM_{2.5}$ generally agree with ambient measured concentrations, but the model performance varies in different regions and seasons. In some regions, such as the northwest of China, the model significantly under-predicted $PM_{2.5}$ concentrations.

The technique of ensemble is often used to reduce uncertainties in model predictions by combining multiple sets of predictions. This technique has been widely used in the climate predictions (Murphy et al., 2004;Tebaldi and Knutti, 2007), and recently adopted in air quality predictions (Delle Monache et al., 2006;Huijnen et al., 2010). A recent study has compared a few anthropogenic emission inventories in China during 2000-2008 (Saikawa et al., 2016), but detailed evaluation of ~~air quality~~ model results based on these inventories ~~for over an extended time period~~ have not been performed. The methods to utilize the strength of different emission inventories to get improved air quality predictions for China have not been reported in the literature. The aim of this study is to create an improved set of air quality predictions in China by using an ensemble technique. First, four sets of one-year air quality predictions were conducted with the WRF/CMAQ modeling system with four different anthropogenic emission inventories for China ~~for the entire year of~~in 2013. In addition to MEIC, the three other emission inventories are the Emissions Database for Global Atmospheric Research (EDGAR), Regional Emission inventory in Asia version 2 (REAS2), and Emission Inventory for China developed by School of Environment at Tsinghua University (SOE). The model performance on $PM_{2.5}$ and O_3 concentrations in 2013 with different emission inventories was then evaluated against available observation data in 60 cities in China. The differences among air quality predictions were also compared and identified. Finally, an ensemble technique was developed to minimize the bias of model predictions and to create improved exposure predictions. To the authors' best knowledge, this is the first ensemble model study in China using multiple emission inventories. The ensemble predictions of this study are available for public health effects analyses.

This paper is organized as follows. The CMAQ model, emissions and other inputs for the model, observational datasets used for model performance evaluation, and the method for ensemble calculation are described in Section 2. Section 3 discusses the model performance on gaseous and particulate pollutants simulated with four emission inventories, as well as the performance of the ensemble predictions in different regions/cities and with different averaging times. ~~Then~~At last, the major findings are summarized in the Conclusion section.

2. Method

2.1 Model description

In this study, the applied CMAQ model is based on CMAQ v5.0.1 with changes to improve the model's performance in predicting secondary organic and inorganic aerosols. The details of these changes could been found in previous studies (Hu et al., 2016a;Hu et al., 2017b) and the references therein, therefore only a brief description is summarized here~~ and more details can be found in the cited publications and the references therein~~. The gas phase photochemical mechanism SARPC-11 was modified to better treat isoprene oxidation chemistry (Ying et al., 2015;Hu et al., 2017b). Formation of secondary organic aerosol (SOA) from reactive uptake of dicarbonyls, methacrylic acid epoxide, and isoprene epoxydiol through surface pathway (Li et al., 2015;Ying et al., 2015) was added. Corrected SOA yields due to vapor wall-loss (Zhang et al., 2014) were adopted. Formation of secondary nitrate and sulfate through heterogeneous reactions of NO_2 and SO_2 on particle surface (Ying et al., 2014) ~~were~~ was also incorporated. It has been ~~showed~~ shown that these modifications improved the model performance on secondary inorganic and organic $PM_{2.5}$ components.

2.2 Anthropogenic emissions

The CMAQ model was applied to China ~~with~~and surrounding countries in East Asia using the horizontal resolution of 36-km. The modeling domain is shown in Figure 1. The anthropogenic emissions are from four~~ different~~ inventories: MEIC, SOE, EDGAR, and REAS2. MEIC was developed by a research group in Tsinghua University (http://www.meicmodel.org). Compared with other inventories for China, e.g. INTEX-B (Zhang et al., 2009) or TRACE-P (Streets et al., 2003), the major improvements include a unit-based inventory for power plants (Wang et al., 2012) and cement plants (Lei et al., 2011b), a county-level high-resolution vehicle inventory (Zheng et al., 2014), and a novel NMVOC speciation approach (Li et al., 2014). The VOCs were speciated to the SAPRC-07 mechanism. As the detailed species to model species mapping of the SAPRC-11 mechanism is essentially the same as the SAPRC-07 mechanism (Carter and Heo, 2012), the speciated VOC emissions in the MEIC inventory were directly used in the simulation.

The SOE emission inventory was developed using an emission factor method (Wang et al., 2011;Zhao et al., 2013). The sectorial emissions in different provinces were calculated based on activity data, technology-based and uncontrolled emissions factors, and penetrations (fractions of pollutants not collected) of control technologies. Elemental carbon (EC) and organic carbon (OC) emissions were calculated based on $PM_{2.5}$ emissions and their ratios to $PM_{2.5}$. The sectorial activity data and technology distribution were obtained using an energy demand modeling approach with various Chinese statistics and technology reports. More details, including the spatio-temporal distributions and speciation of NMVOC emissions, can be found in previous publications (Zhao et al., 2013;Wang et al., 2011;Bin et al., 2013). Since MEIC and SOE emission inventories only cover China, emissions from outside China countries and regions were based on REAS2 (Kurokawa et al., 2013).

The version 4.2 of EDGAR emission (http://edgar.jrc.ec.europa.eu/overview.php?v=42) has a spatial resolution of $0.1^{o}\times0.1^{o}$. The EDGAR inventory contains annual emissions from different sectors based on IPCC designations. REAS2 has a spatial resolution of $0.25^{\circ}\times0.25^{\circ}$ for the entire

Asia. The inventory contains monthly emissions of pollutants from different source categories. Saikawa et al. (2016) compared the major features of different anthropogenic emission inventories for China. Detailed information regarding these inventories can be found in the publications presenting them. Table S1 shows the total emissions of major pollutants within China in a typical workday of each season. In general, large differences exist among different inventories for China. MEIC has the highest CO emissions in ~~January~~ winter while REAS2 has the highest in other ~~three~~ seasons. MEIC has the highest NO_x emissions while REAS2 has the highest emissions of VOCs in all ~~months~~seasons. EDGAR predicts the highest SO_2 emissions, which are approximately a factor of two higher than those estimated by SOE. SOE has highest NH_3 emissions while EDGAR has much lower NH_3 emissions than the other three. EDGAR also has lowest EC and OC emissions, but the total $PM_{2.5}$ emissions are the highest. Standard deviations ~~(SD)~~ indicate that ~~January~~ winter has the largest uncertainties for all species except SO_2 and NH_3. ~~January~~ Winter has the smallest SO_2 uncertainties while ~~July~~ summer has the largest NH_3 uncertainties.

All the emissions inventories were processed with an in-house program and re-gridded into the 36-km resolution CMAQ domain when necessary. Representative speciation profiles based on the SPECIATE 4.3 database maintained by U.S. EPA were applied to split NMVOC of EDGAR and REAS2 into SAPRC-11 mechanism~~.~~ and $PM_{2.5}$ of all inventories was split ~~also speciated~~ into AERO6 species~~ using profiles from the SPECIATE 4.3 database~~. Monthly emissions were temporally allocated into hourly files using temporal allocation profiles from previous studies (Chinkin et al., 2003; Olivier et al., 2003; Wang et al., 2010a). More details regarding EDGAR can be found in Wang et al. (2014a), while those for REAS2 can be found in Qiao et al. (2015).

2.3 Other inputs

The Model for Emissions of Gases and Aerosols from Nature (MEGAN) v2.1 was used to generated biogenic emissions (Guenther et al., 2012). The 8-day Moderate Resolution Imaging Spectroradiometer (MODIS) leaf area index (LAI) product (MOD15A2) and the plant function type (PFT) files used in the Global Community Land Model (CLM 3.0) were applied to generate inputs to MEGAN. The readers are referred to Qiao et al. (2015) for more information. Open biomass burning emissions were generated using a satellite observation based fire inventory developed by NCAR (Wiedinmyer et al., 2011). The dust emission module was updated to be compatible with the 20-category MODIS land use data (Hu et al., 2015a) for in−line dust emission processing and sea salt emissions were also generated during CMAQ simulations.

The meteorological inputs were generated using WRF v3.6.1 (Skamarock et al., 2008). The initial and boundary conditions to WRF were downloaded from the NCEP FNL Operational Model Global Tropospheric Analyses dataset. WRF configurations details can be found in Zhang et al. (2012). WRF performance has been evaluated by comparing predicted 2m above surface temperature and relative humidity, and 10m wind speed and wind direction with all available observational data at ~1200 stations from the National Climate Data Center (NCDC). The model performance is generally acceptable and detailed evaluation results can be found in a previous study (Hu et al., 2016a).

The initial and boundary conditions representing relatively clean tropospheric concentrations were generated using CMAQ default profiles.

2.4 Model evaluation

Model predictions with the four emission inventories were evaluated against available observation data in China. Hourly observations of $PM_{2.5}$, PM_{10}, O_3, CO, SO_2, and NO_2 from March to December 2013 at 422 stations in 60 cities were obtained from CNEMC (http://113.108.142.147:20035/emcpublish/) but~~as~~ no observations were available for January and ~~Feburary~~February. Observations at multiple sites in one city were averaged to calculate the average concentrations of the city. Detailed quality control of the data can be found in previous studies (Hu et al., 2016a;Hu et al., 2014a;Wang et al., 2014b). Statistical matrix of mean normalized bias (MNB), mean normalized error (MNE), mean fractional bias (MFB) and mean fractional error (MFE) were calculated using the Equations (E1)-(E4):

$$MNB = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{C_m - C_o}{C_o}\right) \tag{E1}$$

$$MNE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{C_m - C_o}{C_o}\right| \tag{E2}$$

$$MFB = \frac{1}{N}\sum_{i=1}^{N}\frac{\left(C_m - C_o\right)}{\left(\frac{C_o + C_m}{2}\right)} \tag{E3}$$

$$MFE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|C_m - C_o\right|}{\left(\frac{C_o + C_m}{2}\right)} \tag{E4}$$

where C_m and C_o are the predicted and observed city average concentrations, respectively, and N is the total number of observation data. MNB and MNE are commonly used in evaluation of model performance of O_3, and MFB and MFE are commonly used in evaluation of model performance of $PM_{2.5}$ (Tao et al., 2014). The U.S. EPA previously recommended O_3 model performance criteria of within ± 0.15 for MNB and less than 0.30 for MNE (as shown in Figure 1) and PM model performance criteria of within ± 0.60 for MFB and less than 0.75 for MFE (EPA, 2001b). Figure 2 includes the criteria and goals for PM as a function of PM concentration, as suggested by Boylan and Russell (2006), which have been widely used in model evaluation.

2.5 Ensemble predictions

The four sets of predictions with different inventories were combined linearly to calculate the ensemble predictions, as shown in Equation (E5):

$$C^{pred,ens} = \sum_{m=1}^{N_m} w_m C^{pred,m} \tag{E5}$$

where $C^{pred,ens}$ is the ensemble prediction~~s~~, $C^{pred,m}$ is the predicted concentration from the m^{th} simulation, N_m is the number of simulations in the ensemble (N_m=4), and w_m is the weighting factor of the mth simulation. The weighting factor for each set of predictions was determined by minimizing the objective function Q ~~as shown~~ in Equation (6):

$$Q = \sum_{i}^{N_{city}} \left[C_i^{obs} - \sum_{m=1}^{N_m} w_m C_i^{pred,m} \right]^2 \quad \text{(E6)}$$

where C_i^{obs} is the observed $PM_{2.5}$ or O_3 concentration at the ith city, N_{city} is the total number of cities with observation (N=60), $C_i^{pred,m}$ is the predicted concentration at the ith city from the mth simulation, and N_m is the number of simulations in the ensemble (N_m=4~~),~~). The weight factor ~~and~~ w_m~~'s~~ of the m_{th} simulation~~are weighting factors~~ to be determined ~~under the constraints that 0<w<1~~is within 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 observations data were the same as used in the model evaluation. Ensemble predictions were performed for $PM_{2.5}$ and O_3 in this study. A MATLAB program was developed to solve above equation and determine the weighting factors using the linear least square solver "lsqlin".

3. Results

3.1 Model performance on gaseous and particulate pollutants

Table 1 summarizes the overall model performance on O_3, CO, NO_2, SO_2, $PM_{2.5}$, and PM_{10} with different inventories using the averaged observations in 60 cities in 2013. ~~The U.S. EPA previously recommended O_3 model performance criteria of within ± 0.15 for MNB and less than 0.30 for MNE (as shown in Figure 1) and PM model performance criteria of within ± 0.60 for MFB and less than 0.75 for MFE (EPA, 2001b). Figure 2 includes the criteria and goals for PM as a function of PM concentration, as suggested by Boylan and Russell (2006), which have been widely used in model evaluation.~~ Model performance meets the O_3 criteria for all inventories. O_3 from SOE are 7.2 parts per billion (ppb) lower than the mean observed concentration while the under-predictions of the other three inventories are less than 2 ppb. CO, NO_2, and SO_2 are under-predicted by all inventories, indicating potential uncertainties in the inventories. CO predictions from three inventories (SOE inventory does not include CO) are substantially lower than observations, with the best performance (lowest MNB and MNE) from REAS2. NO_2 overall performance is similar to CO~~;~~. ~~h~~However, MEIC and SOE yield the lowest MNB, and EDGAR yields the highest. SO_2 performance is better than CO and NO_2, and MEIC and SOE yield the lowest MNB, while MNE values of the four inventories are very similar. $PM_{2.5}$ and PM_{10} predictions using all inventories meet the performance criteria with similar MFB and MFE values. REAS2 generally yields slightly better $PM_{2.5}$ and PM_{10} performance, but all inventories under-predict the concentrations generally.

The difference in model performance with the four inventories also varies seasonally and spatially. Figure ~~1~~2 shows the comparison of model performance for hourly gaseous species (O_3, CO, NO_2, and SO_2) in each month from March to December 2013. The MNB values of O_3 in most months are within the criteria for all inventories except for SOE, which under-predicts O_3 concentrations. March has the worst performance for all inventories with MNE values larger than 0.4 for MEIC, SOE, and EDGAR. No significant performance difference among different inventories in different months is found, but large difference exists in various regions of China (see the definition of regions of China in Figure ~~S~~1). O_3 predicted using MEIC, EDGAR~~SOE~~, and REAS2 meets the performance criteria in most regions except for ~~the~~ YRD ~~region~~ by MEIC and PRD by EDGAR. O_3 predicted using SOE only meets the criteria in Northwest (NW) and other region (Other) of China. ~~For~~ CO~~,~~ and NO_2~~,~~ are under-predicted in all regions, with the largest under-predictions in NW and Other. This pattern is similar among the results with all inventories. ~~and~~ SO_2 is generally under-predicted in all regions except, but over-predicted in the Sichuan Basin (SCB) by all inventories. SO_2 is also over-predicted by EDGAR in the PRD region. SO_2 in Northeast (NE) is substantially under-predicted by MEIC and REAS2.~~,~~ ~~m~~In general, model performance in the ~~less~~ more developed regions such as YRD, NCP, and PRD are relatively better, compared to NW and Other regions ~~central (CNT), NW, and Other regions is worse compared to more developed regions~~.

Figure ~~2~~3 illustrates the $PM_{2.5}$ and PM_{10} performance statistics of MFB and MFE as a function of absolute concentrations in different months of 2013 and in different regions. $PM_{2.5}$ predictions based on each inventory are within the performance goal of MFB and between the goal and criteria of MFE in all months. There is no significant difference among inventories. Half of monthly averaged PM_{10} MFB values fall within the goal while the rest are between the goal and criteria. MFE values of PM_{10} are all between the goal and criteria. From the regional perspective~~.~~, $PM_{2.5}$ performance in NE by SOE is out of the MFB criteria, while that in ~~Sichuan Basin (~~SCB~~)~~ by MEIC, SOE, and REAS2 are out of the MFE criteria. MFB values of PM_{10} at all regions meet the criteria except NW, ~~where which is largely affected~~due to under-estimation of ~~by~~ windblown dust emissions in NW.

3.2 Spatial variations in predicted gaseous and particulate pollutants

Figure ~~3~~ 4 shows the spatial distribution of annual averaged gas species, daily maximum 1-hour ~~peak~~ O_3 (O_3-1h), and 8-hour me~~n~~an O_3 (O_3-8h), NO_2, and SO_2 predicted by MEIC and differences between SOE, EDGAR, and REAS2 to MEIC. MEIC ~~predicted~~ predicts annual O_3-1h concentrations are ~60ppb in most parts of China with the highest values of ~70ppb in SCB. SOE predicts lower O_3-1h values ~~for all the domain~~than MEIC, with about 5 ppb differences in the SCB, central China (CNT), and North China Plain (NCP) regions and 2-3 ppb differences in other regions. EDGAR also predicts 2-3 ppb lower O_3-1h in most regions than MEIC but its O_3-1h predictions in the Tibet Plateau, NCP and ocean regions are 2-3 ppb higher than MEIC predictions. REAS2 predicted O_3-1h values are lower than ~~those of~~ MEIC for scattered areas in the NE, NW, and CNT regions and other regions experience slightly higher O_3-1h. MEIC, SOE, and REAS2 have similar results out of China (the difference is generally less than 1 ppb) since the simulations used same emissions for those regions. O_3-8h shows similar spatial distributions as O_3-1h among inventories with slightly less differences. NO_2 concentrations are 10-15ppb in developed areas of the NCP and YRD regions, and greater than 5 ppb at other urban areas as predicted by MEIC. SOE predicts 2-3 ppb lower NO_2 concentrations in most areas except the vast NW region. EDGAR predicts lower NO_2 (more than 5 ppb difference) in urban areas of the NCP and YRD areas but higher concentrations in the entire west part of China by approximately 1-2 ppb. REAS2 has the closest NO_2 with MEIC as the 1-2 ppb underestimation and overestimation are almost evenly distributed in the whole country. SO_2 concentrations are up to 20ppb in the NCP, CNT, and SCB regions while are less than 5 ppb in other regions. SOE mostly predicts 2-3 ppb lower SO_2 in the east half of China with the largest difference of -10 ppb in the CNT region. EDGAR and REAS2 had very similar difference with MEIC, i.e., more than 5 ppb higher concentrations in the NCP and YRD, ~2 ppb higher concentrations in the PRD, 2-3 ppb lower concentrations in the NE and up to 5 ppb lower concentrations in the CEN and SCB.

Figure ~~4~~5 shows the seasonal distribution of $PM_{2.5}$ total mass predicted by MEIC and differences between SOE, EDGAR, and REAS2 to MEIC. In ~~the~~ spring, MEIC predicted $PM_{2.5}$ concentrations are ~50 μg m^{-3} in east and south parts of China, and Southeast Asia has the highest value of ~100 μg m^{-3}. SOE predicts 5-10 μg m^{-3} lower $PM_{2.5}$ in north China and < 5 μg m^{-3} higher values in south China and along the coastline. EDGAR predicts >20 μg m^{-3} lower values in NCP and ~10 μg m^{-3} lower values in NE, CNT, and SCB, but up to 20 μg m^{-3} higher values in PRD. REAS2 predicts higher $PM_{2.5}$ values in most parts of China except under-predictions in NE and SCB. The over-predictions in YRD and NCP ~~were~~ are up to 20-30 μg m^{-3}. In summer, the high $PM_{2.5}$ regions are much smaller compared to spring with ~50 μg m^{-3} ppb concentrations in NCP, north part of YRD and SCB and 20-30 μg m^{-3} in other parts. Generally, SOE predicts <10 μg m^{-3} lower values in most regions. EDGAR predicts lower values in NCP and SCB and 5-10 μg m^{-3} higher values in south part. REAS2 almost predicts higher values in all the regions except some scattered areas in NC~~Y~~P, YRD, and SCB.

In fall, $PM_{2.5}$ concentrations are larger than 50 μg m^{-3} in most regions except NW and are ~100 μg m^{-3} in part of NCP, CNT, and SCB. SOE predicted values are lower in north part and higher in south part. EDGAR predicts up to 30 μg m^{-3} lower values in NCP and SCB while up to 20 μg m^{-3} higher values in YRD. REAS2 again estimates close values to MEIC with less than 5 μg m^{-3} differences in most regions and up to 20 μg m^{-3} higher values in scattered areas in YRD and SCB. In winter, MEIC predicted $PM_{2.5}$ concentrations are up to 200 μg m^{-3} in NC~~Y~~P, CNT, YRD, and SCB, while P~~Y~~RD has concentrations of ~50 μg m^{-3}. SOE severely underestimates by 30 μg m^{-3} in all regions with high $PM_{2.5}$ concentrations and only coast areas experience <10 μg m^{-3} higher values. EDGAR also predicts 30 μg m^{-3} lower $PM_{2.5}$ concentrations in NE, NCP, CNT, and SCB, but the YRD region has 20 μg m^{-3} higher values. The regions with lower values by REAS2 compared to MEIC ~~are much smaller but~~ are at the ~~same~~ regions of NE, NCP, CNT and SCB, similar to EDGAR but with much smaller areas. SOE predicts higher $PM_{2.5}$ in ~~S~~the south parts of YRD and NCP ~~have higher $PM_{2.5}$ values~~ than MEIC.

Figure ~~5~~6 shows the annual averaged concentrations of $PM_{2.5}$ components predicted by MEIC and the differences between other inventories with MEIC. Annual averaged particulate sulfate (SO_4^{2-}) concentrations are 20-25 μg m^{-3} in NCP, CNT, and SCB, and about 10 μg m^{-3} in other regions in the southeast China. SOE predicts ~10 μg m^{-3} lower values in high concentration areas and 2-3 μg m^{-3} lower in other areas. EDGAR predicts ~5 μg m^{-3} higher SO_4^{2-} in southeast China and 2-3 μg m^{-3} lower values in SCB. REAS2 predicted SO_4^{2-} are generally 2-3 μg m^{-3} lower than that of MEIC in areas except the coastal areas. MEIC predicts the highest particulate nitrate (NO_3^-) concentrations of up to 30 μg m^{-3} in NCP and YRD and values in other regions are 5-10 μg m^{-3} except the northwest China. SOE predicts <5 μg m^{-3} lower values in the high concentrations areas and ~2 μg m^{-3} higher values in coastal areas. EDGAR uniformly predicts lower NO_3^- values than MEIC with the largest different of 10 μg m^{-3}. REAS2 has similar results to SOE. Particulate ammonium (NH_4^+) concentrations predicted by MEIC have a peak of 15 μg m^{-3} and are mostly less than 10 µg m^{-3} in the east and south parts of China. SOE predicts slightly lower values except for coastal areas in PRD, where 1-2 µg m^{-3} higher values are observed.

~~Elemental carbon (~~EC~~)~~ concentrations are generally low compared to other components as predicted by MEIC. The highest values are less than 10 µg m^{-3} in NCP, CNT and SCB. All other three inventories predict~~ed~~ 1-2 µg m^{-3} lower EC values throughout the country. Primary organic aerosol (POA) predicted by MEIC are 20-30 µg m^{-3} in NCP, CNT and SCB, and are ~10 µg m^{-3} in other areas in east and south parts of China. SOE predicts up to 5 µg m^{-3} higher values in most areas with scattered places with ~2 µg m^{-3} lower values compared to MEIC. EDGAR and REAS2 predict up to ~10 µg m^{-3} lower values except for coastal areas. SOA concentrations are low in north part of China and up to 10 µg m^{-3} in the whole east and south parts. All three other inventories predict ~2 µg m^{-3} lower SOA values compared to MEIC. For other implicit components (OTHER), the highest concentrations are ~15 µg m^{-3} in NW and NCP, while other regions have lower than 5 µg m^{-3} concentrations. In NW, the major sources of OTHER are windblown dust online generated by CMAQ simulations, thus almost no differences are observed among inventories. SOE and EDGAR predict lower OTHER vales in north part (~2 µg m^{-3}) and slightly higher values in south and east parts (~5 µg m^{-3}). REAS2 predicts higher OTHER values in the whole east part uniformly with up to 10 µg m^{-3} differences in NCP, YRD, and SCB regions.

Additional comparisons of model predictions in different regions and some major cities in China are shown in Figures S~~2~~1-S~~5~~4 in the Supplemental Material.

3.3 Ensemble predictions

Above analyses indicate that model performance with different inventories varies on different pollutants and in different regions. Table S2 shows the observed annual average concentrations of $PM_{2.5}$ in the 60 cities and the predictions from the four inventories as well as the weighted ensemble predictions. The weighting factors for predictions using MEIC, REAS2, SOE and EDGAR are 0.31, 0.36, 0.24 and 0.20, respectively (Table ~~3~~2). The ensemble predictions greatly reduce MFB with a value of -0.11, compared to the MFB values of -0.25 – -0.16 using the annual average concentrations in the individual simulations. Also, the ensemble predictions have an MFE value of 0.24, lower than any~~d~~ MFE values of 0.26 – 0.31 in ~~any~~ individual simulations (Figure ~~6~~7). The ensemble predictions of annual O_3-1h have the MNB and MNE of 0.03 and 0.14, improved from MNB of 0.06 – 0.19 and MNE of 0.16 – 0.22 in the individual predictions, respectively.

To further evaluate the ability of the ensemble method in improving predictions at locations where observational data are not available, ensemble predictions were made using a data withholding method. For each city, the observations at the other 59 cities were used to determine the weighting factors in E6 and the ensemble prediction at the city was calculated. Performance of the ensemble predictions at the city was calculated using the withheld observations to evaluate the performance. The evaluation process was repeated for each of the 60 cities and the performance was compared to that with individual inventories (shown in Table S3~~4~~). 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. Out of the 15 cities that the ensemble $PM_{2.5}$ is only better than one or none of the individual predictions, 10 cities have MFB within ±0.25 and MFE less than 0.25. Out of the 19 cities that the ensemble O_3-1h is only better than one or none of the individual predictions, 14 cities still have MNB within ±0.2 and MNE less than 0.2. The results demonstrate that the ensemble can improve the predictions even at locations with no observational data available.

Previous studies have revealed that CTMs predictions agree more when averaging over longer times (i.e., annual vs. monthly vs. daily averages) (Hu et al., 2014b;Hu et al., 2015b). Ensemble predictions were also calculated with daily and monthly averages for $PM_{2.5}$, in addition to the calculation with annual averages discussed above. The weighting factors and the performance of ensemble predictions are shown in Table ~~3~~2 and Figure ~~6~~7, respectively. The weighting factors vary largely with the averaging times, suggesting that the prediction optimization need to be conducted separately when using different time averages. The ensemble predictions improve the agreement with observations in all averaging time cases, with lower MNB and MNE than any of the individual predictions. In general, EDGAR and REAS have large weights for daily and monthly ensemble calculations, and MEIC and SO~~W~~E have large weights for annual ensemble calculations. This result indicates that the annual total emission rates of MEIC and SOE are likely accurate but the temporal profiles to allocate the annual total emissions rates to specific day/hours need to be improved.

Table 3~~5~~ shows the ensemble prediction performance on $PM_{2.5}$ and O_3-1h in different regions of China using the daily average observations and daily average predictions with individual inventories. The weighting factors vary greatly among regions, reflecting that substantial difference in the spatial distributions of $PM_{2.5}$ and O_3 when using different inventories. The MNB and MNE values of ensemble predictions are reduced in all regions for both pollutants, suggesting the ensemble predictions improve the accuracy and can be better used in further health effects studies. The similar findings are also found with the monthly average observations and predictions (shown in Table S4~~3~~).

Figure 8~~7~~ shows spatial distributions of $PM_{2.5}$ and its components from the ensemble predictions using the weighting factors of annual averages. The ensemble of $PM_{2.5}$ components were calculated using the same weighting factors for $PM_{2.5}$. Over 80 μg m^{-3} annual average $PM_{2.5}$ concentrations are estimated in NCP, CNT, YRD and SCB regions in 2013. Secondary inorganic aerosols (SO_4^{2-}, NO_3^-, and NH_4^+) account for approximately half of $PM_{2.5}$, and exhibit similar spatial patterns. Carbonaceous aerosols (EC, POA, and SOA) account for about 30%, but POA and SOA have quite different spatial distributions. High POA concentrations are mainly distributed in NCP, CNT and SCB, while high SOA concentrations are found in the south part of China. By considering the spatial distributions of population and ensemble $PM_{2.5}$, the population-weighted annual averaged $PM_{2.5}$ concentration in China in 2013 is 59.5 μg m^{-3}, which is higher than the estimated value of 54.8 μg m^{-3} by Brauer et al. (2016).

The products of the current study can be further applied in health effects studies. The first such analysis used the annual $PM_{2.5}$ ensemble predictions to assess~~For example,~~ the spatial distribution of excess mortality due to adult (> 30 years old) ischemic heart disease (IHD), cerebrovascular disease (CEV), chronic obstructive pulmonary disease (COPD) and lung cancer (LC) in China caused by $PM_{2.5}$ exposure (Hu et al., 2017a). Any health studies requiring human exposure information to different pollutants would benefit from this study. Even though the weighted factors vary depending on the regions, averaging times and different study years, the ensemble method proposed in this study is to minimize the difference between predictions and observations and can be applied in different studies. The way to calculate the weighting factors depends on the objectives of specific studies. But in general, more observation data used in the calculation, more accurate the ensemble prediction would be.

4. Conclusion

In this study, air quality predictions in China in 2013 were conducted using the WRF/CMAQ modeling system with anthropogenic emissions from four inventories including MEIC, SOE, EDGAR, and REAS2. Model performance with the four inventories was evaluated by comparing with available observation data from 422 sites in 60 cities in China. Model predictions of hourly O_3 and $PM_{2.5}$ with the four inventories generally meet the model performance criteria, but that model performance with different inventories varies on different pollutants and in different regions. To improve the overall agreement of the predicted concentrations with observations, ensemble predictions were calculated by linearly combining the predictions from different inventories. The ensemble annual concentrations show improved agreement with observations for both $PM_{2.5}$ and O_3-1h. The MFB and MFE of the ensemble predictions of $PM_{2.5}$ at the 60 cities are -0.11 and 0.24, respectively, which are better than the MFB (-0.25 – -0.16) and MFE (0.26 – 0.31) of any individual simulations. The ensemble predictions of annual O_3-1h have the MNB and MNE of 0.03 and 0.14, improved from MNB (0.06 – 0.19) and MNE (0.16 – 0.22) in individual predictions. The ensemble predictions with data withholding method at each city show better performance than the predictions with individual inventories at most cities, demonstrating the ability of the ensemble in improving the predictions at locations where observational data are not available. The ensemble predictions agree better with observations with daily, monthly, and annual averaging times in all regions of China. The study demonstrates that ensemble predictions by combining predictions from individual emission inventories can improve the accuracy in the concentration estimation and the spatial distributions of air pollutants. ~~The products of the current study can be further applied in health effects studies. For example, the spatial distribution of excess mortality due to adult (> 30 years old) ischemic heart disease (IHD), cerebrovascular disease (CEV), chronic obstructive pulmonary disease (COPD) and lung cancer (LC) in China caused by~~ ~~$PM_{2.5}$~~ ~~exposure (Hu et al., 2017a). Any health studies requiring human exposure information to different pollutants would benefit from this study.~~ The data presented in the paper is available for downloading via requests.

Acknowledgement

This project is partly supported by the National Natural Science Foundation of China under contract No. 41675125, Natural Science Foundation of Jiangsu Province under contract No. BK20150904, Jiangsu Distinguished Professor Project 2191071503201, Jiangsu Six Major Talent Peak Project 2015-JNHB-010, and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), Jiangsu Key Laboratory of Atmospheric

Environment Monitoring and Pollution Control of Nanjing University of Information Science and Technology, and Jiangsu Province Innovation Platform for Superiority Subject of Environmental Science and Engineering (No. KHK1201). The authors want to acknowledge the Texas A&M Supercomputing Facility (http://sc.tamu.edu) and the Texas Advanced Computing Center (http://www.tacc.utexas.edu/) for providing computing resources essential for completing the research reported in this paper.

References

Akimoto, H., Ohara, T., Kurokawa, J.-i., and Horii, N.: Verification of energy consumption in China during 1996–2003 by using satellite observational data, Atmospheric Environment, 40, 7663-7667, http://dx.doi.org/10.1016/j.atmosenv.2006.07.052, 2006.

Bin, Z., Shuxiao, W., Xinyi, D., Jiandong, W., Lei, D., Xiao, F., Jiming, H., and Joshua, F.: Environmental effects of the recent emission changes in China: implications for particulate matter pollution and soil acidification, Environmental Research Letters, 8, 024031, 2013.

Boylan, J. W., and Russell, A. G.: PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models, Atmospheric Environment, 40, 4946-4959, 2006.

Brauer, M., Freedman, G., Frostad, J., van Donkelaar, A., Martin, R. V., Dentener, F., Dingenen, R. v., Estep, K., Amini, H., Apte, J. S., Balakrishnan, K., Barregard, L., Broday, D., Feigin, V., Ghosh, S., Hopke, P. K., Knibbs, L. D., Kokubo, Y., Liu, Y., Ma, S., Morawska, L., Sangrador, J. L. T., Shaddick, G., Anderson, H. R., Vos, T., Forouzanfar, M. H., Burnett, R. T., and Cohen, A.: Ambient Air Pollution Exposure Estimation for the Global Burden of Disease 2013, Environ Sci Technol, 50, 79-88, 10.1021/acs.est.5b03709, 2016.

Byun, D., and Schere, K. L.: Review of the Governing Equations, Computational Algorithms, and Other Components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System, Applied Mechanics Reviews, 59, 51-77, 2006.

Carter, W. P. L., and Heo, G.: Development of revised SAPRC aromatics mechanisms. Final Report to the California Air Resources Board, Contracts No. 07-730 and 08-326, April 12, 2012. , 2012.

Delle Monache, L., Deng, X. X., Zhou, Y. M., and Stull, R.: Ozone ensemble forecasts: 1. A new ensemble design, J Geophys Res-Atmos, 111, 18, 10.1029/2005jd006310, 2006.

EPA, U. S.: Guidance for Demonstrating Attainment of Air Quality Goals for PM2.5 and Regional Haze. Draft 2.1, 2 January 2001, in, edited by: The US Environmental Protection Agency, O. o. A. a. R. O. o. A. P. a. S., Research Triangle, NC, 2001a.

EPA, U. S.: Guidance for Demonstrating Attainment of Air Quality Goals for PM2.5 and Regional Haze. Draft 2.1, 2 January 2001, in, edited by: The US Environmental Protection Agency, O. o. A. a. R. O. o. A. P. a. S., Research Triangle, NC, 2001b.

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., and Wang, X.: The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions, Geosci Model Dev, 5, 1471-1492, 10.5194/gmd-5-1471-2012, 2012.

Hu, J., Ying, Q., Chen, J. J., Mahmud, A., Zhao, Z., Chen, S. H., and Kleeman, M. J.: Particulate air quality model predictions using prognostic vs. diagnostic meteorology in central California, Atmos Environ, 44, 215-226, 10.1016/j.atmosenv.2009.10.011, 2010.

Hu, J., Wang, Y., Ying, Q., and Zhang, H.: Spatial and temporal variability of PM2.5 and PM10 over the North China Plain and the Yangtze River Delta, China, Atmospheric Environment, 95, 598-609, http://dx.doi.org/10.1016/j.atmosenv.2014.07.019, 2014a.

Hu, J., Zhang, H., Chen, S.-H., Vandenberghe, F., Ying, Q., and Kleeman, M. J.: Predicting Primary PM2.5 and PM0.1 Trace Composition for Epidemiological Studies in California, Environ Sci Technol, 48, 4971-4979, 10.1021/es404809j, 2014b.

Hu, J., Zhang, H., Chen, S., Ying, Q., Vandenberghe, F., and Kleeman, M. J.: Identifying PM2.5 and PM0.1 Sources for Epidemiological Studies in California, Environ Sci Technol, 48, 4980-4990, 10.1021/es404810z, 2014c.

Hu, J., Wu, L., Zheng, B., Zhang, Q., He, K., Chang, Q., Li, X., Yang, F., Ying, Q., and Zhang, H.: Source contributions and regional transport of primary particulate matter in China, Environmental Pollution, 207, 31-42, 2015a.

Hu, J., Zhang, H., Ying, Q., Chen, S.-H., Vandenberghe, F., and Kleeman, M. J.: Long-term particulate matter modeling for health effect studies in California - Part I: model performance on temporal and spatial variations, Atmos Chem Phys, 15, 3445-3461, 2015b.

Hu, J., Chen, J., Ying, Q., and Zhang, H.: One-year Simulation of Ozone and Particulate Matter in China Using WRF/CMAQ Modeling System, Atmos. Chem. Phys., 16, 10333-10350, 2016a.

Hu, J., Jathar, S., Zhang, H., Ying, Q., Chen, S. H., Cappa, C. D., and Kleeman, M. J.: Long-term Particulate Matter Modeling for Health Effects Studies in California – Part II: Concentrations and Sources of Ultrafine Organic Aerosols, Atmos. Chem. Phys. Discuss., 2016, 1-37, 10.5194/acp-2016-903, 2016b.

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, 2017a.

Hu, J., Wang, P., Ying, Q., Zhang, H., Chen, J., Ge, X., Li, X., Jiang, J., Wang, S., Zhang, J., Zhao, Y., and Zhang, Y.: Modeling biogenic and anthropogenic secondary organic aerosol in China, Atmos. Chem. Phys., 17, 77-92, 10.5194/acp-17-77-2017, 2017b.

Hu, Y. T., Odman, M. T., and Russell, A. G.: Mass conservation in the Community Multiscale Air Quality model, Atmos Environ, 40, 1199-1204, DOI 10.1016/j.atmosenv.2005.10.038, 2006.

Huang, C., Chen, C. H., Li, L., Cheng, Z., Wang, H. L., Huang, H. Y., Streets, D. G., Wang, Y. J., Zhang, G. F., and Chen, Y. R.: Emission inventory of anthropogenic air pollutants and VOC species in the Yangtze River Delta region, China, Atmos. Chem. Phys., 11, 4105-4120, 10.5194/acp-11-4105-2011, 2011.

Huijnen, V., Eskes, H. J., Poupkou, A., Elbern, H., Boersma, K. F., Foret, G., Sofiev, M., Valdebenito, A., Flemming, J., Stein, O., Gross, A., Robertson, L., D'Isidoro, M., Kioutsioukis, I., Friese, E., Amstrup, B., Bergstrom, R., Strunk, A., Vira, J., Zyryanov, D., Maurizi, A., Melas, D., Peuch, V. H., and Zerefos, C.: Comparison of OMI NO2 tropospheric columns with an ensemble of global and European regional air quality models, Atmos Chem Phys, 10, 3273-3296, 2010.

Kurokawa, J., Ohara, T., Morikawa, T., Hanayama, S., Janssens-Maenhout, G., Fukui, T., Kawashima, K., and Akimoto, H.: Emissions of air pollutants and greenhouse gases over Asian regions during 2000–2008: Regional Emission inventory in ASia (REAS) version 2, Atmos. Chem. Phys., 13, 11019-11058, 10.5194/acp-13-11019-2013, 2013.

Laurent, O., Hu, J., Li, L., Kleeman, M. J., Bartell, S. M., Cockburn, M., Escobedo, L., and Wu, J.: A Statewide Nested Case-Control Study of Preterm Birth and Air Pollution by Source and Composition: California, 2001-2008, Environ Health Persp, http://dx.doi.org/10.1289/ehp.1510133, 2016a.

Laurent, O., Hu, J. L., Li, L. F., Kleeman, M. J., Bartell, S. M., Cockburn, M., Escobedo, L., and Wu, J.: Low birth weight and air pollution in California: Which sources and components drive the risk?, Environ Int, 92-93, 471-477, 10.1016/j.envint.2016.04.034, 2016b.

Lei, Y., Zhang, Q., He, K. B., and Streets, D. G.: Primary anthropogenic aerosol emission trends for China, 1990–2005, Atmos. Chem. Phys., 11, 931-954, 10.5194/acp-11-931-2011, 2011a.

Lei, Y., Zhang, Q., Nielsen, C., and He, K.: An inventory of primary air pollutants and CO2 emissions from cement production in China, 1990–2020, Atmospheric Environment, 45, 147-154, http://dx.doi.org/10.1016/j.atmosenv.2010.09.034, 2011b.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., and Pozzer, A.: The contribution of outdoor air pollution sources to premature mortality on a global scale, Nature, 525, 367-371, 10.1038/nature15371, 2015.
Li, J., Cleveland, M., Ziemba, L. D., Griffin, R. J., Barsanti, K. C., Pankow, J. F., and Ying, Q.: Modeling regional secondary organic aerosol using the Master Chemical Mechanism, Atmos Environ, 102, 52-61, http://dx.doi.org/10.1016/j.atmosenv.2014.11.054, 2015.
Li, M., Zhang, Q., Streets, D. G., He, K. B., Cheng, Y. F., Emmons, L. K., Huo, H., Kang, S. C., Lu, Z., Shao, M., Su, H., Yu, X., and Zhang, Y.: Mapping Asian anthropogenic emissions of non-methane volatile organic compounds to multiple chemical mechanisms, Atmos. Chem. Phys., 14, 5617-5638, 10.5194/acp-14-5617-2014, 2014.
Liu, J., Han, Y., Tang, X., Zhu, J., and Zhu, T.: Estimating adult mortality attributable to PM2.5 exposure in China with assimilated PM2.5 concentrations based on a ground monitoring network, Sci Total Environ, http://dx.doi.org/10.1016/j.scitotenv.2016.05.165, 2016.
Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., and Stainforth, D. A.: Quantification of modelling uncertainties in a large ensemble of climate change simulations, Nature, 430, 768-772, 10.1038/nature02771, 2004.
Ostro, B., Hu, J., Goldberg, D., Reynolds, P., Hertz, A., Bernstein, L., and Kleeman, M. J.: Associations of Mortality with Long-Term Exposures to Fine and Ultrafine Particles, Species and Sources: Results from the California Teachers Study Cohort, Environ Health Persp, DOI:10.1289/ehp.1408565, 2015.
Ou, J., Zheng, J., Li, R., Huang, X., Zhong, Z., Zhong, L., and Lin, H.: Speciated OVOC and VOC emission inventories and their implications for reactivity-based ozone control strategy in the Pearl River Delta region, China, Science of The Total Environment, 530–531, 393-402, http://dx.doi.org/10.1016/j.scitotenv.2015.05.062, 2015.
Philip, S., Martin, R. V., van Donkelaar, A., Lo, J. W.-H., Wang, Y., Chen, D., Zhang, L., Kasibhatla, P. S., Wang, S., Zhang, Q., Lu, Z., Streets, D. G., Bittman, S., and Macdonald, D. J.: Global Chemical Composition of Ambient Fine Particulate Matter for Exposure Assessment, Environ Sci Technol, 48, 13060-13068, 10.1021/es502965b, 2014.
Qiao, X., Tang, Y., Hu, J., Zhang, S., Li, J., Kota, S. H., Wu, L., Gao, H., Zhang, H., and Ying, Q.: Modeling dry and wet deposition of sulfate, nitrate, and ammonium ions in Jiuzhaigou National Nature Reserve, China using a source-oriented CMAQ model: Part I. Base case model results, Science of The Total Environment, 532, 831-839, http://dx.doi.org/10.1016/j.scitotenv.2015.05.108, 2015.
Saikawa, E., Kim, H., Zhong, M., Zhao, Y., Janssens-Manehout, G., Kurokawa, J. I., Klimont, Z., Wagner, F., Naik, V., Horowitz, L., and Zhang, Q.: Comparison of Emissions Inventories of Anthropogenic Air Pollutants in China, Atmos. Chem. Phys. Discuss., 2016, 1-41, 10.5194/acp-2016-888, 2016.
Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W., and Powers, J. G.: A Description of the Advanced Research WRF Version 3, NCAR Technical Note NCAR/TN-475+STR, 2008.
Streets, D. G., Bond, T. C., Carmichael, G. R., Fernandes, S. D., Fu, Q., He, D., Klimont, Z., Nelson, S. M., Tsai, N. Y., Wang, M. Q., Woo, J. H., and Yarber, K. F.: An inventory of gaseous and primary aerosol emissions in Asia in the year 2000, Journal of Geophysical Research-Atmospheres, 108, 8809
10.1029/2002jd003093, 2003.

Su, S., Li, B., Cui, S., and Tao, S.: Sulfur Dioxide Emissions from Combustion in China: From 1990 to 2007, Environmental Science & Technology, 45, 8403-8410, 10.1021/es201656f, 2011.

Tao, J., Gao, J., Zhang, L., Zhang, R., Che, H., Zhang, Z., Lin, Z., Jing, J., Cao, J., and Hsu, S. C.: PM2.5 pollution in a megacity of southwest China: source apportionment and implication, Atmos. Chem. Phys., 14, 8679-8699, 10.5194/acp-14-8679-2014, 2014.

Tebaldi, C., and Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections, Philos. Trans. R. Soc. A-Math. Phys. Eng. Sci., 365, 2053-2075, 10.1098/rsta.2007.2076, 2007.

Wang, D., Hu, J., Xu, Y., Lv, D., Xie, X., Kleeman, M., Xing, J., Zhang, H., and Ying, Q.: Source contributions to primary and secondary inorganic particulate matter during a severe wintertime PM2.5 pollution episode in Xi'an, China, Atmospheric Environment, 97, 182-194, http://dx.doi.org/10.1016/j.atmosenv.2014.08.020, 2014a.

Wang, S., Xing, J., Chatani, S., Hao, J., Klimont, Z., Cofala, J., and Amann, M.: Verification of anthropogenic emissions of China by satellite and ground observations, Atmos Environ, 45, 6347-6358, http://dx.doi.org/10.1016/j.atmosenv.2011.08.054, 2011.

Wang, S. W., Zhang, Q., Streets, D. G., He, K. B., Martin, R. V., Lamsal, L. N., Chen, D., Lei, Y., and Lu, Z.: Growth in NOx emissions from power plants in China: bottom-up estimates and satellite observations, Atmos. Chem. Phys., 12, 4429-4447, 10.5194/acp-12-4429-2012, 2012.

Wang, X., Liang, X.-Z., Jiang, W., Tao, Z., Wang, J. X. L., Liu, H., Han, Z., Liu, S., Zhang, Y., Grell, G. A., and Peckham, S. E.: WRF-Chem simulation of East Asian air quality: Sensitivity to temporal and vertical emissions distributions, Atmos Environ, 44, 660-669, 2010.

Wang, Y., Ying, Q., Hu, J., and Zhang, H.: Spatial and temporal variations of six criteria air pollutants in 31 provincial capital cities in China during 2013–2014, Environment International, 73, 413-422, http://dx.doi.org/10.1016/j.envint.2014.08.016, 2014b.

Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., and Soja, A. J.: The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning, Geoscientific Model Development, 4, 625-641, 2011.

Xu, Y., Hu, J., Ying, Q., Wang, D., and Zhang, H.: Current and future emissions of primary pollutants from coal-fired power plants in Shaanxi, China, Science of the total environment, In revision, 2017.

Ying, Q., Cureño, I. V., Chen, G., Ali, S., Zhang, H., Malloy, M., Bravo, H. A., and Sosa, R.: Impacts of Stabilized Criegee Intermediates, surface uptake processes and higher aromatic secondary organic aerosol yields on predicted PM2.5 concentrations in the Mexico City Metropolitan Zone, Atmos Environ, 94, 438-447, http://dx.doi.org/10.1016/j.atmosenv.2014.05.056, 2014.

Ying, Q., Li, J., and Kota, S. H.: Significant Contributions of Isoprene to Summertime Secondary Organic Aerosol in Eastern United States, Environmental Science & Technology, 49, 7834-7842, 10.1021/acs.est.5b02514, 2015.

Yu, S. C., Dennis, R., Roselle, S., Nenes, A., Walker, J., Eder, B., Schere, K., Swall, J., and Robarge, W.: An assessment of the ability of three-dimensional air quality models with current thermodynamic equilibrium models to predict aerosol NO3-, J Geophys Res-Atmos, 110, 2005.

Zhang, H., Li, J., Ying, Q., Yu, J. Z., Wu, D., Cheng, Y., He, K., and Jiang, J.: Source apportionment of PM2.5 nitrate and sulfate in China using a source-oriented chemical transport model, Atmospheric Environment, 62, 228-242, 10.1016/j.atmosenv.2012.08.014, 2012.

Zhang, Q., Wei, Y., Tian, W., and Yang, K.: GIS-based emission inventories of urban scale: A case study of Hangzhou, China, Atmos Environ, 42, 5150-5165, http://dx.doi.org/10.1016/j.atmosenv.2008.02.012, 2008.

Zhang, Q., Streets, D. G., Carmichael, G. R., He, K. B., Huo, H., Kannari, A., Klimont, Z., Park, I. S., Reddy, S., Fu, J. S., Chen, D., Duan, L., Lei, Y., Wang, L. T., and Yao, Z. L.: Asian emissions in 2006 for the NASA INTEX-B mission, Atmospheric Chemistry and Physics, 9, 5131-5153, 2009.

Zhang, X., Cappa, C. D., Jathar, S. H., McVay, R. C., Ensberg, J. J., Kleeman, M. J., and Seinfeld, J. H.: Influence of vapor wall loss in laboratory chambers on yields of secondary organic aerosol, Proceedings of the National Academy of Sciences, 111, 5802-5807, 10.1073/pnas.1404727111, 2014.

Zhang, Y.-L., and Cao, F.: Fine particulate matter (PM2.5) in China at a city level, Scientific Reports, 5, 14884, 10.1038/srep14884 http://www.nature.com/articles/srep14884#supplementary-information, 2015.

Zhao, B., Wang, P., Ma, J. Z., Zhu, S., Pozzer, A., and Li, W.: A high-resolution emission inventory of primary pollutants for the Huabei region, China, Atmos. Chem. Phys., 12, 481-501, 10.5194/acp-12-481-2012, 2012.

Zhao, B., Wang, S., Wang, J., Fu, J. S., Liu, T., Xu, J., Fu, X., and Hao, J.: Impact of national NOx and SO2 control policies on particulate matter pollution in China, Atmospheric Environment, 77, 453-463, http://dx.doi.org/10.1016/j.atmosenv.2013.05.012, 2013.

Zhao, Y., Wang, S., Duan, L., Lei, Y., Cao, P., and Hao, J.: Primary air pollutant emissions of coal-fired power plants in China: Current status and future prediction, Atmospheric Environment, 42, 8442-8452, http://dx.doi.org/10.1016/j.atmosenv.2008.08.021, 2008.

Zhao, Y., Nielsen, C. P., Lei, Y., McElroy, M. B., and Hao, J.: Quantifying the uncertainties of a bottom-up emission inventory of anthropogenic atmospheric pollutants in China, Atmospheric Chemistry and Physics, 11, 2295-2308, 10.5194/acp-11-2295-2011, 2011.

Zheng, B., Huo, H., Zhang, Q., Yao, Z. L., Wang, X. T., Yang, X. F., Liu, H., and He, K. B.: High-resolution mapping of vehicle emissions in China in 2008, Atmos. Chem. Phys., 14, 9787-9805, 10.5194/acp-14-9787-2014, 2014.

Zheng, J., Zhang, L., Che, W., Zheng, Z., and Yin, S.: A highly resolved temporal and spatial air pollutant emission inventory for the Pearl River Delta region, China and its uncertainty assessment, Atmospheric Environment, 43, 5112-5122, http://dx.doi.org/10.1016/j.atmosenv.2009.04.060, 2009.

Table 1. Overall model performance of gas and PM species in 2013 using different inventories. Obs is observation, MFB is mean fractional bias, MFE is mean fractional error, MNB is mean normalized bias, and MNE is mean normalized error. The indices ~~are~~were calculated with hourly observations and predictions.

		Prediction	MFB	MFE	MNB	MNE
O_3	Mean Obs: 51.70 ppb					
	MEIC	49.83	-0.08	0.35	0.02	0.33
	SOE	44.51	-0.2	0.38	-0.09	0.32
	EDGAR	49.82	-0.04	0.28	0.03	0.28
	REAS2	51.17	-0.04	0.33	0.05	0.33
CO	Mean Obs: 0.96 ppm					
	MEIC	0.31	-0.92	0.96	-0.57	0.63
	SOE	/	/	/	/	/
	EDGAR	0.23	-1.12	1.16	-0.66	0.73
	REAS2	0.42	-0.72	0.82	-0.41	0.59
NO_2	Mean Obs: 21.45 ppb					
	MEIC	10.12	-0.79	0.93	-0.41	0.66
	SOE	11.59	-0.65	0.81	-0.33	0.61
	EDGAR	6.82	-1.02	1.07	-0.6	0.67
	REAS2	9.3	-0.81	0.92	-0.46	0.63
SO_2	Mean Obs: 17.21 ppb					
	MEIC	12.5	-0.51	0.87	0.01	0.87
	SOE	12.76	-0.44	0.83	0.06	0.86
	EDGAR	15.86	-0.16	0.73	0.31	0.88
	REAS2	15.15	-0.23	0.74	0.23	0.86
$PM_{2.5}$	Mean Obs: 70.01 μg m^{-3}					
	MEIC	56.39	-0.32	0.64	-0.02	0.63
	SOE	59.77	-0.24	0.61	0.09	0.67
	EDGAR	52.59	-0.3	0.59	-0.05	0.56
	REAS2	60.35	-0.21	0.59	0.08	0.63
PM_{10}	Mean Obs: 118.61 μg m^{-3}					
	MEIC	62.7	-0.63	0.79	-0.32	0.61
	SOE	63.32	-0.6	0.76	-0.3	0.6
	EDGAR	55.76	-0.67	0.78	-0.38	0.58
	REAS2	71.41	-0.49	0.7	-0.21	0.59

~~Table 2. Predicted annual average $PM_{2.5}$ concentrations at 60 cities using different anthropogenic emission inventory, and weighted ensemble based on linear combination of the four simulations, along with observed concentrations. Units are μg m^{-3}.~~

City	MEIC	SOE	EDGAR	REAS2	ENSEMBLE	Observation
Shijiazhuang	86.3	94.2	73.3	104.2	102.1	148.5
Baoding	113.4	79.5	78.1	115.1	111.9	127.9
Handan	97.9	89.7	89.7	135.2	119.0	127.8
Hengshui	103.2	87.9	88.5	113.3	112.3	120.6
Tangshan	84.8	90.6	52.8	80.7	88.2	114.2
Jinan	96.3	95.0	86.6	119.0	113.5	114.0
Langfang	93.5	72.6	70.6	82.2	90.7	113.8
Xi'an	70.9	87.5	69.6	66.1	81.4	104.2
Zhengzhou	107.5	90.5	92.9	105.4	112.3	102.4
Tianjin	84.0	73.8	84.4	95.9	95.7	95.6
Cangzhou	90.6	73.2	67.8	87.4	91.3	93.6
Beijing	62.2	59.2	77.3	71.6	75.2	90.1
Wuhan	94.4	98.5	102.9	89.9	106.7	94.0
Chengdu	52.9	67.8	50.0	52.6	62.0	86.3
Wulumuqi	22.0	39.1	20.1	32.4	32.1	85.2
Hefei	86.6	84.4	74.9	88.5	94.5	84.9
Huai'an	72.4	65.7	66.1	75.8	79.3	80.8
Changsha	87.9	109.8	70.7	82.2	98.0	79.1
Wuxi	64.6	65.6	63.6	74.3	75.7	75.8
Harbin	59.4	150.6	58.6	47.2	84.1	75.7
Nanjing	79.1	79.5	88.9	94.8	96.1	75.3
Xuzhou	100.6	85.8	101.3	102.3	109.6	74.9
Taiyuan	64.5	67.8	61.0	78.1	77.0	74.2
Huzhou	52.8	57.8	63.5	68.2	67.9	73.5
Shenyang	97.3	101.5	75.0	111.6	110.3	72.7
Yangzhou	74.7	67.2	71.5	78.9	82.5	71.1
Suqian	78.1	66.0	69.1	81.9	83.9	70.7
Nantong	77.1	58.8	60.9	70.0	75.9	70.2
Changchun	60.5	55.2	44.3	49.7	59.2	69.2
Nanchang	53.6	82.6	61.6	114.9	90.4	69.1
Jinhua	34.2	39.5	45.2	45.6	45.8	69.0
Lianyungang	66.5	55.0	56.6	66.7	69.6	68.0
Lanzhou	22.9	18.0	28.6	24.1	26.0	67.1
Suzhou	58.2	74.1	69.6	86.0	81.2	67.1
Jiaxing	60.2	59.9	66.4	70.0	71.9	66.9
Quzhou	31.0	34.3	39.5	38.4	39.8	66.5
Shaoxing	47.1	54.0	58.3	59.9	61.2	66.4
Hangzhou	47.2	58.8	63.0	64.6	65.0	66.1
Qinhuangdao	65.5	50.4	39.6	53.9	60.2	65.2
Chongqing	89.2	90.5	80.5	88.5	98.0	63.9
Xining	11.2	11.2	16.3	13.6	14.4	63.2
Qingdao	66.0	62.8	59.5	66.6	71.9	61.7

~~Shanghai~~	~~51.4~~	~~50.0~~	~~65.2~~	~~61.8~~	~~63.6~~	~~60.7~~
~~Huhehaote~~	~~27.5~~	~~20.1~~	~~18.6~~	~~21.5~~	~~25.0~~	~~59.1~~
~~Wenzhou~~	~~26.1~~	~~33.2~~	~~45.3~~	~~47.0~~	~~42.3~~	~~56.5~~
~~Nanning~~	~~37.0~~	~~43.4~~	~~45.0~~	~~43.8~~	~~47.0~~	~~54.7~~
~~Taizhou~~	~~71.3~~	~~62.3~~	~~66.7~~	~~72.1~~	~~76.8~~	~~53.0~~
~~Guangzhou~~	~~31.2~~	~~46.2~~	~~58.1~~	~~35.3~~	~~45.4~~	~~52.2~~
~~Chengde~~	~~40.0~~	~~35.1~~	~~35.0~~	~~49.7~~	~~46.0~~	~~51.5~~
~~Dalian~~	~~41.5~~	~~46.1~~	~~34.4~~	~~52.9~~	~~50.1~~	~~50.7~~
~~Guiyang~~	~~48.9~~	~~60.9~~	~~46.2~~	~~50.8~~	~~57.7~~	~~49.4~~
~~Lishui~~	~~26.2~~	~~30.7~~	~~37.4~~	~~36.9~~	~~36.5~~	~~47.9~~
~~Yinchuan~~	~~18.7~~	~~27.0~~	~~18.6~~	~~19.9~~	~~23.3~~	~~43.7~~
~~Shenzhen~~	~~23.0~~	~~32.8~~	~~45.2~~	~~24.4~~	~~33.1~~	~~39.7~~
~~Zhuhai~~	~~24.0~~	~~32.2~~	~~47.9~~	~~31.4~~	~~36.3~~	~~37.9~~
~~Kunming~~	~~29.4~~	~~32.8~~	~~28.0~~	~~31.8~~	~~34.3~~	~~35.5~~
~~Fuzhou~~	~~22.6~~	~~30.8~~	~~44.0~~	~~27.1~~	~~33.2~~	~~33.2~~
~~Zhoushan~~	~~24.4~~	~~24.1~~	~~26.8~~	~~29.3~~	~~29.4~~	~~32.1~~
~~Lasa~~	~~3.0~~	~~3.4~~	~~3.8~~	~~3.6~~	~~3.9~~	~~26.0~~
~~Haikou~~	~~21.2~~	~~28.2~~	~~29.9~~	~~24.8~~	~~28.4~~	~~25.6~~

Table ~~3~~2. The weighting factors (w) of each inventory in the ensemble predictions of $PM_{2.5}$ when using daily, monthly, or annual averages in the objective function (E5).

	Daily	Monthly	Annual
MEIC	0.07	0.13	0.31
SOE	0.14	0.16	0.24
EDGAR	0.38	0.23	0.20
REAS2	0.49	0.63	0.36

~~Table 4. Comparison of the data-withholding ensemble prediction of $PM_{2.5}$ and O_3-1h at each city with predictions of individual inventories. The ensemble predictions at each city are calculated by using the data in the other 59 cities (i.e., withholding the data at that city) to determine the ensemble weighting factors. Symbol '×' indicates the ensemble prediction performance is better than the performance of a specific inventory (i.e., both MFB (MNB) and MFE (MNE) values are smaller for $PM_{2.5}$ (O_3-1h)); otherwise symbol '–' indicates the ensemble prediction performance is worse.~~

	$PM_{2.5}$				O_3-1h			
City	MEIC	SOE	EDGAR	REAS2	MEIC	SOE	EDGAR	REAS2
Shijiazhuang	×	×	×	–	×	×	×	×
Baoding	–	×	×	–	–	–	–	×
Handan	×	×	×	–	×	×	×	×
Hengshui	×	×	×	–	–	–	–	–
Tangshan	–	–	×	×	–	×	–	–
Jinan	×	×	×	×	–	–	–	–
Langfang	–	×	×	×	×	×	×	×
Xi'an	×	–	×	×	×	×	×	×
Zhenzhou	–	×	–	–	×	–	×	×
Tianjing	×	×	×	×	–	×	–	–
Wuhan	–	–	–	–	×	×	×	×
Cangzhou	×	×	×	×	–	×	–	–
Beijing	×	×	–	×	×	×	×	×
Chendu	×	–	×	×	–	–	–	–
Wulumuqi	×	–	×	–	×	–	×	×
Hefei	–	–	×	–	×	–	×	×
Huai'an	×	×	×	×	–	–	–	–
Changsha	–	×	–	–	×	×	×	×
Wuxi	×	×	×	×	–	–	–	–
Harbin	–	×	–	×	×	×	×	×
Nanjing	–	–	–	–	×	×	×	×
Xuzhou	–	–	–	–	–	–	–	–
Taiyuan	×	×	×	×	–	–	–	×
Huzhou	×	×	×	–	×	×	×	×
Shenyang	–	–	–	–	×	×	×	×
Yangzhou	–	–	–	–	×	×	×	×
Suqian	–	–	–	–	×	×	×	×
Nantong	×	×	×	–	×	×	×	×
Changchun	–	×	×	×	–	–	–	–

~~Nanchang~~	–	–	–	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~
~~Jinghua~~	~~×~~	~~×~~	~~×~~	–	~~×~~	~~×~~	~~×~~	~~×~~
~~Lianyungang~~	–	~~×~~	~~×~~	–	~~×~~	–	~~×~~	~~×~~
~~Lanzhou~~	~~×~~	~~×~~	–	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~
~~Suzhou~~	–	–	–	~~×~~	–	~~×~~	–	~~×~~
~~Jiaxing~~	~~×~~	~~×~~	–	–	~~×~~	–	~~×~~	~~×~~
~~Quzhou~~	~~×~~	~~×~~	–	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~
~~Shaoxing~~	~~×~~	~~×~~	~~×~~	~~×~~	–	–	–	–
~~Hangzhou~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~
~~Qinghuangdao~~	–	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~
~~Chongqing~~	–	–	–	–	~~×~~	~~×~~	~~×~~	~~×~~
~~Xining~~	~~×~~	~~×~~	–	~~×~~	–	–	–	–
~~Qingdao~~	–	–	–	–	~~×~~	~~×~~	~~×~~	~~×~~
~~Shanghai~~	~~×~~	~~×~~	~~×~~	–	~~×~~	~~×~~	~~×~~	~~×~~
~~Huhehaote~~	–	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~
~~Wenzhou~~	~~×~~	~~×~~	–	–	~~×~~	~~×~~	~~×~~	~~×~~
~~Nanning~~	~~×~~	~~×~~	~~×~~	~~×~~	–	–	–	–
~~Taizhou~~	–	–	–	–	–	~~×~~	–	~~×~~
~~Guangzhou~~	~~×~~	–	–	~~×~~	~~×~~	–	~~×~~	~~×~~
~~Chende~~	~~×~~	~~×~~	~~×~~	–	–	–	–	–
~~Dalian~~	~~×~~	~~×~~	~~×~~	~~×~~	–	–	–	–
~~Guiyang~~	–	~~×~~	–	–	–	–	–	–
~~Lishui~~	~~×~~	~~×~~	–	–	~~×~~	~~×~~	~~×~~	~~×~~
~~Yinchuan~~	~~×~~	–	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~
~~Shenzhen~~	~~×~~	–	–	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~
~~Zhuhai~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~
~~Kunming~~	~~×~~	~~×~~	~~×~~	~~×~~	–	–	–	–
~~Fuzhou~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~
~~Zhoushan~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~
~~Lasa~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~	~~×~~
~~Haikou~~	~~×~~	–	~~×~~	–	~~×~~	~~×~~	~~×~~	~~×~~

Table ~~5~~3. Performance of daily $PM_{2.5}$ (MFB and MFE) and O_3-1h (MNB and MNE) in different regions of China based on individual inventories and the ensemble. The weighting factors (w) used to calculate the ensemble of each region are also included.

	Region	MEIC			SOE			EDGAR			REAS2			ENSEMBLE	
	(# of Cities)	w	MFB	MFE	w	MFB	MFE	w	MFB	MFE	w	MFB	MFE	MFB	MFE
$PM_{2.5}$	NE (4)	0.16	-0.23	0.44	0.21	0.38	0.68	0.20	-0.30	0.43	0.43	-0.12	0.43	-0.08	0.42
	NCP (14)	0.00	-0.30	0.47	0.52	-0.34	0.46	0.14	-0.40	0.51	0.56	-0.20	0.41	-0.12	0.40
	NW (6)	0.00	-0.87	0.90	0.20	-0.80	0.84	0.59	-0.85	0.87	1.00	-0.81	0.83	-0.49	0.66
	YRD (20)	0.05	-0.29	0.45	0.00	-0.27	0.43	0.61	-0.23	0.40	0.35	-0.13	0.40	-0.18	0.38
	CNT (5)	0.09	-0.10	0.46	0.18	-0.05	0.41	0.50	-0.27	0.40	0.22	0.09	0.44	-0.14	0.37
	SCB (2)	0.00	0.10	0.48	0.64	0.23	0.48	0.00	-0.10	0.39	0.08	0.07	0.43	-0.15	0.40
	SOUTH (9)	0.10	-0.35	0.51	0.00	-0.18	0.41	0.59	-0.07	0.45	0.30	-0.25	0.44	-0.16	0.41
	CHINA (60)	0.07	-0.34	0.52	0.14	-0.26	0.50	0.38	-0.33	0.49	0.49	-0.22	0.46	-0.20	0.45
		w	MNB	MNE	w	MNB	MNE	w	MNB	MNE	w	MNB	MNE	MNB	MNE
O_3-1h	NE	0.09	0.44	0.50	0.00	0.16	0.34	0.45	0.41	0.47	0.27	0.42	0.48	0.14	0.31
	NCP	0.29	0.33	0.47	0.12	0.23	0.44	0.06	0.46	0.59	0.42	0.47	0.56	0.25	0.43
	NW	0.00	0.65	0.72	0.82	0.54	0.62	0.00	0.70	0.77	0.00	0.68	0.74	0.25	0.46
	YRD	0.00	0.20	0.41	0.53	0.14	0.38	0.00	0.25	0.45	0.45	0.27	0.44	0.17	0.39
	CNT	0.27	0.27	0.47	0.18	0.16	0.43	0.10	0.35	0.53	0.36	0.35	0.52	0.18	0.42
	SCB	0.44	0.59	0.68	0.14	0.42	0.58	0.28	0.59	0.70	0.00	0.60	0.72	0.33	0.53
	SOUTH	0.84	0.39	0.50	0.00	0.29	0.46	0.00	0.38	0.51	0.00	0.42	0.53	0.16	0.37
	CHINA	0.19	0.34	0.49	0.20	0.23	0.44	0.00	0.39	0.54	0.51	0.41	0.53	0.21	0.42

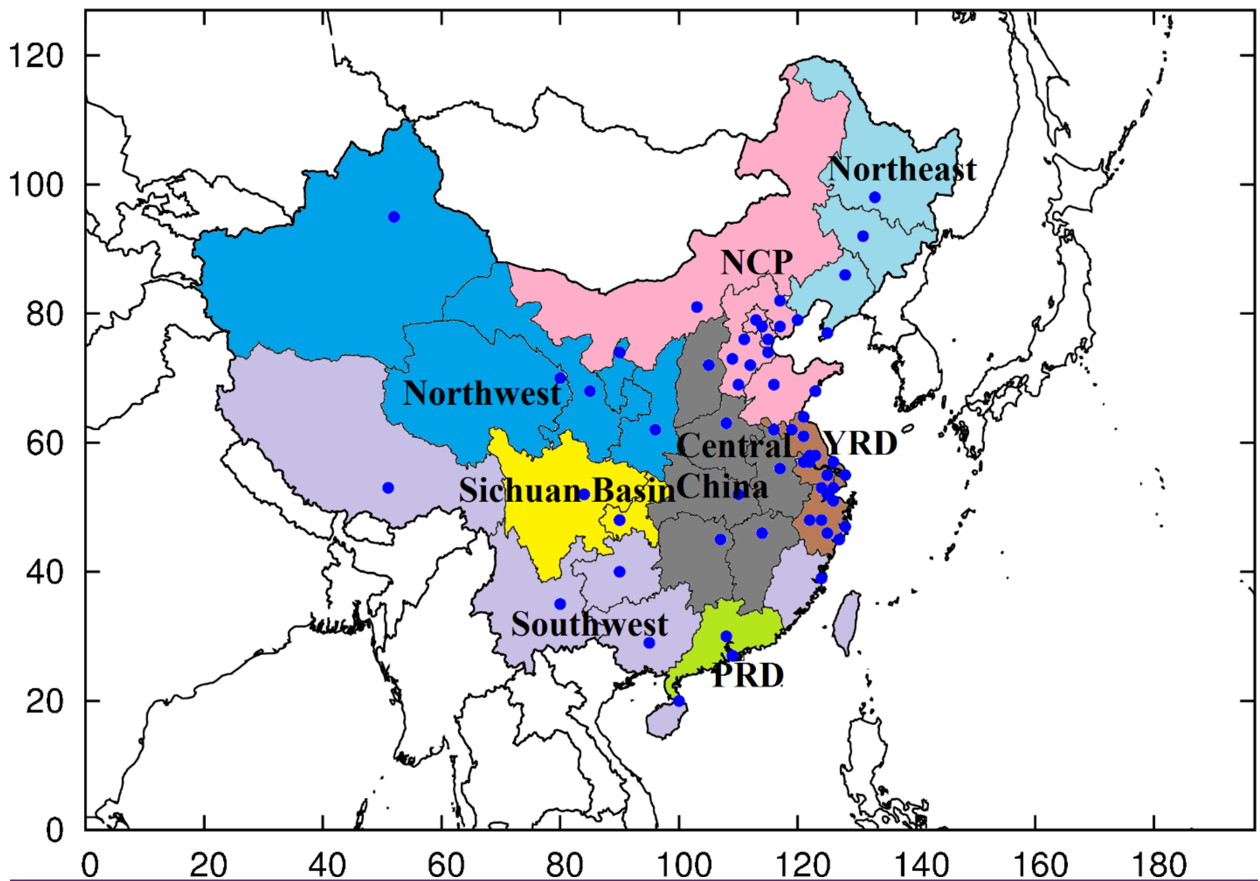


Figure 1. The WRF/CMAQ modeling domain and the regions in China. The dots represent the 60 cities where observational data are available for ensemble analysis.

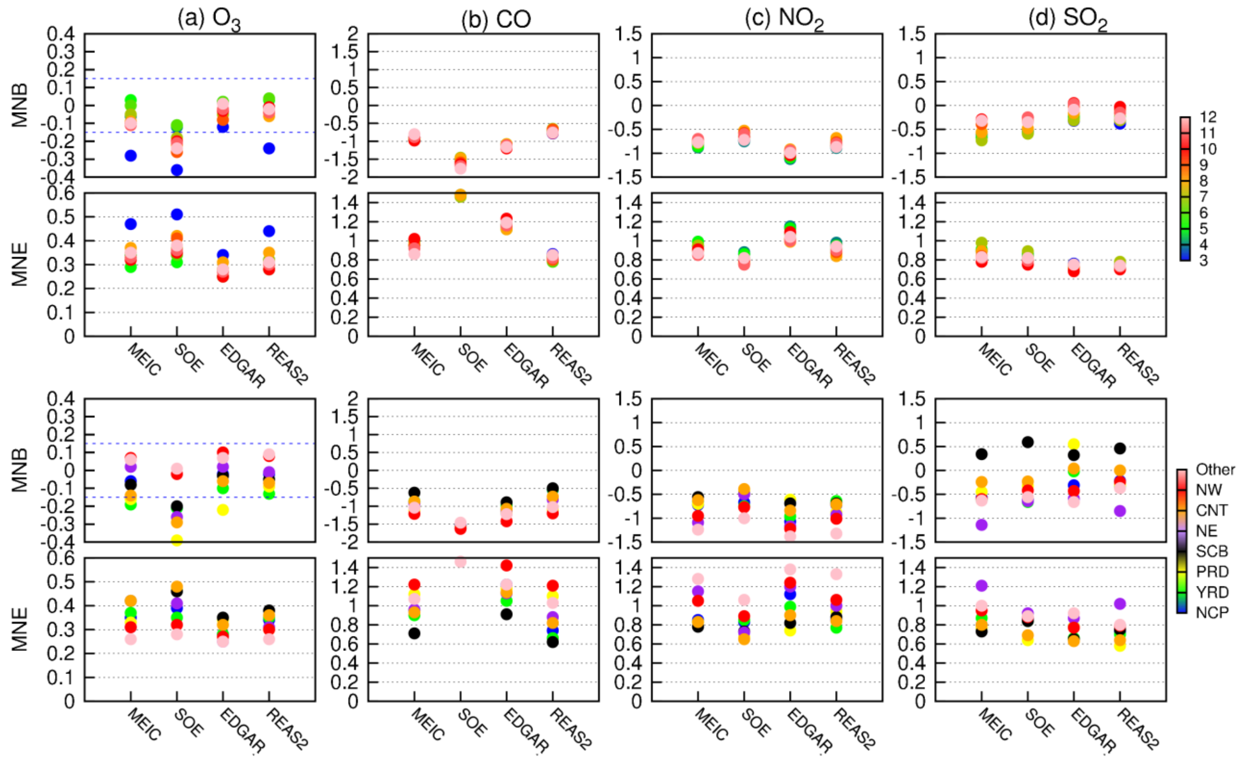


Figure ~~1~~2. Performance of predicted O_3, CO, NO_2, and SO_2 for different months (top two rows) and regions based on simulations with individual inventories. The blue dashed lines on the O_3 plots are +/-0.15 for MNB and 0.3 for MNE as suggested by U. S. EPA (2001a). Changes of colors show the months from March to December in top two rows, while show regions from NCP to Other in the bottom two rows. The same for Figure 2.

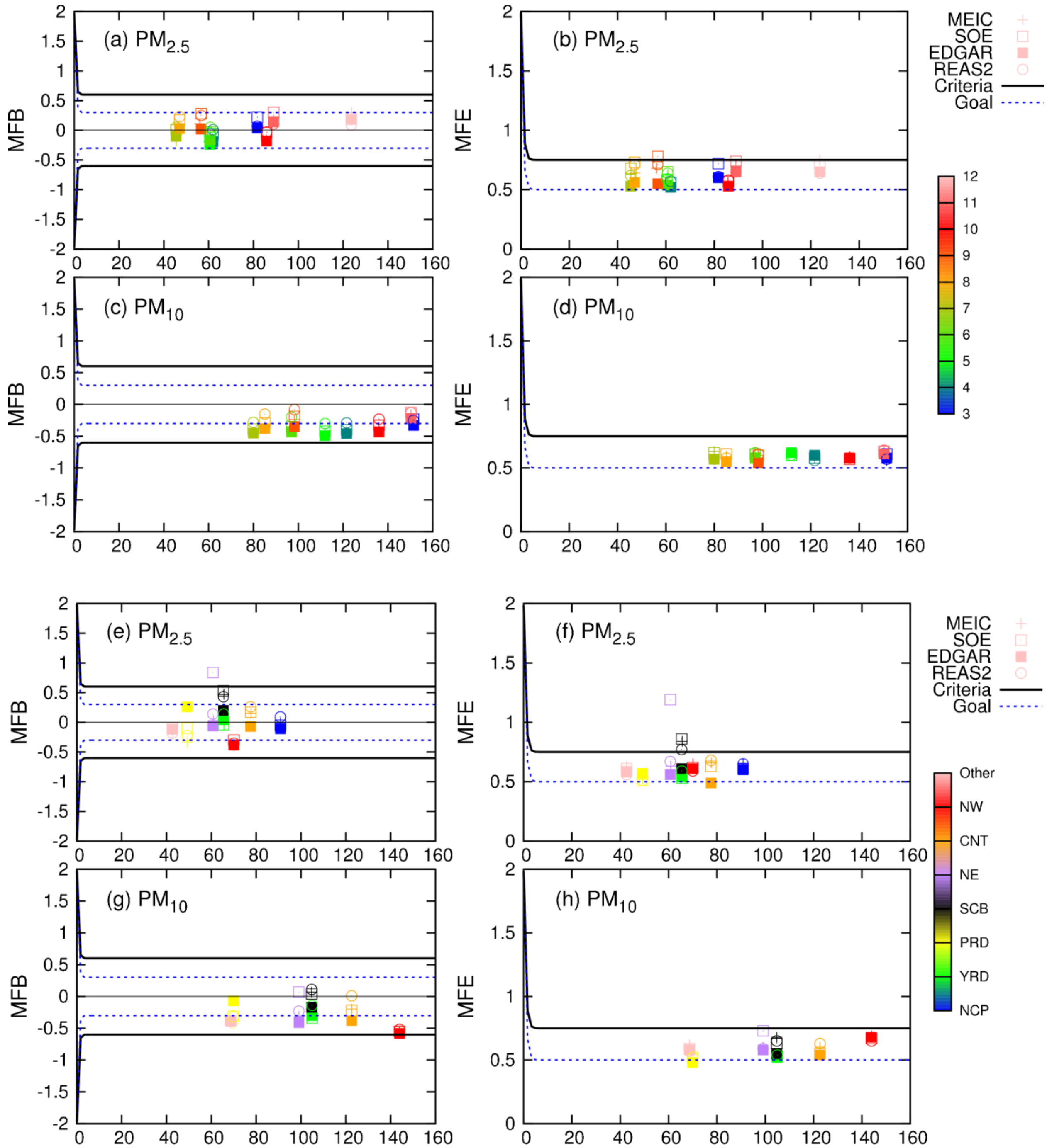


Figure ~~2~~3. Performance of predicted $PM_{2.5}$ and PM_{10} for different months (a-d) and regions (e-h) based on simulations with individual inventories. The x-axis is the observed concentrations. The model performance criteria (solid black lines) and goals (dash blue lines) are suggested by Byun and Russell (2006). 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.

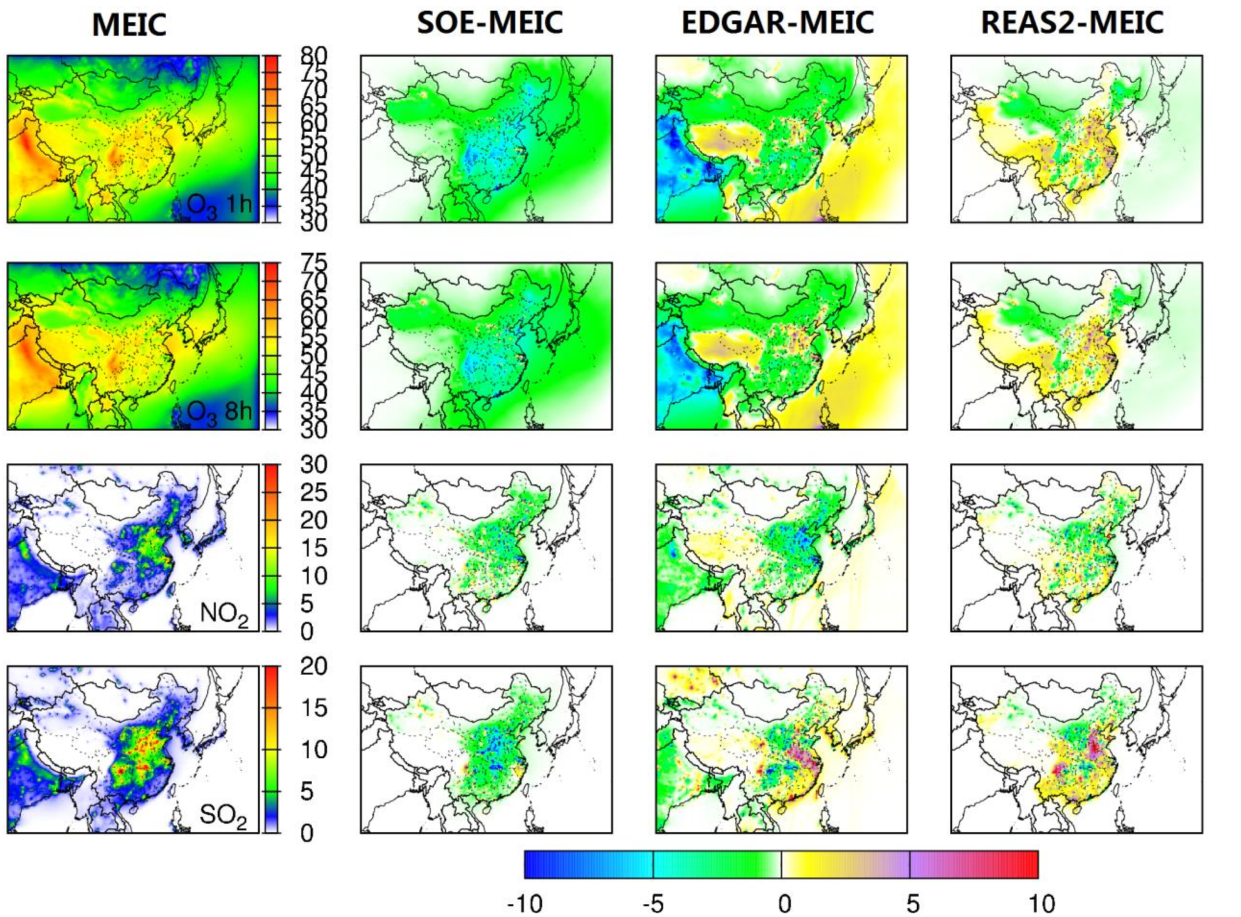


Figure ~~3~~4. Spatial difference of model predicted annual averaged gas species concentrations (in the horizontal panels) with different inventories (in the vertical panels). Units are ppb. The color bars of the first column are different to better show the spatial distribution of different species. White indicates zero while blue, green, yellow and red means concentrations from low to high. The color bar for the other three columns are same, white indicates zero, blue and green mean values less than zero while yellow, purple and red mean values larger than zero. O_3-1h represents daily maximum 1h O_3 and O_3-8h represents daily maximum 8h mean O_3.

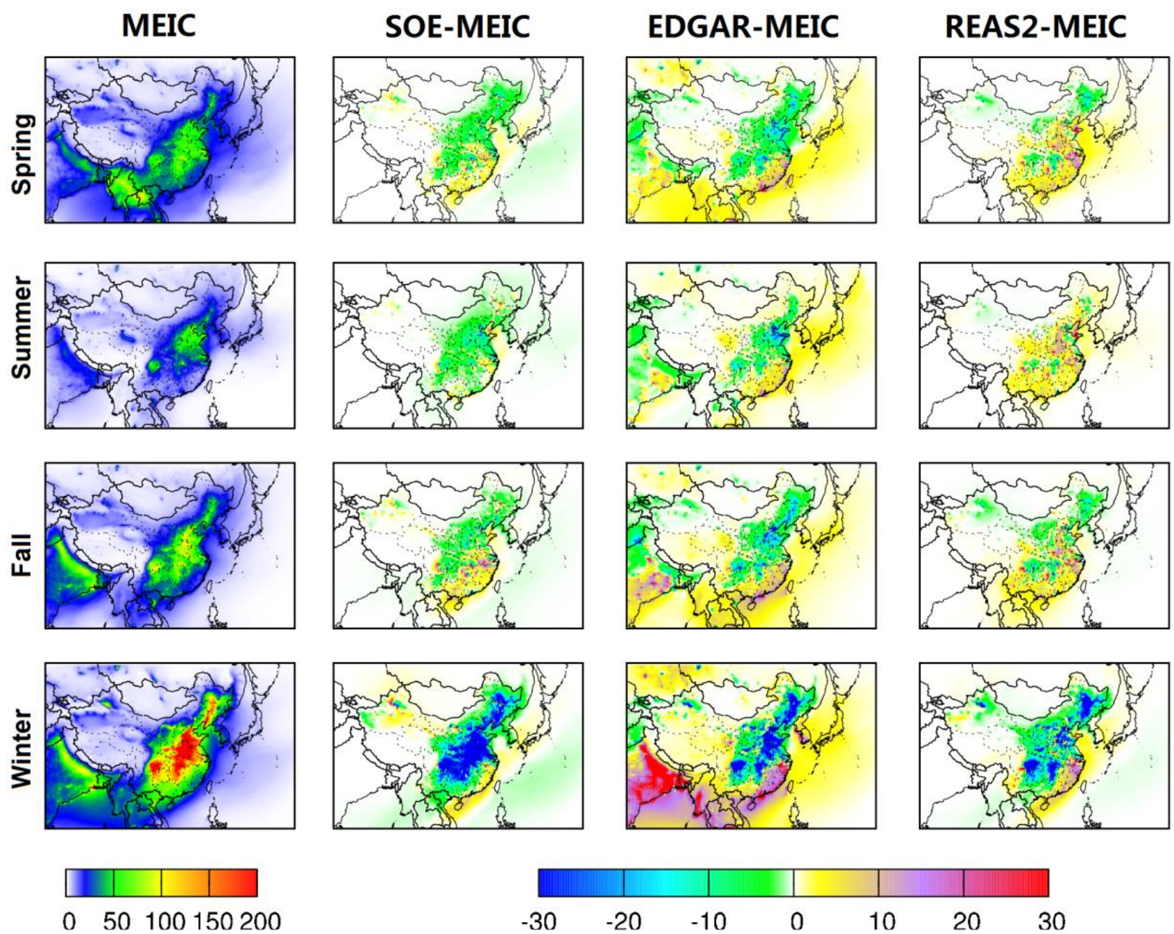


Figure ~~4~~5. Spatial difference of model predicted seasonal averaged $PM_{2.5}$ concentrations (in the horizontal panels) with different inventories (in the vertical panels). Units are μg m^{-3}. In the first column, white indicates zero while blue, green, yellow and red means concentrations from low to high. The color bar for the other three columns are same, white indicates zero, blue and green mean values less than zero while yellow, purple and red mean values larger than zero. The same for Figure ~~5~~6.

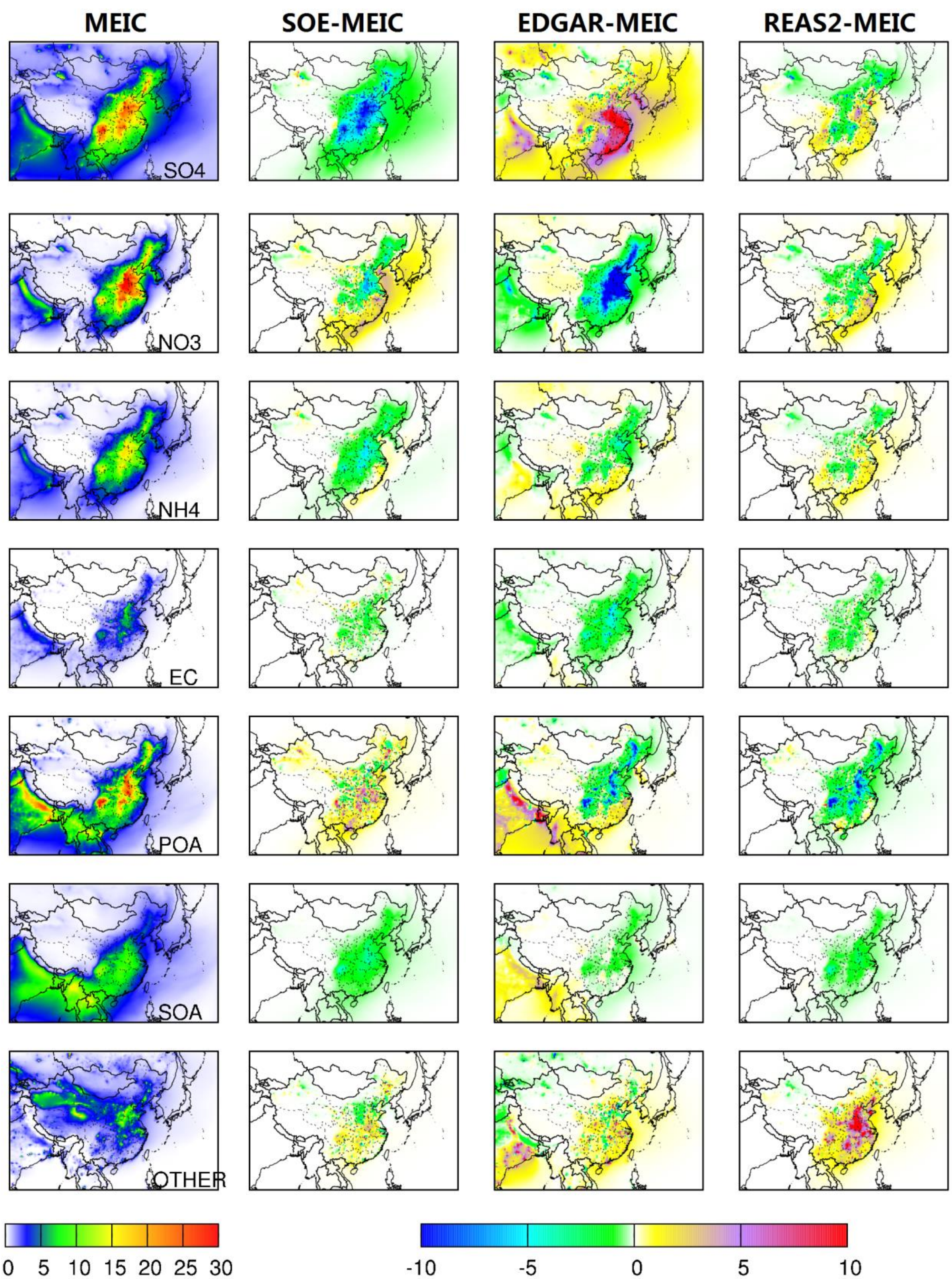
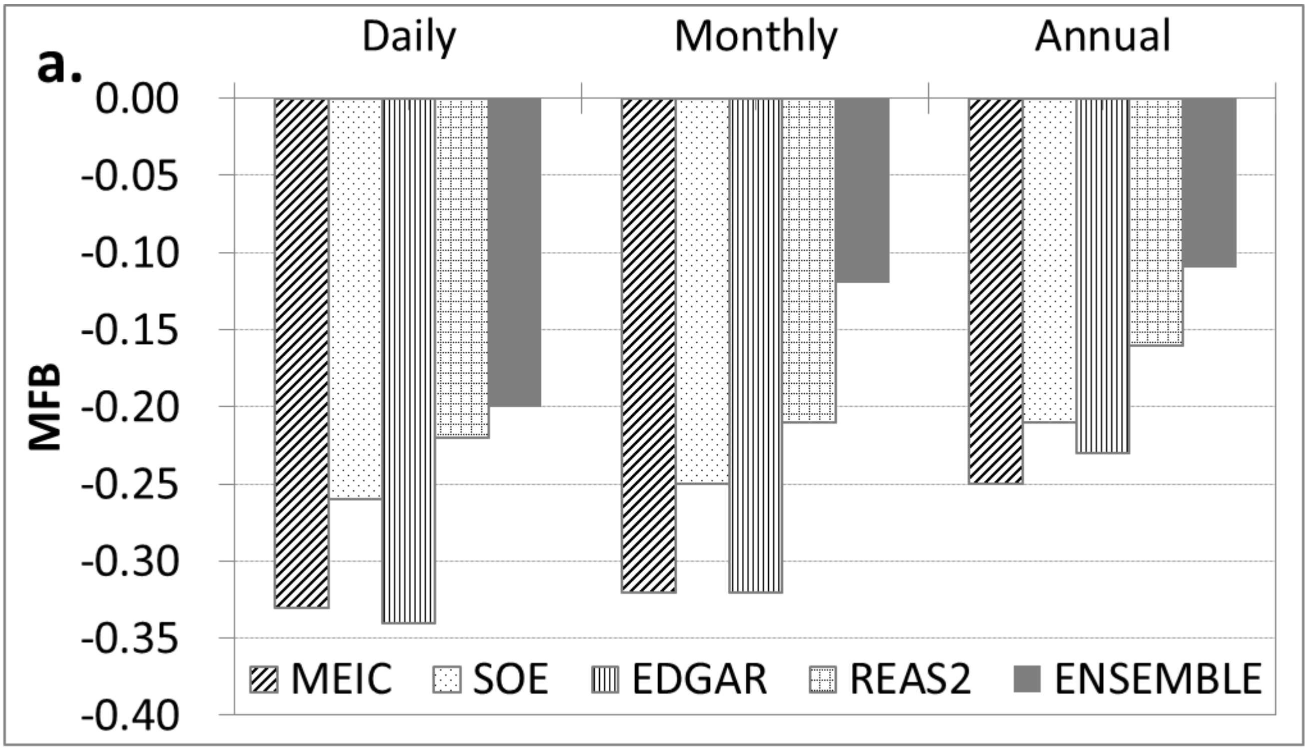


Figure ~~5~~6. Spatial difference of model predicted annual $PM_{2.5}$ components (in the horizontal panels) with different inventories (in the vertical panels). Units are µg m^{-3}.



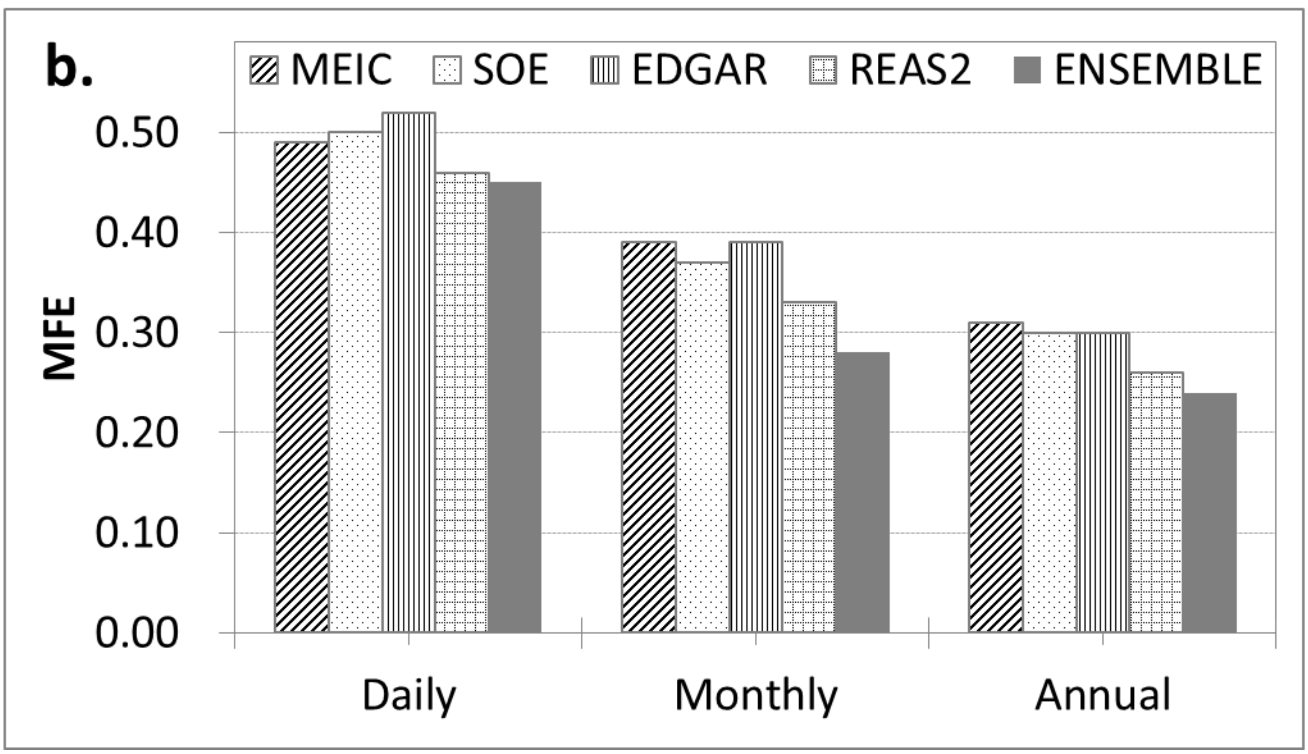


Figure ~~6~~7. MFB and MFE of predicted $PM_{2.5}$ for with an averaging time of 24 hours, 1 month, and 1 year based on the individual inventories and the ensemble.

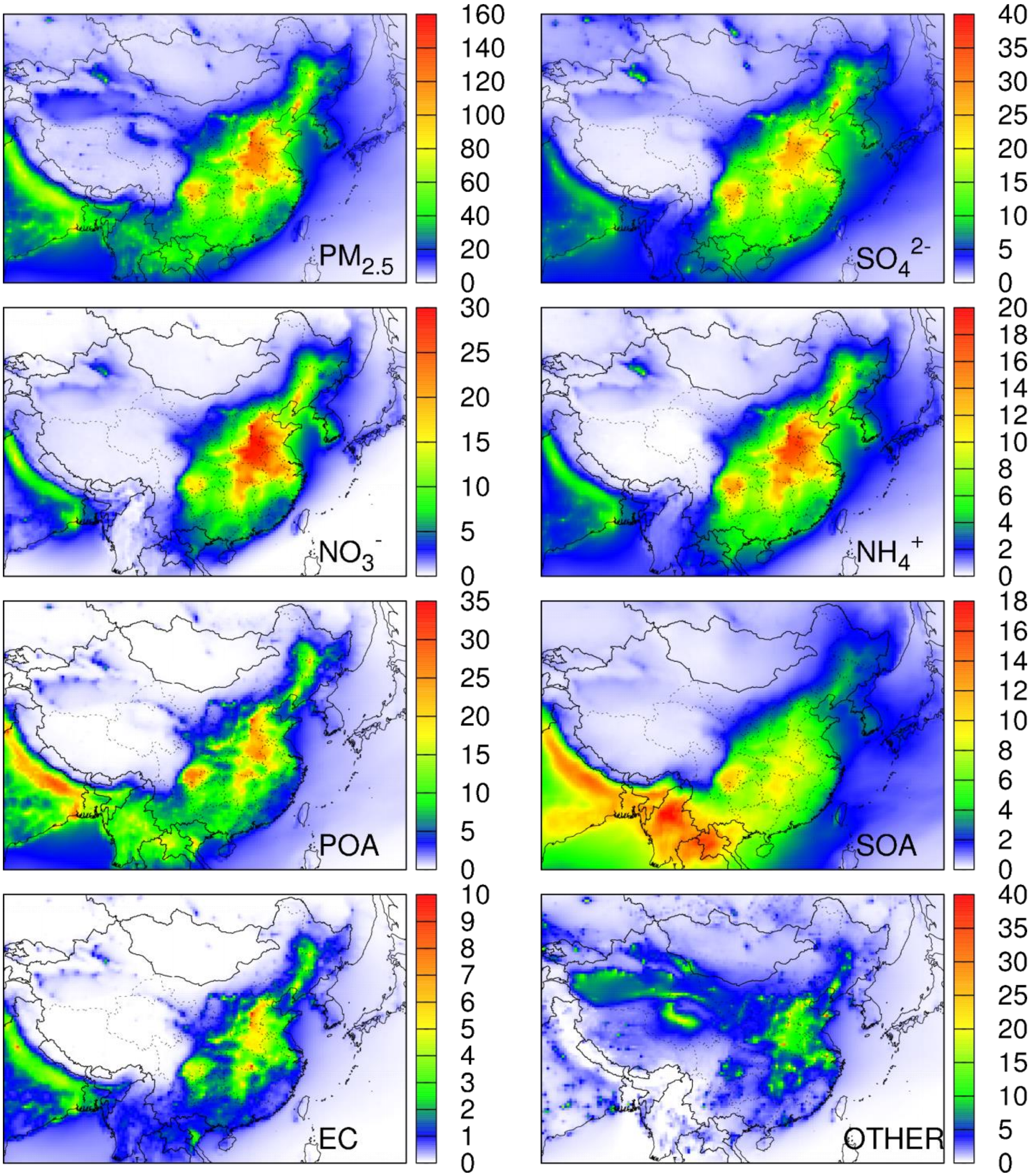


Figure ~~7~~8. Spatial distributions of $PM_{2.5}$ and its components in the ensemble predictions. Units are µg m^{-3}. The scales of the panels are different. White indicates zero while blue, green, yellow and red means concentrations from low to high.