Response to Reviewer 1

Response: Thank you very much for your comments on our paper. We appreciate the time that you have taken to read our manuscript and give us your comments and suggestions. Our replies to your comments are given below in blue.

General Comments
This paper outlines air-quality evaluation of simulations conducted with the WRF-Chem model over China. Sensitivity to emissions is tested by using two commonly used anthropogenic emission inventories (REAS and EDGAR), finding large differences between the two for certain species. This is followed by an evaluation of key species relevant for air quality vs. ground site measurements across China and Japan, using output from the WRF-Chem model run with the REAS emissions.

In general, I think the paper is well written and easy to follow. However, there are several points I would like to see addressed before it would be suitable for publication. I think the paper would benefit from stronger justification (explaining the novelty of the study) and discussion of the impacts and implications of its findings. This is particularly true in the conclusions section, which mostly just summarizes the results and only offers a short last sentence of the form "These findings suggest future work is needed…"as implications of the study.

Response: Based on your helpful feedback, we have strengthened the justification for our study in the introduction and discussed the impacts and implications in the last section. We have revised the abstract, introduction, and conclusion. The newly added sentences in the revised manuscript are in Italics, as given below.

Abstract:
Our study highlights the importance of constraining emissions at the provincial level for regional air quality modeling over East Asia. Our results suggest that future work should focus on the improvement of provincial-level emissions especially estimating primary PM, SO_2 and NO_x.

Introduction section, paragraph 3 and 4:
One of the possible reasons that models underestimate particulate matter (PM) concentrations is the uncertainty in emissions. *Several emissions inventories for Asia have been developed by different groups, each with different purposes and characteristics (JRC and PBL, 2010; Kurokawa et al., 2013; Streets et al., 2003; Klimont et al., 2009). Comparison of the emissions inventories has revealed large differences in these emissions estimates. Kurokawa et al. (2013) compared different emissions inventories for several provinces in China and found that the difference in primary PM_{10} emissions can be as high as 140%. The possible causes for such discrepancies among emissions inventories are differences in estimates of: (1) activity level, (2) level of technologies implemented, and (3) emission factors. Since it is hard to measure emission factors of each individual source at the scale of a province or a country, uncertainties arise when emission factors from one place are applied to another. Activity data or emission factors are often not available at the level of detail required for making insightful comparisons across emissions inventories.*

While comparison of emissions inventories has revealed notable differences in the emissions estimates, few studies have addressed to what extent uncertainty in the emissions inventories really matter for the outcome of air quality modeling studies. Ma and van Aardenne (2004) compared simulated surface O_3 mixing ratios over China using three different emissions inventories as model inputs, and found that surface O_3 differed as much as 30–50% among different model simulations. They also demonstrated that the differences in NO_x and non-methane volatile organic compounds (NMVOCs) among different inventories were dominant factors for the discrepancies in simulated O_3 mixing ratios. Amnuaylojaroen et

al. (2014), on the other hand, studied the effect of different anthropogenic emissions inventories on air quality over Southeast Asia and found only a small difference in simulated O_3 (about 4.5 %) and CO (about 8 %) mixing ratios. However, these studies did not investigate the impact of emissions inventories on other pollutant species such as PM. *Unlike the previous studies, which focused on uncertainties of simulated O_3, CO and NO_x, this study provides quantitative information on how emissions inventories impact PM and other pollutants including SO_2.*

Conclusion section, paragraph 3:
Quantifying uncertainties of simulated air quality at the provincial level due to emission inputs reveals that the uncertainty in emissions inventories leads to significant differences in simulated levels of air pollutants, especially PM_{10}, SO_2 and NO_2. For O_3, on the contrary, different emissions inventories lead to only a moderate variability, showing agreement with the findings of previous studies (Ma and van Aardenne, 2004; Amnuaylojaroen et al. 2014). Our study highlights the importance of better constraining emissions at the provincial level for regional air quality modeling over East Asia, where anthropogenic emissions are high and air pollution is a major environmental and public health challenge. Model evaluation results also indicate that emissions inventories that do not consider local emissions control policies could result in large model discrepancies. Our results suggest that future work should focus on the improvement of provincial-level emissions especially estimating primary PM, SO_2 and NO_x.

Some more details are needed in the description of model setup for the simulations to be reproducible (details given in specific comments below).

Response: We have replied with specific comments about model setup further below and shown our replies here.

We include the aerosol-radiative feedback in our simulation. The rapid radiative transfer model (RRTM) scheme (Mlawer et al., 1997) is used to represent both shortwave and longwave radiation.

The horizontal winds, temperature, and moisture are nudged to 2007 meteorological fields at all vertical levels. The meteorological data are obtained from the National Center for Environmental Prediction (NCEP) Global Forecast System final gridded analysis datasets.

As far as I can see there is no evaluation of the meteorology in the model in the paper. Representing the meteorology reasonably is a first-order pre-requisite for simulating the air-quality accurately. Please show some comparisons of key meteorological variables with some ground sites and/or satellite products (e.g. TRMM), either in the main paper or supplementary material.

Response: We have added comparisons of modeled and observed 2 m temperature, 2 m relative humidity, and 10 m wind speed in Table S4 in the supplementary material and inserted the following text in the main paper, section 4, paragraph 2.

Before evaluating the model performance in simulating air pollutants, we evaluate the simulated meteorological fields, including daily mean 2 m temperature, 2 m relative humidity, and 10 m wind speed against observations from National Climate Data Center of China Meteorological Administration for year 2007 (Table S4). The model reproduces 2 m temperature with a correlation of 0.97 and a negative NMB of -14.57 %. Relative humidity is simulated with a correlation of 0.71 and a positive NMB of 7.02 %. Compared to temperature and relative humidity, the 10 m wind speed has a relatively lower correlation of 0.52 and a higher positive NMB of 59.35 %. Overall, the model performance in simulating these meteorological data is similar to that reported for regional air quality models (Zhang et al., 2015; Tuccella 2012; Tessum et al., 2015).

Table S4 reads as follows:

Table S4. Statistical performance of WRF-Chem-REAS simulations for meteorology. Daily meteorological observations at 194 stations in 2007 are obtained from the website of National Climate Data Center of China Meteorological Administration (http://data.cma.cn/).

Meteorological items	Count	Obs.	Model	*r*	MB	ME	NMB	NMSE
2 m temperature (degree C)	23841	11.51	9.83	0.97	-1.68	2.43	-14.57	0.1
2 m relative humidity (%)	23862	61.77	66.1	0.71	4.34	12.34	7.02	0.06
10 m wind speed (m/s)	23857	2.18	3.48	0.52	1.29	1.49	59.35	0.47

In section 4, there is no evaluation of $PM_{2.5}$ against measurements. I appreciate that for section 3, this is not possible for the emissions sensitivity comparison without $PM_{2.5}$ emissions in EDGAR. However, this is not a limitation in the later half of the paper, where you are just using the REAS case study. It looks as though the PM_{10} results you present are being strongly impacted by dust (given the highest readings are in the North West), so looking at $PM_{2.5}$ should be more representative of anthropogenic pollution. Amalgamating $PM_{2.5}$ comparisons into section 4.1 as a single PM section would be beneficial.

Response: We appreciate your comment; however we are unable to evaluate $PM_{2.5}$ against measurements for 2007 since $PM_{2.5}$ measurement in China started in late 2012. We are working on a model evaluation paper focused on 2013, which will include extensive $PM_{2.5}$ observations. We have added the following text in section 2.3, paragraph 2.

We are unable to evaluate $PM_{2.5}$ against measurements for 2007 since $PM_{2.5}$ measurements in China started in late 2012.

Specific comments

1. Pg 9377, ln 11-13: You should not be presenting your findings in the introduction. Replace this sentence with a reference which says supports your statement, i.e. which highlights the variation in PM emissions between inventories and the uncertainties in emission estimates.

Response: We agree with the reviewer and have replaced our findings with a reference in the introduction section, paragraph 3. It reads as follows:

Several emissions inventories for Asia have been developed by different groups, each with different purposes and characteristics (JRC and PBL, 2010; Kurokawa et al., 2013; Streets et al., 2003; Klimont et al., 2009). Comparison of the emissions inventories has revealed large differences in these emission estimates. Kurokawa et al. (2013) compared different emissions inventories for several provinces in China and found that the difference in primary PM_{10} emissions can be as high as 140%.

2. Pg 9377, ln 23-24: Are there really no other studies investigating PM dependence to emissions inventory in Asia?

Response: To the best of our knowledge, all references we found are discussing the influence of emissions inventories on gaseous pollutants. We would be grateful if the reviewer could point us to such studies that we might have missed.

3. Pg9378, ln 10: Zhang et al., 2015 is at least one paper I know of which compares WRF-Chem model output with a wide array of measurements across China (albeit mostly just for PM2.5). As one of the main justifications of your study was that it is the first WRF-Chem study to compare against an extensive

network in East Asia, please expand on why your study is novel in light of this (and possibly other) papers.

Response: We agree that extensive $PM_{2.5}$ observations were used to evaluate WRF-Chem simulations by Zhang et al. (2015), albeit all the measurements were for time period beginning late 2012. However, our focus is 2007, when $PM_{2.5}$ in China was not available. There are a few studies that use a large array of historical PM_{10} for WRF-Chem study. In addition to PM_{10}, we also compare other gaseous pollutants such as O_3, SO_2 and NO_2. We compare data not only in China but also in Japan and Nepal and investigate the seasonal variability of each pollutant, which has not been performed previously.

Section 2.1: some questions on model setup:

4. Do you run with aerosol-radiative feedback?

Response: Yes, we include aerosol-radiative feedback in our model setup. We have included the following descriptions in Section 2.1, paragraph 1.

We include the aerosol-radiative feedback in our simulation. The rapid radiative transfer model (RRTM) scheme (Mlawer et al., 1997) is used to represent both shortwave and longwave radiation.

5. Do you periodically re-initiate the meteorology, or run with nudged met?

Response: We run the model with nudged meteorology. The following description has been added in section 2.1, paragraph 1.

The horizontal winds, temperature, and moisture are nudged to 2007 meteorological fields at all vertical levels. The meteorological data are obtained from the National Center for Environmental Prediction (NCEP) Global Forecast System final gridded analysis datasets.

6. Do you know how sensitive your model simulations are to the boundary conditions? It seems as though you are using 2010 data from AM3, is that representative of the 2007 period you are running over? Is there any reason why you are not using a product driven by meteorological reanalysis, e.g. MOZART-4 (Emmons et al., 2010)?

Response: We performed an additional simulation in July 2007 with boundary conditions obtained from MOZART-4 for year 2007. The comparisons of simulated PM_{10}, O_3, NO_2 and SO_2 using two different boundary conditions are shown in Table 1 below. We find that the boundary conditions have a small influence on simulated monthly mean O_3 and NO_2 mixing ratios. For these two pollutants, 2007 MOZART-4 boundary leads to lower MFB and NMB, while 2010 AM3 shows a stronger correlation. Based on this analysis we conclude that our model-simulated pollutant concentrations are not significantly sensitive to boundary conditions.

Table 1. Statistical measures calculated from two simulations using 2010 AM3 and 2007 MOZART-4 as boundary inputs.

Pollutants	Boundary	Obs.	Model	*r*	NMB	MFB	MFE	NMSE
PM_{10}	2010 AM3	82.45	80.40	0.37	-2.49	-9.86	48.13	0.36
	2007 MOZART	82.45	79.02	0.39	-4.16	-13.70	50.72	0.38
O_3	2010 AM3	35.18	43.58	0.78	23.87	25.98	30.16	0.10
	2007 MOZART	35.18	39.29	0.64	11.68	15.20	29.28	0.11
SO_2	2010 AM3	4.82	8.25	0.74	71.14	30.77	74.02	1.90
	2007 MOZART	4.82	8.74	0.73	81.35	42.71	77.11	1.99
NO_2	2010 AM3	17.54	12.17	0.73	-30.60	-54.99	62.01	0.47
	2007 MOZART	17.54	13.24	0.63	-24.50	-27.61	49.97	0.51

7. Pg 9380, section 2.2: Do you have any seasonal or diurnal variation in your emissions? If not, you should also mention this when discussing potential reasons for model discrepancy, particularly in terms of different magnitudes of error in different seasons in section 4.

Response: REAS emissions are monthly, while EDGAR emissions are yearly. We considered diurnal variation when we processed the emissions. We have added the following description of seasonal and diurnal variation in section 2.2:

We apply the same diurnal variation to both REAS and EDGAR. REAS emissions inventory provides monthly emissions for each pollutant, while the EDGAR emissions inventory only provides yearly emissions estimates.

8. Pg 9381 ln 25 – pg 9382 ln 3: The sentence "For PM10, we use the performance goals and criteria of Boylan and Russell (2006)." Is redundant. Please delete and put the reference at the start of the next sentence :" Following Boylan and Russell (2006), we set…"

Response: We agree with the reviewer and made changes according to the comments above.

Following Boylan and Russell (2006), we set the performance goals of PM_{10} as: MFB less than or equal to ± 30 % and MFE less than or equal to 50 %. The performance criteria of PM_{10} are MFB ≤ ± 60 % and MFE ≤ 75 %.

9. When you say "…goals and criteria for MFB to be less than or equal to 30 and 60%, respectively", it is not clear what the ±30% and ±60% refer to, as you are only discussing MFB for PM10. Please clarify.

Response: We have rewritten the sentence as follows:

Following Boylan and Russell (2006), we set the performance goals of PM_{10} as: MFB less than or equal to ± 30 % and MFE less than or equal to 50 %. The performance criteria of PM_{10} are MFB ≤ ± 60 % and MFE ≤ 75 %.

10. Add "±" in front of "50%", "75%" and "35%".

Response: These numbers are for mean fractional error (MFE), which is always positive.

$$MFE = \frac{1}{n}\sum_{i=1}^{n}\frac{2|Mi - Oi|}{Mi + Oi} \times 100\%$$

11. Pg 9382 ln 13: Please make explicitly clear here that the similarity in SO_2 emissions is purely coincidental, and you would probably get very different values if the domain was different.

Response: We agree with the reviewer and have added the following sentence in the revised manuscript:

We note that this similarity is purely coincidental and depends on the domain. In certain parts of the domain REAS estimate is higher than EDGAR, while the opposite is true for other parts of the domain. When averaged over the whole domain, both inventories produce similar estimates (Fig. 2).

12. Pg 9383 ln 4-5: While I appreciate doing a detailed analysis on the causes of differences in emissions between inventories is out of the scope of the paper, it would be interesting to have some more detail on why the two are different. This discussion may be better placed in the introduction or section 2.2 than here.

Response: We have added more detail on the cause of uncertainties in different emissions inventories in the introduction, paragraph 3.

The possible causes for such discrepancies among emissions inventories are differences in estimates of: (1) activity level, (2) level of technologies implemented, and (3) emission factors. Since it is hard to measure emission factors of each individual source at the scale of a province or a country, uncertainties arise when emission factors from one place are applied to another. Activity data or emission factors are often not available at the level of detail required for making insightful comparisons across emissions inventories.

13. Pg 9383 ln 22: change "large" to "largest".

Response: Corrected.

14. Pg 9383 ln 3-4: It would be helpful for those with less knowledge of the geography of China to have a visual representation of the different Chinese regions, rather than just a list in the supplement. Perhaps boxes drawn over the different regions in Fig. 1, or a similar map in the supplementary material?

Response: We have added a supplemental figure, Fig. S1, to highlight the geographic location of each region.

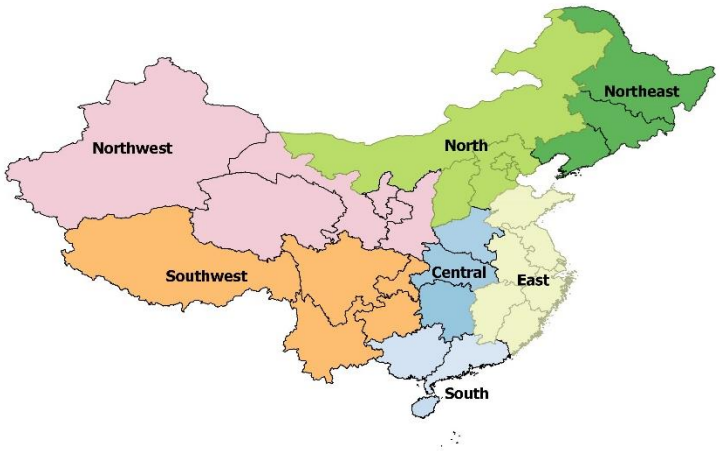

Figure S1. The seven regions in mainland China used for analysis in this paper.

15. pg 9384, ln 11: I am confused by the line "More detailed comparisons, assessing the impacts due to various inventories, will be conducted in our future work." This stuck me as a strange way to structure a set of papers. I agree that having more emission inventories to compare would be of benefit, but I would have thought the best place to present that would be in this paper, rather than another one. This would make the emissions sensitivity section much more interesting. A second paper focused on a more thorough evaluation with the chosen inventory (REAS) and/or investigating other scientific questions (e.g. impacts on health/climate) would make an interesting follow up. Please comment as to why you have written the papers in this form.

Response: We have deleted this sentence. We thank the reviewer for suggestions on how to construct a set of papers, which we hope to adopt in the future.

16. Pg 9385, ln 24-25. I see you have measurements from Lhasa, please can you comment on whether your modelled seasonal trends of PM_{10} in Tibet are seen in the measurements as well. If they diverge, what does this say about the dust emission scheme?

Response: The following table compares monthly mean observed and modeled PM_{10} from Lhasa. In terms of the trend, the modeled PM_{10} agrees with the observations, with the highest value in January, followed by April, October, and July. In April, July, and October, the model underestimates 50-60 % of the measurement.

Table 2. Comparison between observed and modeled monthly mean PM_{10} concentration at Lhasa, Tibet.

PM_{10} ($\mu g/m^3$)	January	April	July	October
Observed	102	82	55	68
Modeled	116	40	24	27

17. Pg 9386, ln 15-17. This can be evaluated by comparing the wintertime model meteorology with measurements.

Response: We further compared the simulated and observed wind speed in four months of 2007 and found that wind speed in the winter tends to be overestimated as shown in the following Table 3. This result illustrates that WRF-Chem has difficulties simulating stagnant conditions in winter.

Table 3. Statistical measures calculated from REAS drived simulations for wind speed.

Month	Obs.	Model	r	MB	ME	NMB	NMSE
Jan.	1.97	3.38	0.49	1.41	1.59	71.40	0.62
Apr.	2.55	3.88	0.56	1.33	1.53	52.13	0.37
Jul.	2.13	3.24	0.39	1.11	1.34	51.97	0.41
Oct.	2.08	3.42	0.57	1.34	1.51	64.04	0.51

18. Pg 9388, ln 2-3. Please comment on the seasonal variation of your emissions in the model. If there is none, the difference in modeled concentrations must be due to meteorological changes between the seasons.

Response: We have checked the monthly emissions of SO_2 in China, as shown in the following Table 4, and realized that January is not the highest emission month among the four, although the simulated SO_2 mixing ratio is the highest. We agree with the reviewer that the modeled difference in SO_2 concentrations is mostly due to meteorological changes between the seasons. We have deleted the previous discussion on emissions.

Table 4. Monthly total SO_2 (kt/month) emissions in China.

January	April	July	October
2872	2686	2883	2898

19. Table 3, 4, 5 and 6. Please say in the caption that you are using data from the WRFChem-REAS simulation. I know you say this in the text, but it is good to be reminded here to avoid confusion.

Response: We have made changes accordingly.

20. Figures 4, 7, 9 and 10. These figures will look clearer if lowest value of the scale is set to the first value above 0 (e.g. 20 instead of 0 for figure 4), as this will make the background white and make it easier to pick out regions of higher loadings .

Response: We have created figures based on the reviewer's suggestion, as seen below. However, after comparing those with our original figures, we still prefer the latter since regions in white background appear as if there are missing values.

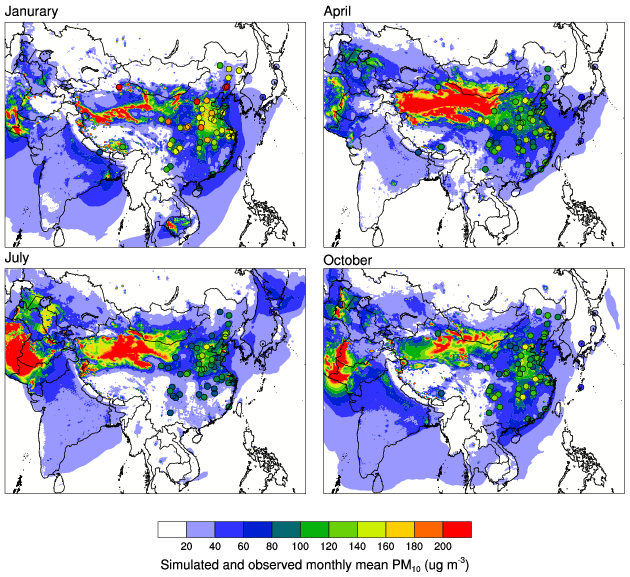

Figure 4. Simulated and observed monthly average surface PM_{10} in 2007 using WRF-Chem-REAS. The filled circles indicate the observed monthly average values

References

Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S. L. and Kloster, S.: Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4), Geosci. Model Dev., 3(1), 43–67, doi:10.5194/gmd-3-43-2010, 2010.

Zhang, B., Wang, Y. and Hao, J.: Simulating aerosol–radiation–cloud feedbacks on meteorology and air quality over eastern China under severe haze conditions in winter, Atmos. Chem. Phys., 15, 2387–2404, doi:10.5194/acp-15-2387-2015, 2015.