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. A detailed response to each comment is provided in this file with comments from referees in black, author's response in red, and author's changes in manuscript in blue.

Report #2 by anonymous referee #1:
In this paper, the author presents a yearlong air quality simulation using a chemical transport model to provide detailed temporal and spatial distribution of O3, PM2.5 and PM2.5 chemical compositions in China. The topic is important, the method is generally sound, and the results are generally reasonable. I suggest this manuscript be accepted with revisions described below.

General comments:
(1) The main objective of this study is to provide detailed temporal and spatial distribution of O3, PM2.5 and its chemical components, which supplements the current observational network in China. The key to success is to ensure that the model well reproduces the magnitude and spatiotemporal distribution of these pollutants. However, the author only compared simulated O3 and PM2.5 concentrations with surface observations. To better evaluate the modeling results, I suggest the author also compare with satellite observations, such as AOD, NO2 column, SO2 column, and tropospheric ozone residual. Moreover, although the observations of PM2.5 chemical components are not publicly available, some data can be found in the literature. It will be very beneficial if the author can compare the simulated chemical components with some available chemical component data, because the spatiotemporal distribution of chemical components is a major focus of this study.

Response: To address the reviewer's comments, we did some comparison between model predicted and satellite observed NO2 columns. An example of the comparison for August 2013 is shown in Figure R1. The spatial patterns of predicted and observed NO2 columns are consistent in general. While the satellite observations are useful in evaluating regional distribution of air pollutants, we choose not to include these in the present analysis and focus on comparison with ground-level observations in this study. A lot of assumptions were used in generating these vertical column density (VCD) products (gridded level 3 products), such as the vertical distribution of target species. To allow an "apple-to-apple" comparison of the model predictions, it is necessary to use a lower level product (e.g. level 2 (L2) products) and adjust the satellite VCD using modeled vertical distributions. The adjusted L2 products will have to be gridded to compare with model predictions (Duncan et al., 2014). The aerosol optical depth (AOD) is not directly comparable with surface PM2.5. Deducing surface PM2.5 from AOD is no trivial task due to spatial and temporal variation of the aerosol composition and weather conditions (CIESIN, 2013). We agree fully with the reviewer that comparison with satellite products is complex yet very important, and thus warrants a more detailed discussion in a separate manuscript.

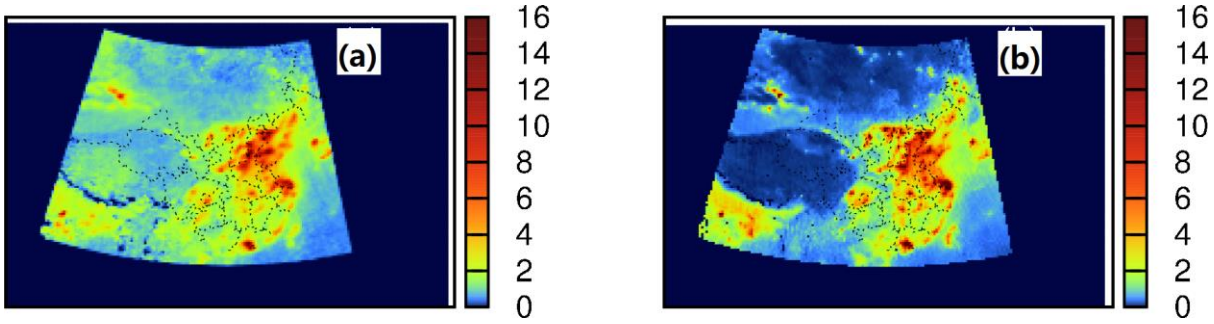

Figure R1. OMI satellite observed (a) and model predicted (b) NO2 columns for August 2013. Units are $x10^{15}$ molecules/cm^2

While aerosol composition data are available to us at a few locations, they are far from complete to provide a holistic view of the uncertainties in aerosol chemical composition throughout the country. Case studies to compare predicted and observed aerosol composition at selected locations will be more suitable to document these findings rather than including them in the current paper that focuses on the national level. In a published paper with similar major model settings, we compared some primary components at different locations and included the results in a previous paper (Hu et al., 2015). We cited this paper in this manuscript.

References:

Duncan, B.N., et al., Satellite data of atmospheric pollution for US air quality applications: Examples of applications, summary of data end-user resources, answers to FAQs, and common mistakes to avoid. Atmospheric Environment, 2014. 94: p. 647-662.

Battelle Memorial Institute and Center for International Earth Science Information Network - CIESIN - Columbia University, Global Annual Average PM2.5 Grids from MODIS and MISR Aerosol Optical Depth(AOD). 2013, Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

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

Changes in manuscript: no changes were made for this comment.

Furthermore, the comparison results indicate that PM2.5 concentrations are underestimated significantly in some months (e.g., MFB=-48% in July), and the model performance can be quite different in different regions. The author needs to comment how these temporally and spatially variant biases affect the simulation results of spatiotemporal distribution of O3, PM2.5 and PM2.5 chemical composition.

Responses: The temporally and spatially variant biases of PM2.5 do affect the simulated spatiotemporal distribution of PM2.5 and its chemical composition. PM2.5 is more underpredicted in summer when the concentrations are lower, so the predicted seasonal variation of PM2.5 is stronger. PM2.5 is more underpredicted in NW where the concentrations are lower, so the predicted spatial difference between NW and eastern China region (i.e., NCP, YRD, etc) is likely stronger. It should also affect the spatiotemporal distribution of PM2.5 chemical compositions, but no detailed information can be obtained due to the lack of detailed PM2.5 composition observations. The biases of O3 exhibit much

less variation temporally and spatially, so the predicted spatiotemporal distribution of O3 is more accurate than PM2.5.
Changes in manuscript: We added the above discussion in Section 3.3 in lines 396-404 in the revised manuscript.

(2) Introduction: The author suggests that most modeling studies focus on a specific pollution episode and extensive model performance evaluation is lacking. In fact, as far as I know, quite a few studies have been done to evaluate the model performance in China for a full year or several representative months (e.g., Gao et al., 2014; Zhang et al., 2016; Liu et al., 2016; Zhao et al., 2013; Wang et al., 2011; Liu et al., 2010), and there are more. The author should review these long-term modeling studies because they highly resemble the work presented here.

Gao, Y., Zhao, C., Liu, X. H., Zhang, M. G., and Leung, L. R.: WRF-Chem simulations of aerosols and anthropogenic aerosol radiative forcing in East Asia, Atmos Environ, 92, 250-266, DOI 10.1016/j.atmosenv.2014.04.038, 2014.

Zhang, Y., Zhang, X., Wang, L., Zhang, Q., Duan, F., and He, K.: Application of WRF/Chem over East Asia: Part I. Model evaluation and intercomparison with MM5/CMAQ, Atmos Environ, 124, 285-300, 10.1016/j.atmosenv.2015.07.022, 2016.

Liu, X. Y., Zhang, Y., Zhang, Q., and He, M. B.: Application of online-coupled WRF/Chem-MADRID in East Asia: Model evaluation and climatic effects of anthropogenic aerosols, Atmos Environ, 124, 321-336, 10.1016/j.atmosenv.2015.03.052, 2016.

Zhao, B., Wang, S. X., Wang, J. D., Fu, J. S., Liu, T. H., Xu, J. Y., Fu, X., and Hao, J. M.: Impact of national NOx and SO2 control policies on particulate matter pollution in China, Atmos Environ, 77, 453-463, DOI 10.1016/j.atmosenv.2013.05.012, 2013.

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

Liu, X.-H., Zhang, Y., Cheng, S.-H., Xing, J., Zhang, Q., Streets, D. G., Jang, C., Wang, W.-X., and Hao, J.-M.: Understanding of regional air pollution over china using CMAQ, part I performance evaluation and seasonal variation, Atmos Environ, 44, 2415-2426, 10.1016/j.atmosenv.2010.03.035, 2010.

Responses: We thank the reviewer for pointing out the references. The focus of the current manuscript is to evaluate the model performance of surface level O3 and PM2.5. Although a full year or several representative months' air quality simulations have been conducted previously, model performance on temporal and spatial variations of air pollutants were mostly evaluated against available surface observation at a limited number of sites. In addition, the surface observations were mostly based on the MEP's air pollution index (API) numbers, which could be used to calculate the concentrations of the major pollutants of SO2, NO2 or PM10. Therefore, it is still true that no studies have reported "the

detailed model performance of O3 and PM2.5 for an entire year". We have modified the introduction section to include the above facts and cited the above references.
Changes in manuscript: We added above discussion in Section 1 lines 103-110 in the revised manuscript.

Specific comments:
(1) Line 222-225: Why does the author filter out these data? How are the thresholds determined?
Responses: We performed quality control checks on the raw observation data, and filtered out data that are either unrealistically high or show abnormal temporal variations, which could greatly bias the model performance analysis. For the extreme values, we choose 250 ppb for hourly O3 and 1500 μg m^{-3} for hourly PM2.5. These cut-off values were chosen based on past experience in regional air quality modeling. While locally, these extreme values might be possible, they were likely not representative of the regional average concentrations for a 36x36 km^2 grid cell. We also removed days that had standard deviation less than 5 ppb for O3 or 5 μg m^{-3} for PM2.5. This is also based on the general understanding of the typical diurnal variations of O3 and PM2.5 in polluted urban areas, and examination of the data at all the monitoring stations collected for this study.
Changes in manuscript: no changes were made for this comment.

(2) Line 228: There should be a comma before "PM2.5"
Responses: Corrected it.
Changes in manuscript: add a comma before "PM2.5" in line 235.

(3) Line 284-291: The PM2.5 concentrations are underpredicted significantly in some months. The author should explain the reason for the underestimation. In addition, the author attributes the underestimation in PM10 to natural and anthropogenic dust emissions. How is wind-blown dust emissions calculated in CMAQ (any reference)? Are there any previous studies showing that the dust emission module embedded in CMAQ underpredict wind-blown dust emissions?
Responses: PM2.5 is generally more underpredicted in the warm months (April-July) than in the cold months (November-March). We think that SOA underprediction is likely an important reason for this phenomenon and we have discussed it in lines 488-498.

Reviewer 2 also commented on the dust module (comment #8) and details about the dust module can be found in our response to that comment. Briefly we used the CMAQ inline dust module with modifications to use the land use types in MODIS rather than the BEIS, as it only works for the United States. As previous studies by Fu et al. (2013) and Dong et al. (2015) reported that the dust module significantly underestimated the emission of total dust, it is possible that dust emissions were estimated in the current study as well.
Changes in manuscript: We added the PM10 results in Table 3 and Table 4. We added the discussions of PM10 in line 235, lines 297-299, and line 305 in the revised manuscript.

(4) Line 197: The author states that the benchmarks are adapted from Emery et al. (2012). However, the author indicates that these benchmarks are from Emery et al. (2001) in the title of Table 1. The two papers seem quite different and the latter one appears the correct source. Please confirm.
Responses: It is Emery et al. (2001). We have corrected it.

Changes in manuscript: We have corrected it in line 204 in the revised manuscript.