

Reply to comments from Dr. Yves Balkanski on "Observations and modeling of air quality trends over 1990–2010 across the Northern Hemisphere: China, the United States and Europe" by Xing et al.

We would like to thank the editor Dr. Yves Balkanski for a very thoughtful and detailed review of our manuscript that helped to improve the paper.

[Comment]: *In response to reviewer's 3 comment about WRF performances, this answer is very factual and too short. What does this evaluation imply for the results presented in the paper? Does such comparison insure that the meteorological fields used by the model have no bias? Please give 2 or 3 sentences that give credence in your comparison.*

[Response]: We added additional description in the revised manuscript as below:

"The ranges of biases meet the model performance criteria recommended by Emery et al. (2001) for retrospective regional-scale model applications which is  $\leq \pm 0.5$  K,  $\leq \pm 0.5$  m s<sup>-1</sup> and  $\leq \pm 10$  degree respectively, suggesting that meteorology simulations in this study are acceptable. The evaluation of WRF performances ensures that there is no significant bias in the meteorological fields used in the coupled model."

[Comment]: *Please change the sentence: "The lightening NOx emissions used in this study (Price et al, 1997) are likely overestimated by 0.5 to 5 times compared to more recent study (Schumann and Huntrieser et al., 2007) and may contribute to some extent to the overestimation of NOx, O3 and nitrate concentrations." To "The lightening NOx emissions used in this study (Price et al, 1997) are likely overestimated compared to more recent study (Schumann and Huntrieser et al., 2007) and may contribute to some extent to the overestimation of NOx, O3 and nitrate concentrations." It is far fetched to give this range of 0.5 to 5 times, I do not understand where it comes from.*

[Response]: We agree with the comment and we deleted "by 0.5 to 5 times" in the revised manuscript.

Reply to comments from Referee #1 on "Observations and modeling of air quality trends over 1990–2010 across the Northern Hemisphere: China, the United States and Europe" by Xing et al.

We would like to thank the referee for a very thoughtful and detailed review of our manuscript. Incorporation of the reviewer's suggestions has led to a much improved manuscript. Below we provide a point-by-point response to the reviewer's comments and how we have addressed them in the revised manuscript.

### General

[Comment]: *The authors repeatedly invoke "coarse spatial resolution" as a reason for many of the model's shortcomings in comparison with observations. This well may be the case, however, the authors should give some thought (and some discussion in the Conclusions) about exploring this limitation in future work, possibly via finer-scale simulations nested over one or more of the focus domains.*

[Response]: we agree with the reviewer and have provided additional description about limitation of coarse spatial resolution in the revised manuscript (P26 L4-7), as below:  
"To future explore the limitation of coarse spatial resolution, we are currently conducting a study with a finer-scale simulation over the CONUS domain for the same simulated period as from 1990 to 2010. A detailed description and comparison will be provided in a separate paper (Gan et al., in preparation)

Gan, C.-M., Pleim, J., Mathur, R., Hogrefe, C., Long, C.N., Xing, J., Wong, D., Gilliam, R., Roselle, S.J. and Wei, C.: Assessment of long-term simulations with various observations for better understanding of aerosol effects on radiation "brightening" in the United States, in preparation"

[Comment]: *Figures 3-9 need more detailed captions to explain the identities of each Figure component. These graphics are dense with information, but it is not immediately obvious to the reader exactly what is being presented. It is possible to infer from the text what each component of the Figure represents, but a more informative caption would make for a better presentation for the average reader.*

[Response]: we thank the reviewer for pointing this out; more detailed captions for each graphics in Figures 3-9 are provided in the revised manuscript, as below:

Figure 3 captions:

- (a) Simulated SO<sub>2</sub> trend from WRF-CMAQ (unit:  $\mu\text{g m}^{-3} \text{ yr}^{-1}$ )
- (b) Upper-Color map: simulated SO<sub>2</sub> trend in East China overlaid with observed SO<sub>2</sub> trend from China-API, dot represents each observation site, computed on the basis of annual means over the 2005–2010 period with a linear least square fit method, dot size is determined by the significance of trend, i.e., larger symbols denote more significant trends at 0.05 level (unit:  $\mu\text{g m}^{-3} \text{ yr}^{-1}$ )  
Lower-Scatter plot: observed and simulated SO<sub>2</sub> concentration, network-mean for each year corresponding grid cells from model simulation are selected for comparison (unit:  $\mu\text{g m}^{-3}$ )
- (c) same as (b) for Europe - AIRBASE
- (d) same as (b) for Europe - EMEP
- (e) same as (b) for the U.S. - AQS
- (f) same as (b) for the U.S. – CASTNET

[Comment]: *A final general suggestion is that the text be further proofread for acceptable English grammar and usage. Some edits are noted in the specific comments below, but further changes may be needed.*

[Response]: We have reworked all the sections of the manuscript to improve the written English and editorial quality.

[Comment]: *p. 3, lines 15-16: It's debatable that this is the "ultimate" goal of any country. Possibly, the authors meant something like "an important goal for any country."*

[Response]: As the reviewer suggested, we changed "ultimate" to "important" in the revised manuscript (P3 L15).

[Comment]: *p. 6, lines 18-19: It would be appropriate to include a brief summary of results from the WRF performance evaluation here, in particular noting any biases that may have an impact on the results presented in this manuscript (e.g., temperature, precipitation, etc.).*

[Response]: We have included a brief summary of WRF evaluation in the revised manuscript (P6 L17-P7 L1), as below:

"WRF performance for the simulation of hourly surface temperature (T), relative humidity, wind speed and direction was evaluated through comparison with observations from NOAA's National Climatic Data Center (NCDC) Integrated Surface Data (ISD with lite-format) which provides hourly (or with 3-hour interval) meteorological observations over a long historical period across the globe. The mean bias of T, wind-speed and direction over the simulation domain is -0.4 K, 0.4 m s<sup>-1</sup> and -3 degree respectively, within the benchmark range suggested by Emery et al. (2001) for retrospective regional-scale model applications which is  $\leq \pm 0.5$  K,  $\leq \pm 0.5$  m s<sup>-1</sup> and  $\leq \pm 10$  degree respectively"

[Comment]: *p. 7, lines 13-15: It is puzzling to the reviewer why BVOC emissions were kept constant over all simulated years, although it likely does not significantly impact the results obtained. A rationale for choosing constant BVOC emissions should be provided.*

[Response]: We agree with the reviewer that detailed BVOC emissions with high temporal resolution will definitely improve the accuracy of the results. Unfortunately, at the time this study began, there were no available BVOC emissions covering such spatial and temporal scale as simulated here. We have clarified such limitation in the revised manuscript (P26 L1-2), as below: "The trend of biogenic emissions, which hasn't been considered in this study, might also impact the analysis."

[Comment]: *p. 11, lines 12-16: A more detailed explanation should be offered for the difference in sulfate bias between the U. S. networks and the European network, which is an interesting result. Does the reference to "uncertainty in precipitation" refer to something found in the WRF evaluation of this time period? Are there differences in precipitation biases between the U. S. and Europe? If so, they should receive more discussion here.*

[Response]: We appreciate the suggestion from the review, and we further investigated the WRF performance of the precipitation. However, there are no significant differences in the biases between the U.S. and Europe which are -0.14mm and -0.10mm for 6h-duration precipitation

respectively. The difference in sulfate bias between the U.S. networks and the European network might be associated with different SO<sub>2</sub> biases in these two regions, i.e., a moderate bias (NMB=-9.4%) in US-CASTNET but a relatively larger bias (NMB=+67%) in EU-EMEP. The transition rate from SO<sub>2</sub> to SO<sub>4</sub><sup>2-</sup> is likely underestimated in both regions, leading to the underestimation of SO<sub>4</sub><sup>2-</sup> in the U.S. and the better estimates of SO<sub>4</sub><sup>2-</sup> in Europe.

We added some discussion about this in the revised manuscript (P12 L5-10), as below:

“Better performance is shown at EU-EMEP, with NMB within ±10%. The difference in sulfate biases between the U.S. networks and the European network might be associated with the different SO<sub>2</sub> biases, i.e., a moderate bias (NMB=-9.4%) in US-CASTNET but a relatively larger bias (NMB=+67%) in EU-EMEP. The transition rate from SO<sub>2</sub> to SO<sub>4</sub><sup>2-</sup> is likely underestimated in both regions, leading to the underestimation of SO<sub>4</sub><sup>2-</sup> in the U.S. and the better estimates of SO<sub>4</sub><sup>2-</sup> in Europe.”

[Comment]: p. 9, line 2: Should be: “... considered during periods of missing ...”; p. 10, line 21: Should be: “... worst ...” not “worse.”; p. 16, line 11: Should be: “... trends in observations in the urban network ...”. p. 16, line 12: Should be: “... that causes the model to fail to represent ...”. p. 24, line 13: Should be: “... in Europe and North America has been ...”. p. 25, line 15: Should be: “... this relative ratio could potentially ...”.

[Response]: These typos have been corrected in the revised manuscript (P9 L12; P11 L7; P17 L7; P17 L8; P25 L11; P26 L21).

Reply to comments from Referee #2 on "Observations and modeling of air quality trends over 1990–2010 across the Northern Hemisphere: China, the United States and Europe" by Xing et al.

We would like to thank the referee for a very thoughtful and detailed review of our manuscript that helped to improve the paper. Below we provide a point-by-point response to the reviewer's comments and how we have addressed them in the revised manuscript

[Comment]: *The limitations of this paper, e.g., O3 chemistry over China is discussed without evaluation due to the unavailability of O3 data, should be discussed in the conclusion section.*

[Response]: we agree with the reviewer and have provided additional description about this limitation in the revised manuscript (P26 L4-6), as below:

"The lack of long-term observations in Asia, particularly over China and India limits a robust model performance evaluation as well as O3 and PM chemistry assessment in these polluted areas"

[Comment]: *Seven cities are selected in CN-API network in model evaluation. It is better to add an explanation of the reasons to choose only those seven cities for China.*

[Response]: we have added the reason for choosing only those seven cities in the revised manuscript (P8 L21-P9 L3) as below:

"CN-API is the average of observed air pollutant concentrations from urban monitoring sites in each city and represents records in 7 Chinese cities (i.e., Beijing, Shanghai, Guangzhou, Xi'an, Wuhan, Guiyang, Guilin which are located in north China plain, Yangtze-river delta, Pearl-river delta, northwest China, central China and south China respectively) where long-term observations are available starting from 2005."

[Comment]: *The heading rows of Table 3 are confusing (% , emission, and concentration). Please revise it into a more readable format.*

[Response]: we thank the reviewer for pointing this out; more detailed heading rows of Table 3 are provided in the revised manuscript, as below:

	Eastern China		Eastern US		Europe	
Emission	kg km <sup>-2</sup> yr <sup>-1</sup>	% yr <sup>-1</sup>	kg km <sup>-2</sup> yr <sup>-1</sup>	% yr <sup>-1</sup>	kg km <sup>-2</sup> yr <sup>-1</sup>	% yr <sup>-1</sup>
Concentration	μg m <sup>-3</sup> yr <sup>-1</sup>	% yr <sup>-1</sup>	μg m <sup>-3</sup> yr <sup>-1</sup>	% yr <sup>-1</sup>	μg m <sup>-3</sup> yr <sup>-1</sup>	% yr <sup>-1</sup>

Reply to comments from Referee #3 on "Observations and modeling of air quality trends over 1990–2010 across the Northern Hemisphere: China, the United States and Europe" by Xing et al.

Reply to comments from Referee #3 on "Observations and modeling of air quality trends over 1990–2010 across the Northern Hemisphere: China, the United States and Europe" by Xing et al.

We would like to thank the reviewer for a very thoughtful and detailed review of our manuscript that helped to improve the paper. We address all the points raised by the reviewer as follows. We basically followed all the comments and revised manuscript accordingly.

[Comment]: *The manuscript says that an evaluation of the WRF meteorology will be the subject of a separate paper. I think that there should be at least some evaluation of WRF in the current paper. This could be simply be some summary statistics of model performance.*

[Response]: we agree with the reviewer and have provided additional description about the WRF performance evaluation in the revised manuscript (P6 L17-P7 L1), as below:

“WRF performance for the simulation of hourly surface temperature (T), relative humidity, wind speed and direction was evaluated through comparison with observations from NOAA's National Climatic Data Center (NCDC) Integrated Surface Data (ISD with lite-format) which provides hourly (or with 3-hour interval) meteorological observations over a long historical period across the globe. The mean bias of T, wind-speed and direction over the simulation domain is -0.4 K, 0.4 m s<sup>-1</sup> and -3 degree respectively, within the benchmark range suggested by Emery et al. (2001) for retrospective regional-scale model applications which is  $\leq \pm 0.5$  K,  $\leq \pm 0.5$  m s<sup>-1</sup> and  $\leq \pm 10$  degree respectively.”

[Comment]: *The authors are comparing CMAQ model output (which is I think at 108 km resolution) with the AQS and EU-AIRBASE data, which are primarily from urban areas. Each of these sites are representative of much smaller regions. I do not think this comparison is appropriate. The AQS and EU-AIRBASE data should be averaged over the 108 km grid cells before comparing with the model to obtain a more valid analysis.*

[Response]: we agree with the reviewer that the averaged AQS and AIRBASE data are more appropriate for the comparison against with simulations on a 108km resolution. We have reworked these two networks and updated all the numbers in the revised manuscript. In most of cases, the performance gets slightly improved. The NMBs for SO<sub>2</sub> in AQS/AIRBASE are changed from -46%/-12% to -38%/-18%; the NMBs for NO<sub>2</sub> in AQS/AIRBASE are changed from -54%/-57% to -48%/-54%. Such updates have been noted in the revised manuscript (P8 L18-L21), as below: “Sites in US-AQS and EU-AIRBASE are typically closer to urban areas and may be impacted by local pollution and features sub-grid to the model resolution, thus are representative of much smaller regions. To obtain a more valid analysis, the US-AQS and EU-AIRBASE data were averaged over the 108 km grid cells before comparing with the model.”

[Comment]: *p. 25457, lines 17-19 and p. 25458, line 8: the earlier text mentions nested regional domains at finer resolution, and then in the later text specified the three sub-regions used in the analysis. However, no specific finer resolution is mentioned. This leaves the reader unclear as to*

*whether the sub-regions are or are not nested. I've assumed they are not. Please clarify the text.*

**[Response]:** the three sub-regions are not nested. We have clarified it in the revised manuscript (P6 L8-9), as below:

“We selected three sub-regions, i.e., eastern China (20–40 N, 100–125 E), eastern US (28–50 N, 100–70 W) and Europe (35–65 N, 10W–30 E), for further analysis and comparison with measurements. These three sub-regions are parts of the original northern hemispheric domain and no nested simulations were conducted.”

**[Comment]:** *p. 25459, lines 16-17: lightning NO<sub>x</sub> emissions are said to be from Price et al., 1997. This paper indicates the total global emission is 12.2 Tg/yr. This amount is well above the most well-accepted values of 2 - 8 Tg/yr (Schumann and Huntrieser et al, 2007, ACP). Please provide some indication of what the impact of this likely too large emission value is on NO<sub>x</sub>, O<sub>3</sub>, and nitrate.*

**[Response]:** We have included a brief discussion about this bias in the revised manuscript (P25 L20-P26 L3), as below:

“However, the model estimates still suffer from uncertainties in emissions (in regards to temporal variation and speciation), coarse spatial resolution and subsequent impacts on representation of non-linear atmospheric chemistry. The lightning NO<sub>x</sub> emissions used in this study (Price et al, 1997) are likely overestimated by 0.5 to 5 times compared to more recent study (Schumann and Huntrieser et al., 2007) and may contribute to some extent to the overestimation of NO<sub>x</sub>, O<sub>3</sub> and nitrate concentrations.”

**[Comment]:** *p. 25460, lines 5 - 8: define the acronyms*

**[Response]:** the acronyms have been defined in the revised manuscript, but in the previous section (P 4).

**[Comment]:** *p. 25463, line 7-8: Some statistics on model precipitation vs. observed should be provided. Then, the authors could more definitively say whether precipitation bias is the reason for the underestimation.*

**[Response]:** We appreciate the suggestion from the reviewer, and we further investigated the WRF performance of the precipitation. The precipitation was underestimated domain-wide by from 4% (in summer) to 65% (in winter). We provided the statistics on the performance of precipitations and clarified the reason for the bias in the revised manuscript (P11 L22-P12 L5) as below:

“Some studies also found similar under-prediction in their simulations and they attributed such low biases to the uncertainty in precipitation and overestimation of wet-scavenging. However, precipitation simulated in this study is underestimated domain-wide by 4% (in summer) to 65% (in winter). Wang et al (2009) found similar underestimation of precipitation from -31% to -41%, but SO<sub>4</sub><sup>2-</sup> was over-predicted because higher SO<sub>2</sub> emissions were used. Future investigation of the low bias in predicted SO<sub>4</sub><sup>2-</sup> is still necessary.”

**[Comment]:** *p. 25464, lines 4-7: Were these previous modeling studies at much finer resolution? If so, then resolution may not be the issue.*

**[Response]:** These previous modeling studies were conducted at finer resolutions of 36km/12km.

We agree with the reviewer that the original statement is vague. We have rephrased that in the revised manuscript (P13 L7-L15), as below:

“The correlation between the observed and simulated EC concentrations is high with  $R > 0.5$ , though the model significantly underestimates the concentrations. NMB up to  $-74\%$  which is worse than previous modeling studies utilizing relatively higher spatial resolution (Zhang et al., 2009; NMB=  $-15.4$  to  $8\%$ ; Eder and Yu, 2006; NMB=  $-6\%$ ), but the magnitude of NMB is comparable with Wang et al. (2009) (NMB=  $101.7\%$ ) which also utilized coarse spatial resolution. Some previous CMAQ modeling studies (Tesche et al., 2006; Appel et al., 2008) with higher spatial resolution also found the similar underestimation of EC, indicating other factors besides model resolution, such as uncertainties of PM speciation profiles used to estimate the EC emissions might also contribute to such low biases.”

[Comment]: p. 25465, line 8: maybe ‘eastern AQS’ instead of ‘mid-east AQS’. ‘Mid-east’ is not a commonly used term to describe locations in the US.

[Response]: as the reviewer suggested, we replaced the “mid-east AQS” by “eastern AQS” in the revised manuscript (P14 L13).

[Comment]: p. 25466, line 13: “...capture these trends, yielding trends more similar to those of the emissions”

[Response]: as the reviewer suggested, we modified this sentence into “the model was unable to capture these trends, yielding trends more similar to those of the emissions” in the revised manuscript (P15 L21-L22).

[Comment]: p. 25467, line 24: should “NO<sub>x</sub>- and VOC-limited regimes” be reversed?

[Response]: we thank the reviewer for pointing this out; this typo has been fixed in the revised manuscript (P17 L9), as below:

“a likely switch of O<sub>3</sub> chemistry from VOC- to NO<sub>x</sub>-limited regime which usually goes along with the transition from urban to rural area”

[Comment]: p. 25472, lines 5 - 7: The authors should note that in China the rate of O<sub>3</sub> increase was much smaller during 1995-2002, which was the period when VOC emission growth was much greater than that of NO<sub>x</sub> emissions. This result indicates greater sensitivity of ozone to NO<sub>x</sub> emissions than VOC emissions.

[Response]: we thank the reviewer for this good suggestion. We have included this finding in the revised manuscript (P22 L9-L12), as below:

“The ratio suggested is less than 1 indicating greater sensitivity of ozone to NO<sub>x</sub> emissions than VOC emissions. It’s also obvious to see that the rate of O<sub>3</sub> increase was much smaller during 1995-2002 which was the period when VOC emission growth was much greater than that of NO<sub>x</sub> emissions in China.”