Anonymous Referee #1

This paper compares the modelling surface NO_2 concentrations from a regional CTM CHIMERE driven by the satellite-derived DECSO and the bottom-up MIX emission inventories, against in-situ measurements over China. The detailed evaluation is well demonstrated, and the model performance seems sound. Nonetheless, this work lacks of impressive new findings or insights. Substantial improvements are needed before publication.

Response: We thank you for the comments. To our knowledge, this study is the first validation of modelling NO_2 results with the widespread in situ network over China, which became recently available. The validation results reveal the bias of the spatial distribution of bottom-up NO_x emissions inventories and for the first time point out that such bias can be reduced significantly by applying the spatial distribution of NO_x emissions in the satellite-derived inventory as the spatial proxy for the bottom-up inventory. In addition, this study explores a reasonable scheme to classify MEP in-situ stations over China, which helps to draw more solid conclusions on the ability of the model to reproduce the measurements. The ground-based air quality monitoring network of MEP in China is a valuable database to validate model simulations, however it lacks provision of detailed information about the stations, including classification and height, and background stations which better represent regional pollution levels. Using measurements from all stations blindly without classification may bias the conclusive statements of the validation results, as the measurements may not represent the mean of a grid box of the simulations. The classification scheme developed in this study aiming to select measurements better representing grid-mean is expected to be very useful for following users of the MEP measurements.

Other reviewers have further affirmed the novelty of this study. As pointed out by other two reviewers, “The paper is of scientific importance because this is the first significant publication of data from this network.” and “The comparison and description of the NO_2 observations is very good and this data set will be of interest to the science community. The conclusions that the MIX inventory totals for 2010 along with the updated 2015 satellite derived spatial distribution provide the most accurate prediction of 2015 surface NO_2 concentrations is interesting.”

All the other comments have been addressed carefully below.

Major comments

1) Pg4, Ln26: The authors used the MIX 2010 data, and declared that it was suitable for the NOx simulations spanning over the year of 2015, as DECSO NOx emissions are highly similar in year 2010 and 2015. This reasoning is not rigorous, if this work focuses on the comparisons of different emission inventories. Would the NOx emissions in bottom-up inventories also be highly similar in year 2010 and 2015? Furthermore, how about the changes in anthropogenic emissions of SO2, CO, VOCs and other reactive species between 2010 and 2015 both in DECSO and MIX? Both the differences in emissions for NOx (reactive species) and other species (e.g., CO, VOCs) between DECSO and MIX should impact the concentrations of NO_2 through interfering with gas phase oxidation, particles through aerosol-radiation interactions and heterogeneous oxidation of NOx).

Response: We used the MIX inventory for 2010 instead of 2015, because the MIX inventory was developed only for the years 2008 and 2010 and this is the latest available gridded bottom-up inventories for this region. However, the usage of the 2010 inventory is not expected to bring significant bias, as the magnitude of NO_x emissions for 2010 is comparable to that for 2015 in both the bottom-up inventory MEIC (Liu et al, 2016) and the satellite-derived inventory DECSO, even though there is a five-year lag. For the period of 2010–2012, the NO_x emissions of China experienced a rapid growth. A sharp decline of NO_x emissions was observed in the years of 2013–2015, with a peak around 2012. We have clarified this in Section 2.2 of the manuscript as follows:

"We focused on 2015 as the most recent year with available DECSO emission estimates and in situ measurements, but we used the MIX inventory for 2010, because the year 2015 is not available yet. However, the use of the 2010 MIX inventory without scaling is not expected to bring significant bias, as the similarity of NO_x emissions for 2010 and 2015 has been reported by both the bottom-up inventory MEIC (Liu et al, 2016a) and the satellite-derived inventory DECSO. For the period of 2010–2012, the NO_x emissions of China experienced a rapid growth. A sharp decline of NO_x emissions was observed in the years of 2013–2015, with a peak around 2012 (Liu et al., 2016a). As a result, the inventory for 2010 is comparable to that for 2015, even though there is a five-year lag. Figure 1 compares DECSO NO_x emissions for 2015 (left) and 2010 (middle), which are consistent in both total amount (21.5 vs 21.6 Tg) and spatial distribution (r= 0.83)."

Concerning the emissions of other species, we used the same 2010 MIX emissions in this study, as they are the most recent emissions that are public available. From 2010 to 2015, the anthropogenic emissions of SO_2, CO and NMVOC for China has decreased by 2%, 5% and increased by 21%, respectively (Li et al., 2017). We agree that the differences in emissions for other species between 2010 and 2015 should affect the concentrations of NO_2. However, it is reasonable to use the 2010 MIX inventory without scaling, because the percentages of emission changes are far less than the uncertainty ranges of emission estimates and scaling without considering changes of spatial distributions will bring larger errors and make the validation less reliable. We have added the discussion in Section 3.1 of the manuscript, as follows:

"Note that the use of the 2010 MIX emissions for other species including SO_2, CO and NMVOC, due to the lack of the 2015 inventory, may introduce uncertainties to the simulating NO_2. The anthropogenic emissions of SO_2, CO and NMVOC for China have been reported to decrease by 2%, 5% and increase by 21% from 2010 to 2015, respectively (Li et al., 2017b). In gas-phase chemistry, the principal sink of NO_x is oxidation to HNO_3. The influence of the growth in NMVOC on the oxidizing power of the atmosphere is partially compensated by the reduction in CO, as CO and hydrocarbons play similar roles in depleting oxidants following the HO_x-NO_x-CO-hydrocarbon chemical mechanisms (Jacob, 2000). Additionally, SO_2 contributes to influence NO_2 concentrations by forming aerosols, concentrations of which have impact on photolysis rates and thus photochemical reaction rates associated with NO_x (Mailler et al., 2017). Nevertheless, we still believe it is reasonable to use the 2010 inventory without scaling for other species, considering the emission changes are rather small compared to the uncertainties in bottom-up estimates and scaling without taking the changes of spatial distributions into account may bring larger errors to make the validation less reliable."

2) I am not sure whether this work focuses primarily on the comparisons of different emission inventories? If yes, the strengths and limitations of satellite-derived DECSO and the bottom-up MIX emission inventories have not been well discussed. Which emission inventory is better and recommended? Would the evaluation results change if a different model or chemical mechanism is used?

Response: This work aims not only to compare bottom-up and satellite-derived inventories to identify their inconsistency, but also to gain better knowledge of the reasons for the

inconsistency and then shed light on improved emission quantification approaches. Both inventories have their own strengths and limitation, which have been detailed in our previous work (Ding et al., 2017). We have summarized them in Section 2.2 of the revised manuscript, as follows:

"Both inventories show comparable spatial distributions at a national and regional scale, but distinctions between urban and rural areas (see Sect.3.1). The strength of the MIX inventory is that it includes detailed source-category information (e.g., power plant and transportation sector) for emissions, which is useful for driving atmospheric models and designing emission mitigation policies, but not included in DECSO. The advantage of the DECSO inventory is that emissions are timely updated (as soon as the satellite observations are available); while bottom-up inventories usually lag behind a few years and are outdated by the time they become available. In addition, the spatial information in DECSO is based on OMI NO_2 observations, while MIX relies on spatial proxies like GDP to allocate emissions due to the lack of data. An in-depth comparison between inventories has been described by Ding et al. (2017)."

It is difficult to determine which inventory is better. The model accurately reproduces the spatial variability of NO_2 from in-situ measurements, with a spatial correlation coefficient of over 0.7 for simulations based on both inventories, indicating the reliability of both inventories. A negative and positive bias is found for the simulation with the satellite-derived and the bottom-up inventory respectively, suggesting an underestimation and overestimation of NO_x emissions from corresponding inventories. However, we do find the consistency with in-situ measurements improves when correcting the spatial distribution of NO_x emissions in the bottom-up inventory by that in the satellite-derived inventory, which suggests a promising approach to improve inventories in the future. We have clarified this in the conclusion section, as follows:

"The bias between observed and modelled concentrations was reduced significantly with the slope decreasing from 1.3 to 1.0, when the spatial distribution of NO_x emissions in the DECSO inventory is applied as the spatial proxy for the MIX inventory. The reduced bias suggests an improvement of the distribution of emissions between urban and suburban/rural areas in the DECSO inventory compared to that used in the bottom-up inventory, which shed light on addressing the spatial errors in bottom-up inventories."

The validation results may change, as the simulated concentrations of NO_2 depend on the model and chemical mechanism used. However, we don't expect fundamental changes in our

conclusions. Firstly, the CHIMERE model is considered as a state-of-the-art model and its good performance has been confirmed by many previous studies (Marécal et al., 2015). For China, the CHIMERE results are always close to the 9-model ensemble that was developed within the EU FP7 programme two collaborative research projects, MarcoPolo and Panda (available at http://www.marcopolo-panda.eu/forecast/). Second, the modelling results with the available gas phase chemical mechanisms implemented in CHIMERE (including MELCHIOR and SAPRC07) produce similar quantities of HO_x radicals and ozone (Menut et al., 2013), indicating quite comparable levels of NO_2. Third, the positive bias of the bottom-up inventory has been confirmed by other models in certain regions of China as well, e.g., CMAQ in Jiangsu (Zhou et al., 2017) and Hebei (Zheng et al., 2017). It is a pity that we don't have the capacity to perform an in-depth comparison of multiple models with variable chemistry schemes using both inventories, but the current evidences give us confidence in the accuracy of the CHIMERE simulations.

3) This paper made considerable efforts in classification of a total of 1413 air quality monitoring sites (e.g., urban sites, suburban sites, and sites share a grid cell). This information might not be that important, and should be presented briefly (or put into supporting information). A 0.25 resolution is a bit coarse in terms of mesoscale air quality simulations. If more sites appear in one cell, the errors would counteract. Would the evaluation results change if a different resolution is used?

Response: We treat the classification scheme as a highlight of this work. As mentioned previously, the ground-based air quality monitoring network of MEP in China is a valuable database to validate model, however it lacks provision of detailed information about the stations, including classification and height, and background stations which better represent regional air pollution levels. Using measurements from all stations blindly without classification may bias the conclusive statements of the validation results, as the measurements may not represent the mean of a grid box of the simulations. The classification scheme developed in this study aiming to select measurements better representing grid-mean is expected to be very useful for following users of the MEP measurements.

Performing simulations at higher resolutions requires finer-resolution emission inventories. MIX is developed at the resolution of 0.25° × 0.25° by allocating emissions using spatial proxies.

DECSO is developed at the same resolution, determined by the comparable resolution of OMI NO_2 observations. Finer-resolution inventories deriving from downscaling without additional information on locations of emitting facilities are expected to have significantly larger uncertainties compared to the original inventories (Zheng et al., 2017). The uncertainties will be propagated into biases of modelling results and make the validation results less reliable. Thus, we chose the resolution of 0.25° × 0.25°, consistent with that of MIX and DECSO, to make the validation more reliable.

4) Table 2 in Pg19: DECSO and MIX underestimated and overestimated the observed NOx, respectively. Then the corrected DECSO, scaled from the original DECSO using the ratio of MIX NOx to DECSO NOx over the simulation domain, simulated better results. What is the point in corrected DECSO (quite arbitrary) and its evaluation statistics? Note that in Figure 2, NOx emissions from MIX and DECSO show large discrepancies in spatial distribution especially in North China Plain.

Response: The aim of the scaling is to better compare the spatial distributions of the two inventories and identify the sensitivity of simulating NO_2 on spatial distributions of NO_x emissions, considering the complicated influence of emissions on modelling NO_2 as pointed out by the reviewer in the previous comment. We have clarified this in Section 3.1 of the revised manuscript, as follows:

"In order to better compare the spatial distributions of the two inventories and identify the sensitivity of model performance on spatial distributions of emissions, we further evaluate the impact of the spatial distribution of emissions on simulating NO_2 by applying the same spatial proxy for NO_x emissions in both inventories."

5) Description regarding the model configuration (especially the gas phase mechanism) should be added.

Response: We thank you for the suggestion and add the model configuration in Section 2.1, as follows:

"We use the CHIMERE model v2013b over East Asia (18°N to 50°N and 102°E to 132°E) with a resolution of 0.25° following the configuration in Ding et al. (2015). The CHIMERE simulation was driven by operational meteorological data from the European Centre for

Medium-Range Weather Forecasts (ECMWF) with a horizontal resolution of 0.25°. Atmospheric variables were simulated in 8 layers from the surface to 500 hPa. Tropospheric photochemistry is represented using the reduced MELCHIOR chemical mechanism (Derognat et al., 2003), including about 120 reactions and 44 gaseous species. Aerosol module accounting both for inorganic and organic species of primary or secondary origin is included according to Bessagnet et al. (2004). Boundary conditions for the model domain were derived from monthly mean climatology based on Model for OZone And Related chemical Tracers (MOZART) second-generation (Horowitz et al., 2003) for gases, the Laboratoire de Météorologie Dynamique Zoom – Interaction avec la Chimie et les Aérosols (LMDz-INCA; Folberth et al., 2006) for nitrate and ammonium, and the Georgia Tech/Goddard Global Ozone Chemistry Aerosol Radiation and Transport (GOCART, Ginoux et al., 2001) for other aerosols."

Minor comments

1) Pg4, Ln3: the speciation of NOx needs further discussion. The recommended ratios are from Generation of European Emission Data for Episodes (GENEMIS), then their applicability over China? How would these ratios impact the NOx simulations?

Response: We agree that the speciation of NO_x may have impact on simulated NO_2. However, there is no widely accepted recommendation for the NO_x speciation over China and the speciation approach is rarely mentioned by the previous air quality modelling work over China. We do find limited descriptions about the speciation, which are very close to our choice. For example, Fu et al. (2009) divides the inventory NO_x into 90% NO and 10% NO_2. Thus we do not expect significant changes to our findings. A sensitivity study of how the speciation of NO_x will influence the simulation is expected in the future, but this is out of the scope for this work.

2) Pg5, Ln16: the reviewer also uses the data from air quality monitoring sites of the MEP network. How to treat the monitoring sites with abundant and even overwhelming missing data?

Response: We are not aware of any consensus about discarding monitoring sites based on the quantity of missing data. Thus, all the measurements in the dates with 24-hour valid measurements (lager than 0) are used for the analysis in this study. We have clarified it in the Sect. 2.3 of the revised manuscript, as follows:

"Note that only the measurements for the dates with 24-hour valid measurements (larger than 0) are used for the following analysis in this study."

3) Table 2, add the unit for RMSE. Please add NME.

Response: Thanks. We have added them in the revised manuscript.

4) Figure 10, please add "LST" or "UTC".

Response: Thanks. We have added "LST" in the revised plot.

5) Pg11, Ln3: the data source for the daily profile of NOx emissions.

Response: Thanks. The data source is the Selected Nomenclature for sources of Air Pollution Prototype (SNAP) diurnal profiles (Menut et al., 2012). We have added it in the revised manuscript.

6) Pg11, Ln6, and also in the Abstract: discussions regarding the boundary layer mixing need more verification (figures or tables).

Response: We attributed the nocturnal bias to uncertainties in boundary layers based on the earlier model evaluations of CHIMERE. Supporting information including figures and tables can be derived from Lampe et al. (2009), which is out of the scope of this study. We have clarified this in Sect. 3.3 of the revised manuscript, as follows:

"NO_2 concentrations simulated by the model cannot reproduce the observed temporal pattern at night, but present constantly high values, probably caused by unrealistically low boundary layer heights and too little vertical turbulence in the model (Bessagnet et al., 2016). This has been further confirmed by the earlier evaluation of the diurnal cycle of trace gases as modelled by CHIMERE in Lampe et al. (2009)."

References

Bessagnet, B., Hodzic, A., Vautard, R., Beekmann, M., Cheinet, S., Honoré, C., Liousse, C., and Rouil, L.: Aerosol modeling with CHIMERE—preliminary evaluation at the continental scale, Atmos. Environ., 38, 2803–2817, 2004.

Bessagnet, B., Pirovano, G., Mircea, M., Cuvelier, C., Aulinger, A., Calori, G., Ciarelli, G., Manders, A., Stern, R., Tsyro, S., García Vivanco, M., Thunis, P., Pay, M. T., Colette, A., Couvidat, F., Meleux, F., Rouïl, L., Ung, A., Aksoyoglu, S., Baldasano, J. M., Bieser, J., Briganti, G., Cappelletti, A., D'Isidoro, M., Finardi, S., Kranenburg, R., Silibello, C., Carnevale, C., Aas, W., Dupont, J. C., Fagerli, H., Gonzalez, L., Menut, L.,

Prévôt, A. S. H., Roberts, P., and White, L.: Presentation of the EURODELTA III intercomparison exercise – evaluation of the chemistry transport models' performance on criteria pollutants and joint analysis with meteorology, Atmos. Chem. Phys., 16, 12667–12701, doi: 10.5194/acp-16-12667-2016, 2016.

Derognat, C., Beekmann, M., Baeumle, M., Martin, D., and Schmidt, H.: Effect of biogenic volatile organic compound emissions on tropospheric chemistry during the Atmospheric Pollution Over the Paris Area (ESQUIF) campaign in the Ile-de-France region, J. Geophys. Res., 108, D17, doi: 10.1029/2001JD001421, 2003.

Ding, J., van der A, R. J., Mijling, B., Levelt, P. F., and Hao, N.: NO_x emission estimates during the 2014 Youth Olympic Games in Nanjing, Atmos. Chem. Phys., 15, 9399–9412, doi:10.5194/acp-15-9399-2015, 2015.

Ding, J., Miyazaki, K., van der A, R. J., Mijling, B., Kurokawa, J. I., Cho, S., Janssens-Maenhout, G., Zhang, Q., Liu, F., and Levelt, P. F.: Intercomparison of NO_x emission inventories over East Asia, Atmos. Chem. Phys., 17, 10125–10141, doi: 10.5194/acp-17-10125-2017, 2017.

Folberth, G. A., Hauglustaine, D. A., Lathière, J., and Brocheton, F.: Interactive chemistry in the Laboratoire de Météorologie Dynamique general circulation model: model description and impact analysis of biogenic hydrocarbons on tropospheric chemistry, Atmos. Chem. Phys., 6, 2273–2319, doi: 10.5194/acp-6-2273-2006, 2006.

Fu, J. S., Streets, D. G., Jang, C. J., Kebin He, j. H., He, K., Wang, L., and Zhang, Q.: Modeling Regional/Urban Ozone and Particulate Matter in Beijing, China, J. Air Waste Manage. Assoc., 59, 37–44, 2009.

Ginoux, P., Chin, M., Tegen, I., Prospero, J. M., Holben, B., Dubovik, O., and Lin, S.-J.: Sources and distributions of dust aerosols simulated with the GOCART model, J. Geophys. Res., 106, 20255–20273, doi: 10.1029/2000JD000053, 2001.

Horowitz, L. W., Walters, S., Mauzerall, D. L., Emmons, L. K., Rasch, P. J., Granier, C., Tie, X., Lamarque, J.-F., Schultz, M. G., Tyndall, G. S., Orlando, J. J., and Brasseur, G. P.: A global simulation of tropospheric ozone and related tracers: Description and evaluation of MOZART, version 2, J. Geophys. Res., 108, 4784, doi: 10.1029/2002JD002853, 2003.

Jacob, D. J.: Heterogeneous chemistry and tropospheric ozone, Atmos. Environ., 34, 2131–2159, 2000.

Lampe, A.: Evaluation of the diurnal cycle of trace gases over Beijing as modelled by CHIMERE, KNMI report / Master Thesis, University of Twente, 2009.

Li, M., Liu, H., Geng, G., Hong, C., Liu, F., Song, Y., Tong, D., Zheng, B., Cui, H., Man, H., Zhang, Q., and He, K.: Anthropogenic emission inventories in China: a review, National Science Review, nwx150, 2017.

Liu, F., Zhang, Q., Ronald, J. v. d. A., Zheng, B., Tong, D., Yan, L., Zheng, Y., and He, K.: Recent reduction in NO_x emissions over China: synthesis of satellite observations and emission inventories, Environ. Res. Lett., 11, 114002, 2016.

Mailler, S., Menut, L., Khvorostyanov, D., Valari, M., Couvidat, F., Siour, G., Turquety, S., Briant, R., Tuccella, P., Bessagnet, B., Colette, A., Létinois, L., Markakis, K., and Meleux, F.: CHIMERE-2017: from urban to hemispheric chemistry-transport modeling, Geosci. Model Dev., 10, 2397–2423, doi: 10.5194/gmd-10-2397-2017, 2017.

Marécal, V., Peuch, V. H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., Benedictow, A., Bergström, R., Bessagnet, B., Cansado, A., Chéroux, F., Colette, A., Coman, A., Curier, R. L., Denier van der Gon, H. A. C., Drouin, A., Elbern, H., Emili, E., Engelen, R. J., Eskes, H. J., Foret, G., Friese, E., Gauss, M., Giannaros, C., Guth, J., Joly, M., Jaumouillé, E., Josse, B., Kadygrov, N., Kaiser, J. W., Krajsek, K., Kuenen, J., Kumar, U., Liora, N., Lopez, E., Malherbe, L., Martinez, I., Melas, D., Meleux, F., Menut, L., Moinat, P., Morales, T., Parmentier, J., Piacentini, A., Plu, M., Poupkou, A., Queguiner, S., Robertson, L., Rouïl, L., Schaap, M., Segers, A., Sofiev, M., Tarasson, L., Thomas, M., Timmermans, R., Valdebenito, Á., van Velthoven, P., van Versendaal, R., Vira, J., and Ung, A.: A regional air quality forecasting system over Europe: the MACC-II daily ensemble production, Geosci. Model Dev., 8, 2777–2813, doi: 10.5194/gmd-8-2777-2015, 2015.

Menut, L., Goussebaile, A., Bessagnet, B., Khvorostiyanov, D., and Ung, A.: Impact of realistic hourly emissions profiles on air pollutants concentrations modelled with CHIMERE, Atmos. Environ., 49, 233–244, 2012.

Menut, L., Bessagnet, B., Khvorostyanov, D., Beekmann, M., Blond, N., Colette, A., Coll, I., Curci, G., Foret, G., Hodzic, A., Mailler, S., Meleux, F., Monge, J. L., Pison, I., Siour, G., Turquety, S., Valari, M., Vautard, R., and Vivanco, M. G.: CHIMERE 2013: a model for regional atmospheric composition modelling, Geosci. Model Dev., 6, 981–1028, doi: 10.5194/gmd-6-981-2013, 2013.

Zheng, B., Zhang, Q., Tong, D., Chen, C., Hong, C., Li, M., Geng, G., Lei, Y., Huo, H., and He, K.: Resolution dependence of uncertainties in gridded emission inventories: a case study in Hebei, China, Atmos. Chem. Phys., 17, 921–933, doi: 10.5194/acp-17-921-2017, 2017.

Zhou, Y., Zhao, Y., Mao, P., Zhang, Q., Zhang, J., Qiu, L., and Yang, Y.: Development of a high-resolution emission inventory and its evaluation and application through air quality modeling for Jiangsu Province, China, Atmos. Chem. Phys., 17, 211–233, doi: 10.5194/acp-17-211-2017, 2017.