

We thank Reviewer for his/her constructive comments.

Response to the Specific comments.

General comments: This paper presents overview about Phase III of the chemical transport model inter-comparison study MICS-ASIA for East Asia region. The atmospheric models participating in Phase III and its simulation framework have greatly improved from the previous MICS-ASIA Phase II. And, the calculation results are compared with the observations in industrial China, which was not done in the Phase II. So, this paper introducing MICS-ASIA Phased III is believed to have certain academic value. However, in the manuscript at the present time, there are many problems such as the sentences being too long, and the lack of the necessary information to convince the authors' interpretation to the results. Then, the manuscript should be revised according to the following comments as well as many other specific comments before the publication in ACP.

Reply: We totally agree with the reviewer. In the new manuscript, we accepted all comments suggested by the reviewer.

Comment 1: About the length of the manuscript. it seems that the manuscript is too long compared to its contents. The things to be claimed should be focused (probably on what is stated in summary or the abstract), and the descriptions not related to those should be removed or simplified. The figures or their contents which are not necessary for the main line should be also omitted.

Reply: We totally agree. In the revised manuscript, words have been cut back by 15-20%. 25% figures (Fig. 5, 6 and 11) and related discussions n (i.e. emissions) were also deleted. The revised manuscript included "1 Introduction; 2. Model validation(annual and monthly variation of surface O₃, NO and NO₂, surface O₃ diurnal variation, and O₃ vertical profiles); 3. Spatial distribution of O₃ and its comparison with MICS-Asia II, 4. Discussion (comparison with observed dry velocity and boundary layer height, relationships between O₃ with NO_x), 5. Summary"

Comment 2. On the comparison of model results and measured values. Most models have rough resolution (horizontal direction: 45 km, vertical direction: 58m near the ground), and it is not shown whether the observed values to be compared represent the extent of that range. If many measuring stations are unevenly distributed in a grid cell at locations with high NO_x emissions, the effect of titration there is greater than the grid cell average. So, actually the models overestimating the measured ozone concentration may be correct.

Reply: We agree with the reviewer that the rough resolution may affect the model evaluation. In this study, observation data were taken from 1) Chinese Ecosystem Research Network (EA1); 2) Pearl River Delta Regional Air Quality Monitoring Network (PRD RAQMN) (EA2); 3) the Acid Deposition Monitoring Network in East Asia (EANET) (EA3). Observations were rarely affected by the very local emissions

around sites, and were used to represent the regional air quality.

- As listed in Table R1 in this reply, most stations are located in rural, remote and clear urban regions in EA1. Fig. R1 presents the scatter plots of NO emissions in 45 and 3km emission inventory. Emission errors resulting from coarse grids were not significant in most stations. This implied that observation generally represents the 45km averages of ozone.

Table R1 site descriptions in Chinese Ecosystem Research Network

Site	Site characteristics	Longitude, latitude
Xinglong	Remote	117.576 40.394
Lingshan	Remote	115.431 39.968
Yangfang	Rural	116.11 40.13
Xianghe	Suburban	116.962 39.754
Langfang	Suburban	116.689 39.549
Zhuozhou	Suburban	115.99 39.46
Datong	Suburban	113.389 40.089
Zhangjiakou	Suburban	114.918 40.771
Cangzhou	Suburban	116.779 38.286
Yanjiao	Suburban	116.824 39.961
Beijing	Urban	116.372 39.974
Baoding	Urban	115.441 38.824
Shijiazhuang	Urban	114.529 38.028
Chengde*	Urban	117.925 40.973
Tianjin	Urban	117.206 39.075
Tanggu*	Urban	117.717 39.044
Caofeidian*	Urban	118.442 39.270
Tangshan	Urban	118.156 39.624
Qian'an*	Urban	114.800 40.100

*cities are clear, and annual $PM_{2.5} < 35 \mu g/m^3$

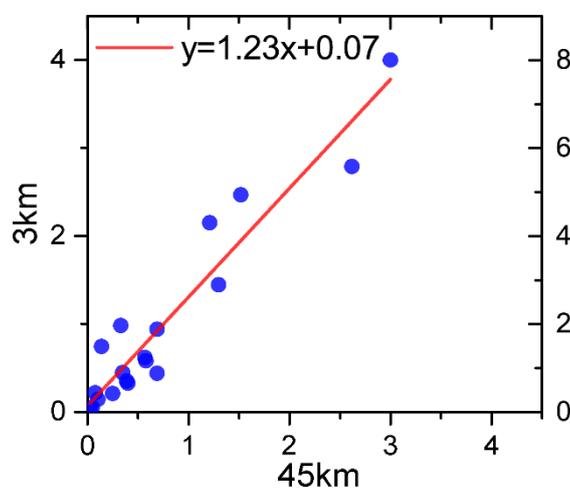


Fig.R1 Scatter plots of NO emission rates ($\mu g/m^2/s$) at observation sites in EA1 in 45km and 3km

resolution emission inventory (MEIC)

- Pearl River Delta Regional Air Quality Monitoring Network (PRD RAQMN) was jointly established by the Guangdong Provincial Environmental Monitoring Centre (GDEMC) and the Environmental Protection Department of the Hong Kong Special Administrative Region (HKEPD) from 2003 to 2005. The PRD RAQMN was to probe the regional air quality, assess the effectiveness of emission reduction measures and enhance the roles of monitoring networks in characterizing regional air quality and supporting air quality management (Zhong et al.,2013). So sites are rarely affected by the local emissions near them. Fig. R2 showed the Spatial distribution of average concentrations of NO₂ and O₃ in the PRD-RAQMN Network. Concentrations of pollutants were smooth. The effect of very local emissions was rarely seen.

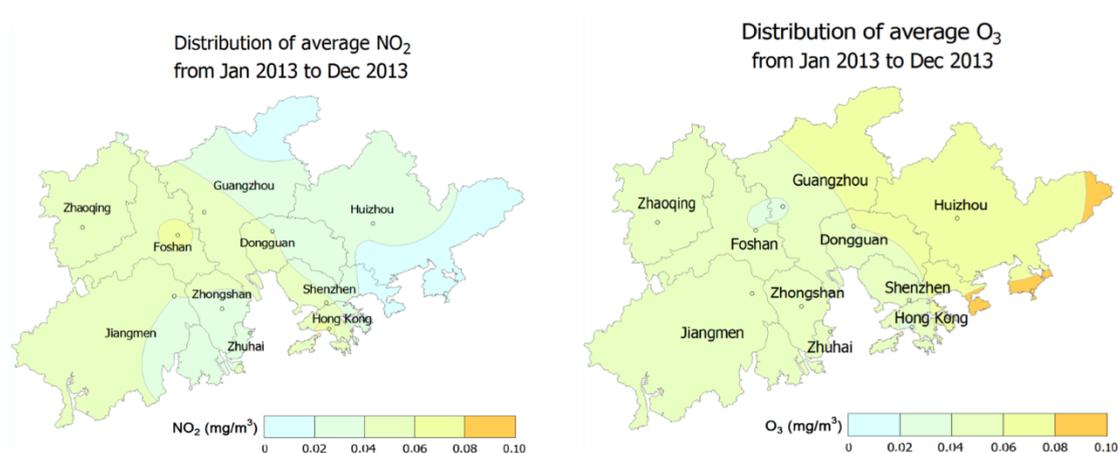


Fig.R2 Spatial distribution of average concentrations of NO₂ and O₃ in the PRD-RAQMN Network, figure is annual report of Pearl River Delta Regional Air Quality Monitoring Network in 2013 (https://www.epd.gov.hk/epd/sites/default/files//epd/english/resources_pub/publications/files/PRD_2013_report_en.pdf)

- Sites in EANET are mostly located in oceanic regions (Hedo, Ogasawara and Oki) and remote regions (Rishiri, Ochiishi, Yusuhara, Sado-seki, Happo). More information can be found in Ban et al. (2016).

Changes in the revised manuscript: Page 7 Line 20-25; Page 8 Line 1-5, Line 11.

Comment 3. About the investigation of intermodel variability on O₃ (chapter.4) In phase II of the MICS-ASIA, because input data (weather, emissions, boundary condition) are different, it was not possible to specify how much each process of chemistry, vertical diffusion, and dry deposition in the model contributed to calculated ozone variation among models. In the Phase III of this time, although common input data were provided to avoid it, it seems in this paper that the contribution of each of the above processes could not be specified again because the post process of these data differs between models. If the above guess is true, it seems better to clearly state it and

to give up the brute forth evaluation of the contribution of each of the above process inspections 4.3-4.5. On the other hand, if you stick to say that you could specify the contribution of each of the above processes, you should add thoroughly the information described in the following so that the reader can understand its rationality.

Reply: We totally agree. In MICS-Asia III, we found that there were significant model biases and intermodel variability in summer ozone in North China Plain and Western Pacific. These findings were not revealed in phase II of MICS-Asia. This point is beyond we expected before MICS-Asia III. Hence, one issue we are facing is to explain the bias causes or provide a future direction on analysis for MICS-Asia IV. We agree the reviewer that quantifying the contribution of each process processes (vertical mixing, horizontal advection, gaseous and heterogeneous chemistry, dry and wet deposition, emissions and model resolution...) is important to explain model bias. Sensitivity simulation is a good way. But this requires a tremendous amount of computational cost and data space for 14 models. Designing sensitivity simulating scenarios with acceptable costs is essential to next studies. The MICS-Asia III has not directly output the contribution of each process, so we did a qualitative analysis on potential causes by comparison between models and observations to narrow sensitivity simulating scenarios for MICS-Asia IV. We believe that this is also helpful for other model developers to improve model performance in East Asia. In MICS-Asia II, related discusses were mostly based on guesses because meteorology, emissions, model domain, boundary conditions were quite different. In MCIS-Asia III, common input data provide a good chance for this qualitative analysis.

We agree with the reviewer that brute forth evaluation of the contribution of processes may cause errors or uncertainties. In the revised manuscript, we collected observation data on key parameters of potential processes as much as possible. Our focus was the model evaluation on these parameters, which has not been conducted by previous phase of MICS-Asia. So we changed the title from “Investigation of intermodel variability on O₃” to “Discussion”.

As shown in Fig. R3, ensemble average dry deposition velocity of O₃ underestimated observations in August-September by 30-50% in EA1. This underestimation decreased the deposition amounts of surface O₃ and partly explained the overestimation of ensemble simulated O₃ in summer. This is consistent with intermodel comparison between M11 with M1-M6. M11 reproduced observed surface O₃ in EA1 in May-July. The higher dry deposition velocities in M11 between May-July (0.3 cm/s) contributed to low surface O₃ than M1-M6. This implied that we should conducted the sensitivity analysis on dry deposition to quantify its impact on EA1 surface ozone in MICS-Asia IV. In EA4, simulated dry deposition velocity agreed well with observations, so there could be other reasons responsible for overestimation in EA4.

Previous studies revealed that O₃ precursors are mostly constrained within the boundary layer (Quan et al., 2013). The model evaluation on PBLH and turbulent kinetic energy is essential for the interpretation of model biases with observations.

Unfortunately, few observations on turbulent kinetic energy were directly measured in East Asia. Fig. R4 presents the comparison between simulated and observed PBLH. In EA1, all the selected models exhibited the spring-maximum and winter-minimum season cycle, which captured the major pattern of climatology of PBLH observations (Guo et al.,2016). The Ense on PBLH was 100-200 m higher than radiosonde measurements. This is likely caused by the inconsistency of samples between models and measurements. The simulation was the mean value of 12 hours (08:00-20:00), while the average of measurements was calculated based on 3 hours (08:00, 14:00 and 20:00). In MICS-Asia IV, more model evaluation on turbulent kinetic energy is urgent.

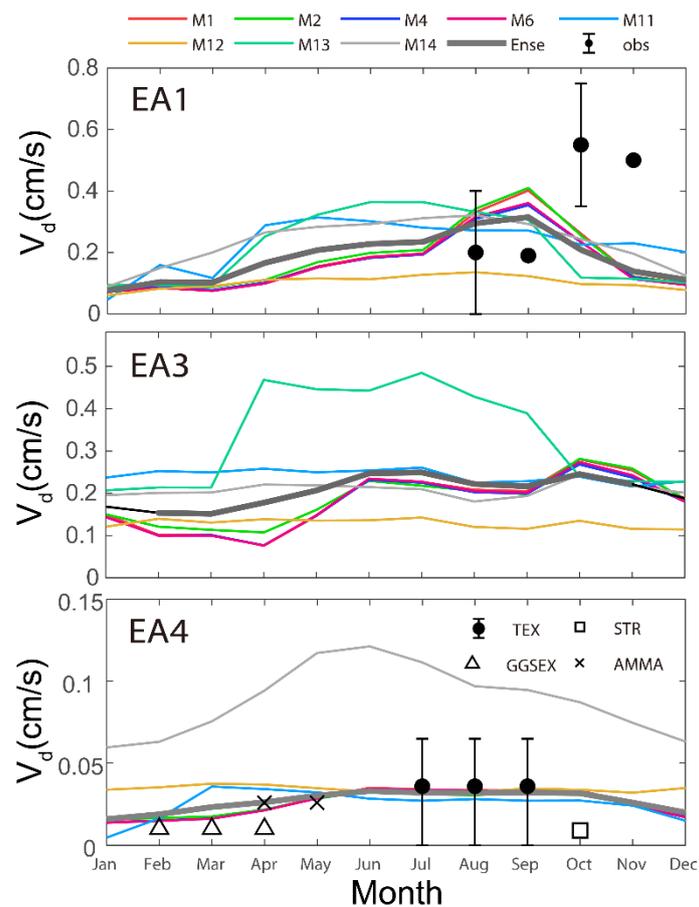


Fig. R3 Simulated and observed monthly dry deposition velocity

In the revised manuscript, we moved vertical profile of O₃ into the section “model evaluation”, and observations in EA3 and EA4 were added. In general, ensemble means (Ense) presented an underestimation and overestimation for EA3 O₃ in middle (500-800 hpa) and lower (below 900 hpa) troposphere, respectively. In winter, the underestimation even extended to 200hpa in winter. The magnitudes of underestimation and overestimation reached 10-40 ppbv and 10-20 ppbv. In EA4, Ense reproduced the vertical structure of ozone in both summer and winter. An overestimation existed below 800 hpa in summer, with a magnitude of 10-20 ppbv.

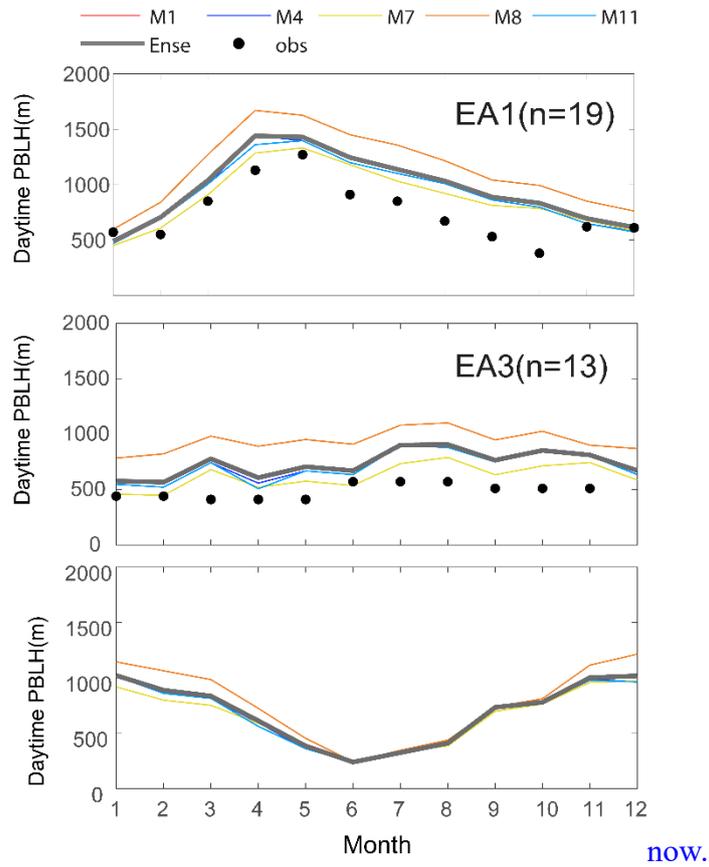


Fig. R4 Simulated and observed monthly daytime PBLH

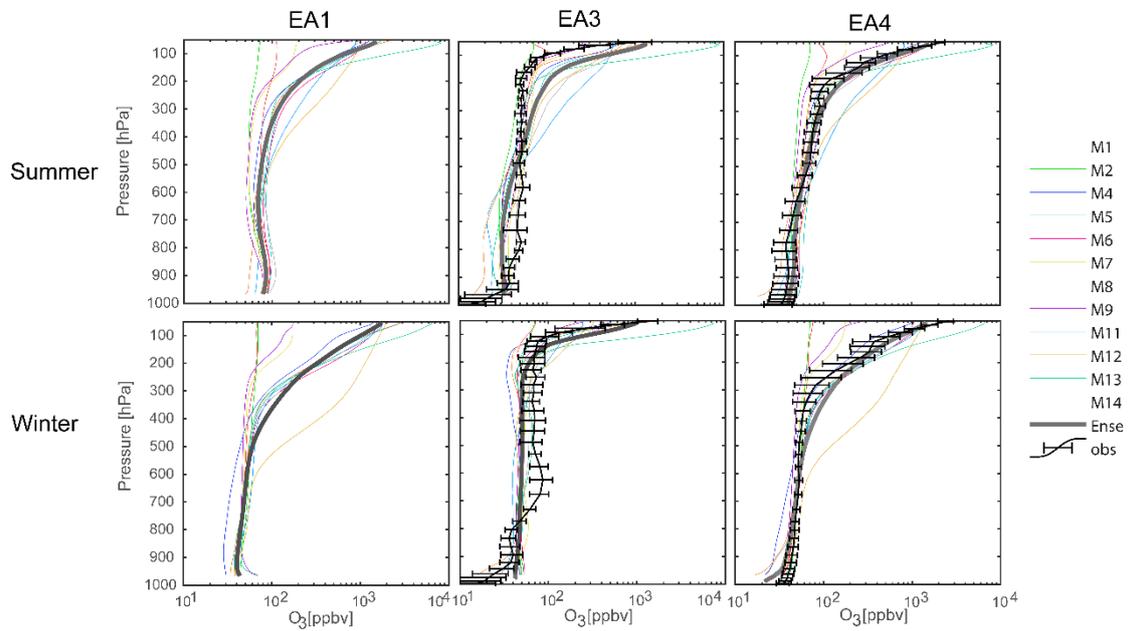


Fig. R5 Simulated and observed O_3 profiles in summer and winter of 2010, averaged over all observed stations in three subregions over East Asia (EA1: left column, EA3: middle column, EA4: bottom column). The ozonesonde data observe in 2010 was taken from the data base stored by World Ozone and Ultraviolet Radiation Data Centre (WOUDC).

The evaluation on chemistry in models is a difficult problem all along. As far as we know, there are no direct measurements on ozone production rates in East China till now. The relationships between O₃ with its precursors usually was regarded as an effective index on chemistry. We realized that the simple comparison between O₃ with NO_x could bring errors or uncertainties. Hence, the relationship only was used to qualitative analyze the intermodel variability on chemistry, more quantitative analysis will be conducted in MICS-Asia IV. We believe that this qualitative analysis is helpful to model developer. For example, we found that the slope and intercept between O₃ and NO_x in M11 (the best performance of O₃ in EA1) were closer to observations. The lower slope (-1.02) in M11 than M1-M6 (-1.31 - -2.25) indicated a weaker ozone chemical production intensity. This is validated by Akimoto et al. (2019) in which ozone chemical production in M11 was 60% of M1.

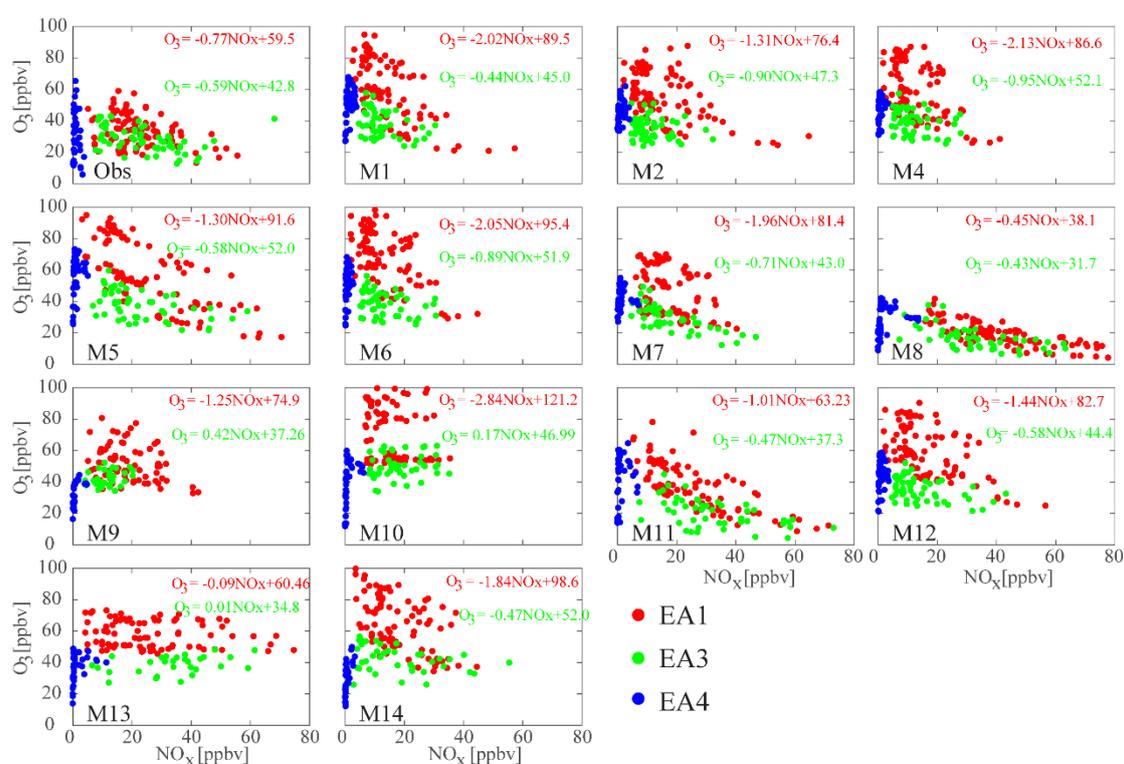


Fig. R9 Scatter plots between monthly daytime (08:00-20:00) surface NO_x and O₃ at each station over EA1 (red), EA3 (green) and EA4 (blue) in May-October, for observations (Obs) and models

Changes in the revised manuscript: Section 3.5; Section 5.

Comment 4. About authors' interpretation of the results. Many parts cannot be convinced about the interpretation of the results by the author mainly because the differences among each model (e.g., differences of boundary conditions estimated with Mozart, Chaser, and by default settings, differences in dry deposition model, differences in sub-grid scale parameterization such as convection, differences in PBL model, and differences in spatiotemporal distribution of emissions) are not specifically mentioned. For relevant parts other than chapter 4, I will point out each of the following "other

specific comments”

Reply: We understand the reviewer. The large divergence on parameterizations and emissions among models is always a difficult problem in air quality model intercomparison projects. Hence, some intercomparison projects like HTAP v1 conducted by United Nations, CityDelta by Europe Union and AQMEII employed models with different resolutions and various meteorology. Sometimes, different lateral boundary conditions were used in regional models (CityDelta, AQMEII). This increased the difficulty of interpretation. In MICS-Asia III, most models employed the same emissions, meteorology and resolution, which provide a good chance to explore the impact of parameterization on ozone.

As mentioned by the reviewer, no specifying the contribution of processes could bring errors or uncertainties to the interpretation of the results. So we moved our focus from interpretation of the results to the model evaluation on key parameters of processes by collecting their observations (dry deposition velocity, PBLH, vertical profiles) as much as possible. We hope our analysis is helpful to detailed model intercomparison in next studies and other model developers in East Asia.

We revised our manuscript according to your flowing comments.

“Other specific comments:

Comment 5: p.5 L2-3 Is the problem (3) really addressed? I don’t think so, as I already mentioned in the general comments

Reply: In the revised manuscript, we deleted the problem3.

Changes in the revised manuscript: Page 5 Line 1-4.

Comment 6: p.5 L10-11 You mean to interpolate model outputs to locations of observations both horizontally and vertically? If yes, please show that method in detail. It may get rid of my concern mentioned in the general comments.

Reply: Firstly, we determine the model grid cell indexes of observation sites from their longitude, latitude, and height above sea levels. If there are two or more sites in one grid, we will select their mean values to compare with model outputs in this grid.

Changes in the revised manuscript: Page 5 Line 9-10.

Comment 7: p.5 L24 Fig.1 does not introduce WRF model.

Reply: In the revised manuscript, we added a description “The domain of meteorological fields is shown in Fig.1”.

Changes in the revised manuscript: Page 5 Line 23-23.

Comment 8: p.6 L28-p.7L1 Please identify which model adopt the projection by themselves.

Reply: M13 and M14 made the projection by themselves

Changes in the revised manuscript: Page 6 Line 27-28.

Comment 9: p.7 L5 I think two references should be moved after the names of the universities are introduced in L6

Reply: We revised it.

Changes in the revised manuscript: Page 7 Line 3-4.

Comment 10: p.7 L9 Are the models making boundary conditions depending on their own previous experience denoted by "default" in table 1? If yes, I think the phrase such as "their own" is better in table 1.

Reply: We revised it.

Changes in the revised manuscript: Page 34 Table 1.

Comment 11: p.9 L4 Is the word "total" necessary?

Reply: We deleted it.

Changes in the revised manuscript: Page 9 Line 16.

Comment 12: p.9 L5 M12 seems also an exception as well as M11.

Reply: We agree, and revised it in the new manuscript.

Changes in the revised manuscript: we delete this sentence.

Comment 13: p.9 L11-12 Is a two-peak seasonal cycle for O₃? If yes, I see there are three peaks but not two. And I see observations show three-peak but not one-peak.

Reply: We revised this sentence. "In EA3, most models (except M7, M8 and M11) exhibited high O₃ concentrations in March-May and September-November. Observed O₃ showed that the highest concentrations appeared in October-November."

Changes in the revised manuscript: Page 9 Line 18-21.

Comment 14: p.9 L22 "Similar results have been found in MICS-Asia II" seems contradict to the statement in L5-L7 of p.4.

Reply: Thanks. In L5- L7 of P4, the underestimation of simulated O₃ appeared in spring (March) and winter (December) during the MCS-Asia II. In this study, our reported overestimation of O₃ was in May-October (L22 P9). The periods in P4 and P9 are different.

Changes in the revised manuscript: None.

Comment 15: p.10 L24-25 Show the evidence for the slight overestimation of 10 ppbv in M11 due to difficulties in dealing with vertical mixing.

Reply: In M11, the minimum of vertical diffusivity was set to be $0.5 \text{ m}^2 \text{ s}^{-1}$. This value is a little higher than other models (e.g. CAMx: $0.1 \text{ m}^2 \text{ s}^{-1}$). In the stable boundary layer on nighttime, the higher vertical diffusivity may transport high ozone in upper layer to the surface, and also uplifted surface NO. The lower NO weakens the ozone titration.

We realized that vertical mixing may be not the only reason of nighttime ozone overestimation in M11. We needed more observed evidence to support our conclusion. So, we deleted it in the revised manuscript.

Changes in the revised manuscript: We deleted it.

Comment 16: p.10 L25-26 Show the evidence for the significant improvement of the model performance in winter, compared to in summer, due to the weak intensity of photochemical reactions.

Reply: Thanks. As shown in Table R2, ensemble simulated ozone (Ense) in winter was closer to observations than summer. The ratio between Ense and Observation was 1.28, much lower than 1.69 in summer. The intensity of overestimation increased from winter to summer, with the increase of solar radiation. This implied that the treatment of photochemical reactions in models may play an important role in this overestimation.

Table R2 Observed and ensemble simulated ozone (Ense) in EA1

Season	Observation	Ense	Ense/Obs
Winter (Dec-Feb)	12.6	16.1	1.28
Spring (Mar-May)	25.6	34.6	1.35
Summer (Jun-Aug)	38.0	64.4	1.69
Autumn (Sep-Nov)	14.9	23.6	1.58

Changes in the revised manuscript: None.

Comment 17: p.11 L17 Add explanation how to derive the statics in table 2, 3 and 4 to clarify which part of the spatiotemporal deviations from the observations are included in the static

Reply: We add the definition of these statics in Appendix A in the revised manuscript.

Changes in the revised manuscript: Page 21 Line 25.

Comment 18: p.12 L12-13 Show the evidence for that the treatment of models on chemistry, vertical diffusion and dry deposition have contributed to the underestimation of NO.

Reply: Thanks.

Changes in the revised manuscript: We delete this sentence.

Comment 19: p.13 L8-10 I can't understand why you selected the PBLH, emissions fluxes, dry deposition velocities, relationships between NO_x and O₃, and the vertical profiles of O₃ and its precursors to compare.

Reply: Thanks for your comments. In the revised manuscript, we collected related observations to evaluate the model performance, as discussed in Comment 3.

Changes in the revised manuscript: Section 3.5; Section 5.

Comment 20: p.16 L23-L24 Jin et al (2015) perhaps showed the ozone formation regime at 1330 LST (overpass time of OMI) while you show that between 1000-1800 LST. Also, your results include NO_x titration effect while Jin et al (2015)'s results did not. So, I think it is not appropriate to compare them directly.

Reply: We agree.

Changes in the revised manuscript: we deleted this reference

Comment 21: p/17 L8-9 In M11, O₃ does not seem positively correlated with NO_x.

Reply: Sorry. M9 and M10 were positively correlated with NO_x, instead of M8 and M11. In the revised manuscript, we revised it.

Changes in the revised manuscript: Page 18 Line 21

Comment 22: p.18 L17-18 Show the evidence that difference of concentrations are related to the treatments of convection and cloud activity among models.

Reply: Thanks. Fig. R5 showed the simulated and observed O₃ profiles in EA3. Clearly, the most significant underestimation and inter-variability of models appeared in 950-700 hpa (~0.5-2.5 km). The climatology of ozone sounding revealed a high relative humidity (about 80%) and enhanced ozone layer in this layer (0.5-2 km) in summer (Leung et al., 2003). Leung et al. (2004) stated that the ozone in this layer was likely from convection of photochemical production in the polluted boundary layer, based on the simultaneous occurrence of high ozone mixing ratio and high relative humidity. In MICS-Asia III, horizontal resolution is 45 km, which was not enough to explicitly simulate the convection. So sub-grid parameterization in models may played an important in the underestimation and inter-variability. We realized that these are not direct evidence because impact of convections in models were not output. Hence, we delete this sentence in the revised manuscript.

Changes in the revised manuscript: we delete this sentence in the revised manuscript.

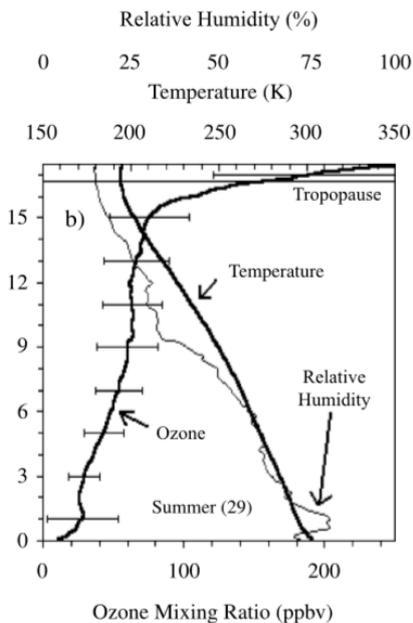


Fig. R10 Seasonally averaged ozone profiles in the troposphere above Hong Kong summer

Comment 23: p.19 L22-23 The locations of the place names shown in the text are not known for the foreign readers. So, you should show these place names in Fig.10.

Reply: We plotted place names in Fig. S1 in the revised manuscript.

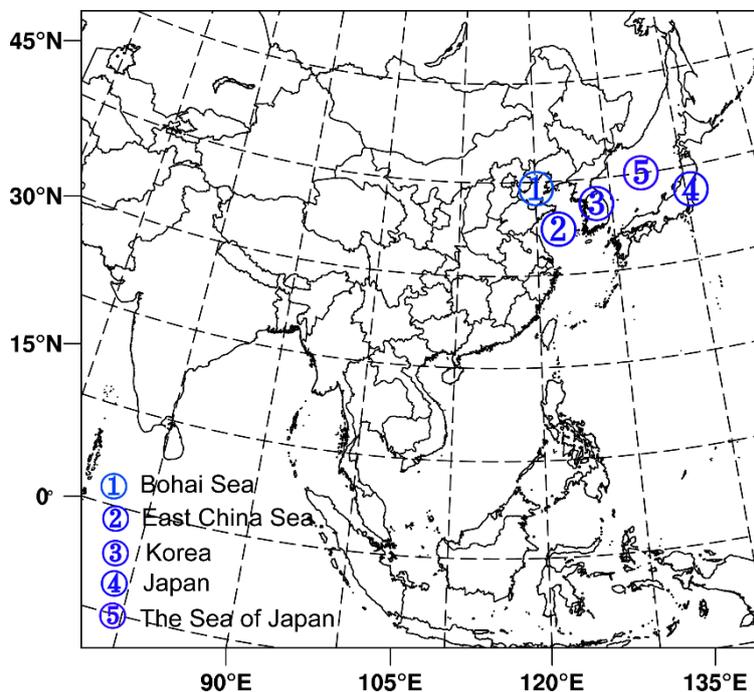


Fig. R11 Locations of related regions

Changes in the revised manuscript: Page 14 Line 19.

Comment 24: p.20 L16-17 Before you have the statement in L16-17, you should show that the wind fields are actually the same between the models which estimate 30 ppbv

or higher O₃ mixing ratio and those which estimate lower O₃ mixing ratio. And, how do you think about the difference of emissions that was discussed in section 4.2

Reply: We agree. In the revised manuscript, we showed the simulated wind fields by models. Winds between models were similar. In section 4.2, we found that EA1 emissions in M1, M4 and M11 are similar, but the simulated O₃ between these three models the western Pacific Ocean showed a O₃ discrepancy. So, there could be other causes responsible for this discrepancy, besides emissions in source regions.

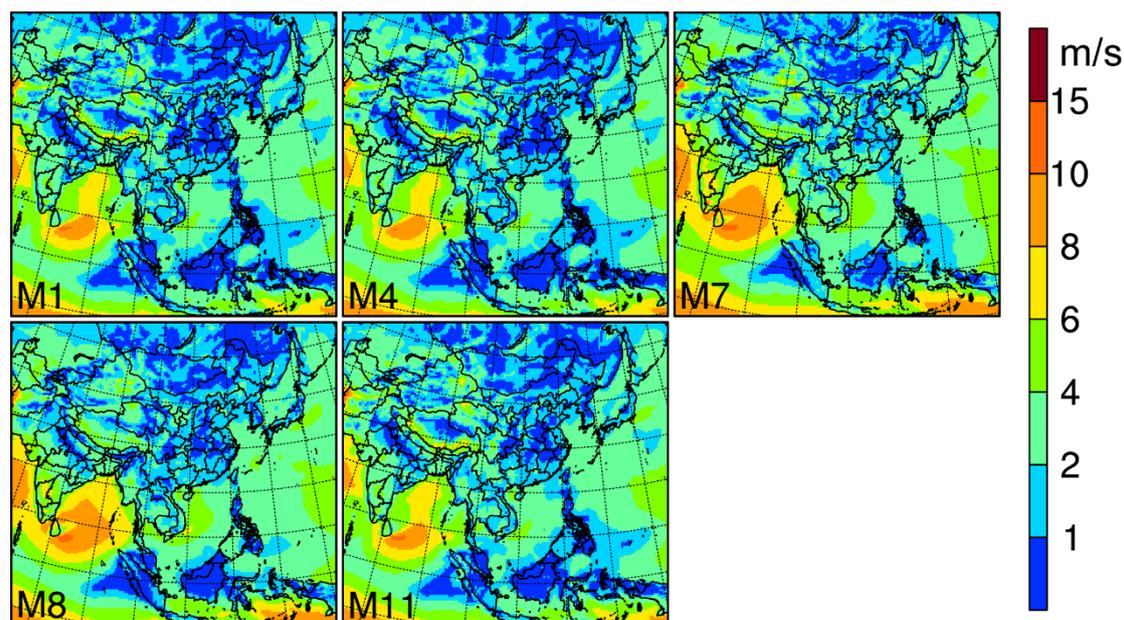


Fig.R12 simulated surface wind velocities(m/s) in MICS-Asia III

Changes in the revised manuscript: Page 14 Line 26.

Comment 25 p.33 L9 I guess the meteorological model used for providing meteorological fields with most models also use the domain in Fig.1. If yes, please mention about that too.

Reply: We added this point in the revised manuscript.

Changes in the revised manuscript: Page 32 Line 20.

Comment 26 p.33 L14 Please add a description of the symbol such as "+" or "-" in Fig.2.

Reply: We added a description in the caption of Fig.2.

Changes in the revised manuscript: Figure 2.

Comment 27: p.46 Fig.3 and p.47 Fig.4 The kinds of color of the curve in the figures is too many to distinguish. Are all the models need to be distinguished by different colors?

Reply: Sorry for trouble you in Fig.2 and 3. An aim of MICS-Asia III is to examine the models' performance for O₃ in East Asia, and provide useful information to improve model ability. As the first step, we need discuss the strengths of individual models and tell the readers as much as possible. Then we will compare the parametrization of this model with others and explore why it exhibit a better performance. In this respect we need label each model in Fig.2 and 3. We listed the performance of individual models in section 3.2. For example, we mentioned that M11 was closer to O₃ observations in EA1. In our another manuscript, we compared M11 parametrization of transport, vertical diffusion and heterogeneous chemistry with M1 and M6. This is helpful to improve the model.

Changes in the revised manuscript: None.

Technical corrections:

Comment 28: p.3 L15 You need space between "2013" and "(Wang et al., 2017)". You can find the similar mistake to miss spaces elsewhere in the manuscript.

Reply: We revised it

Changes in the revised manuscript: We revised it.

Comment 29: p.10 L4"4)" should be removed.

Reply: We revised it

Changes in the revised manuscript: we delete.

Comment 30: p.19 L23 I think "predicated" should be "predicted".

Reply: We revised it

Changes in the revised manuscript: we corrected it.

Comment 31: p.20 L1"EA1" should be moved right after "source regions

Reply: We revised it

Changes in the revised manuscript: we corrected it.

References:

Akimoto, H., Nagashima, T., Li, J., Fu, J. S., Ji, D., Tan, J., and Wang, Z.: Comparison of surface ozone simulation among selected regional models in MICS-Asia III – effects of chemistry and vertical transport for the causes of difference, Atmos. Chem. Phys., 19, 603-615, <https://doi.org/10.5194/acp-19-603-2019>, 2019.

Ban, S. , Matsuda, K. , Sato, K. , & Ohizumi, T. . (2016). Long-term assessment of nitrogen deposition at remote EANET sites in japan. *Atmospheric Environment*, 146, 70-78.

Guo, J., Miao, Y., Zhang, Y., Liu, H., Li, Z., Zhang, W., He, J., Lou, M., Yan, Y., Bian, L., and Zhai, P.: The climatology of planetary boundary layer height in China derived from radiosonde and reanalysis data, *Atmos. Chem. Phys.*, 16, 13309-13319, <https://doi.org/10.5194/acp-16-13309-2016>, 2016

Han, Z., Sakurai, T., Ueda, H., Carmichael, G. R., Streets, D., Hayami, H., Wang, Z., Holloway, T., Engardt, M., Hozumib, Y., Parkh, S.U., Kajinoi, M., Sarteletj, K., Funk, C., Bennetg, C., Thongboonchooc, N., Tangc, Y., Changk, A., Matsudal, K., Amannm, M. : MICS-ASIA II: model intercomparison and evaluation of ozone and relevant species, *Atmos. Environ.*, 42(15), 3491-3509,2008.

Leung Y K , Chang W L , Chan Y W . Some characteristics of ozone profiles above Hong Kong. *Meteorology and Atmospheric Physics*, 87(4):279-291, 2004.

Liuju Zhong, Peter K.K. Louie*, Junyu Zheng, K.M. Wai, Josephine W.K. Ho, Zibing Yuan, Alexis K.H. Lau, Dingli Yue, Yan Zhou, The Pearl River Delta Regional Air Quality Monitoring Network – Regional Collaborative Efforts on Joint Air Quality Management, *Aerosol and Air Quality Research*, 13: 1582–1597, 2013

Quan, J., Tie, X., Zhang, Q., Liu, Q., Li, X., Gao, Y., and Zhao, D.: Evolution of planetary boundary layer under different weather conditions, and its impact on aerosol concentrations, *Particuology*, 11(1), 34-40, 2013.

Zhong, L., Louie, P., Zheng, J., Wai, K. M., Josephine W.K. Ho, Yuan, Z., Lau, A. K. H., Yue, D., Zhou, Y.: The Pearl River Delta Regional Air Quality Monitoring Network – Regional Collaborative Efforts on Joint Air Quality Management, *Aerosol and Air Quality Research*, 13: 1582–1597, 2013

We thank Reviewer for his/her constructive comments.

Response to the Specific comments.

General comments: This paper describe the ability of an ensemble of regional chemistry-transport models to reproduce surface ozone pollution in East Asia as well as NO_x concentrations. Indeed, recent observations do show that surface ozone concentrations are still in-creasing in China which underline the necessity to have good forecasting tools and means to set-up and control mitigation policies. This intercomparison is conducted in the framework of the Model Inter-Comparison Study for Asia phase III (MICS-ASIA III) which is the follow-up of MICS-ASIA II (2003) and MICS-ASIA I (1998). 13 models are cross compared for a one-year simulation (2010). The simulation suits are based on state-of-the-art CTMs. Simulations are compared to available observations with specially observations available on industrialized China which was not the case of MICS-ASIA II. Also, the dispersion of the simulations are investigated to understand what reasons could explain models differences. Compared to European or American are as, the models have more difficulties to reproduced observed concentrations and the median of the ensemble do not always over skilled single models like it is the case for European ensembles. Such exercises have been proven useful to improve modelling suits and for this reason this paper is interesting for the community. The work conducted in that case is important and this study deserved to be published in ACP journal but corrections are probably needed to make the paper more efficient and to fulfill the high level standard of quality of the journal. I will list the comments and questions I still have on this work and that could help, i hope, to improve it.

Reply: Thanks a lot for your insightful comments. We accept all your comments in the revised manuscript.

Comment 1: The analysis of the skills of an ensemble is always complicated. To be more clear and to have stronger messages, i suggest you to first analysis skills using the average of the ensemble and then to discuss the single models. By this way, it will allow to clearly identify the main biases either for seasonal analysis either for diurnal analysis and then discuss singularities.

Reply: We totally agree. We firstly evaluate the ensemble performance in each section of the revised manuscript.

In section 3.1,

“The O₃ NMB and RMSE of ensemble mean were significantly less than the ensemble median in most situations (Table 1). Therefore, we only presented the results of multi-model mean ensemble (Ense). In general, the majority of models significantly overestimated annual surface O₃ compared with the observations in EA1, EA3 and EA4 (Fig. 2). Ense overestimated surface O₃ by 10-15 ppbv in these subregions. Ense NO₂ was generally close to the observations to within $\pm 20\%$ in all subregions. In EA1 and EA3, Ense NO was 5-10 ppbv lower than observation, and showed a reasonable

performance in EA4.”

In section 3.2,

“From the perspective of monthly variation, the overestimation of O₃ mostly appeared in May-September in EA1. Ense O₃ was 10-30 ppbv higher than observations, 30-70% of observed values. In the same period (May-September), Ense NO and NO₂ appeared to be consistent with observations, attaining mean biases of < 3 ppbv. This suggests that the intercomparison on O₃ production efficiency per NO_x with observations is needed. In EA3, Ense O₃ agreed well with observed high autumn O₃, but overestimated from January to September by 5-15 ppbv (15-60% of observations). This maximum of overestimation appeared in March-April (15ppbv), which led to a spring peak in simulated O₃ which was not found in observations. This overestimation was partly related to the underestimation of NO in the same months, which decreased the titration effect. For NO₂, Ense agreed well with observed values in June-December, and slightly underestimated observations in January-May. In EA4, a significant overestimation of O₃ and underestimation of NO existed in June-October. Both observations and Ense NO were lower than 0.5 ppbv, so impact of by NO underestimation on O₃ are needed to be further explored. The ensemble NO₂ was generally close to the observations to within ±0.5 ppbv.”

In section 3.3,

“In general, model results for three sub-regions exhibited a larger spread with a magnitude of 10-50 ppbv throughout the diurnal cycle than that in Europe and North America (Solazzo et al., 2012). The Ense O₃ in summer exhibited a systematic overestimation (20 ppbv) throughout the diurnal cycle in EA1. This indicated that models had difficulty dealing with O₃ in North China Plain. Compared with summer, there was only a slight systematic overestimation of Ense O₃ in other seasons (3-5 ppbv). In EA3, Ense O₃ generally agreed with the observations in summer, autumn and winter. In particular, the O₃ maximum around noon was reproduced, reasonably. There was only a 3-5 ppbv overestimation during 16:00-23:00 and early morning (6:00-10:00). In spring, a systematic overestimation of Ense O₃ existed in the whole diurnal cycle (5-10 ppbv). In EA4, Ense captured the small diurnal variation of O₃ in four seasons, but significantly overestimated observations in summer and autumn (5-20 ppbv). In spring and winter, differences between Ense and observations were within 5 ppbv.”

In section 3.4,

“In general, Ense performed a better performance level than individual models for representing NO₂ in East Asia, reproducing the observed seasonal cycle and magnitudes. However, Ense did not always exhibited a superior performance for O₃ over certain individual model in East Asia, which was in contrast to its performance in Europe. M11 and M7 agreed well with observations in EA1 and EA3, while ENSE tended to overestimate O₃ concentrations in May-September in EA1 and January-September in EA3. Loon et al. (2007) indicated that ENSE exhibited a superior performance level

only when the spread of ensemble-model values was representative of the uncertainty of O₃. This indicated that most models did not reflect this uncertainty or missed key processes in MICS-Asia III.”

In section 3.5,

“In general, ensemble means (Ense) presented an underestimation and overestimation for EA3 O₃ in middle (500-800 hpa) and lower (below 900 hpa) troposphere, respectively. In winter, the underestimation even extended to 200hpa in winter. The magnitudes of underestimation and overestimation reached 10-40 ppbv and 10-20 ppbv. In EA4, Ense reproduced the vertical structure of ozone in both summer and winter. An overestimation existed below 800 hpa, with a magnitude of 10-20 ppbv.”

Changes in the revised manuscript: Page 8 Line 23-27; Page 9 Line 16-25; Page 10 Line 20-Page 11 Line 2; Page 13 Line 2-7.

Comment 2: Maybe also it would nice to have a more explicit but still short reminder of the physical processes driving the variability in each sub-region (i.e late maxima of ozone in EA3 quite different than EA1 and even EA4).

Reply: We totally agree. In the revised manuscript, we discussed the physical factors driving variability of each region on seasonal cycle.

“The East Asia monsoon played an important role in seasonal cycle of O₃ in subregions by the long-range transport. Besides local intensive photochemical productions, the O₃ summer maxima in EA1 were also affected by regional transport from Yangtze River Delta under prevailed summer southern monsoon (~20%) (Li et al., 2016). In EA3, a late maximum of O₃ in September-November was quite different from EA1 and EA4. This is largely attributed to the long-range transport of O₃ and its precursors in the polluted continental air masses from northern China and photochemical formation under dry and sunny weather conditions in autumn (Zheng et al., 2010). In EA4, the seasonal change of O₃ concentrations was characterized by two peaks in spring and autumn. The first and second peak in March–April and May-June were mainly influenced by the inflow from outside of East Asia and chemically produced O₃ by regional emissions, respectively. In the next studies, we will conduct the intermodel comparison on transport fluxes of O₃ between sub-regions over East Asia.”

Changes in the revised manuscript: Page 19 Line 15-25.

Comment 3: More informations about the nature of the stations and specifically about their representativity is needed. It is a key element of the model skills. Also, for NO₂ it exist sometimes biases (especially for stations far from sources) in the measurements when using molybden convertors devices since all nitrogen oxydes are measured instead of just NO₂, do you have checked this?

Reply: We agree. In this study, stations are taken from from 1) Chinese Ecosystem Research Network (EA1); 2) Pearl River Delta Regional Air Quality Monitoring

Network (PRD RAQMN) (EA2); 3) the Acid Deposition Monitoring Network in East Asia (EANET) (EA3). Observations were rarely affected by the very local emissions around sites, and were used to represent the regional air quality.

- As listed in Table R1 in this reply, most stations are located in rural, remote and clear urban regions in EA1. Fig. R1 presents the scatter plots of NO emissions in 45 and 3km model grid cell. Clearly, emission errors resulting from coarse grids were not significant in most stations. This implied that observation generally represents the 45km averages of ozone.

Table R1 site descriptions in Chinese Ecosystem Research Network

Site	Site characteristics	Longitude, latitude
Xinglong	Remote	117.576 40.394
Lingshan	Remote	115.431 39.968
Yangfang	Rural	116.11 40.13
Xianghe	Suburban	116.962 39.754
Langfang	Suburban	116.689 39.549
Zhuozhou	Suburban	115.99 39.46
Datong	Suburban	113.389 40.089
Zhangjiakou	Suburban	114.918 40.771
Cangzhou	Suburban	116.779 38.286
Yanjiao	Suburban	116.824 39.961
Beijing	Urban	116.372 39.974
Baoding	Urban	115.441 38.824
Shijiazhuang	Urban	114.529 38.028
Chengde*	Urban	117.925 40.973
Tianjin	Urban	117.206 39.075
Tanggu*	Urban	117.717 39.044
Caofeidian*	Urban	118.442 39.270
Tangshan	Urban	118.156 39.624
Qian'an*	Urban	114.800 40.100

*cities are clear, and annual $PM_{2.5} < 35 \mu\text{g}/\text{m}^3$

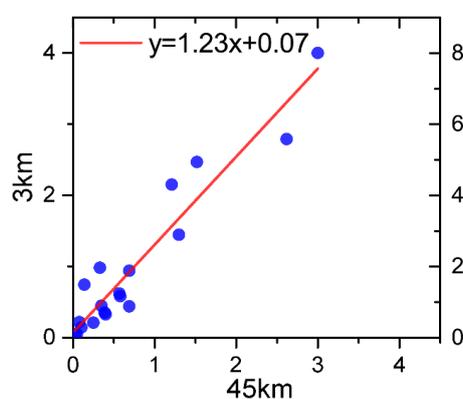


Fig.R1 Scatter plots of NO emission rates ($\mu\text{g}/\text{m}^2/\text{s}$) at observation sites in EA1 in 45km and 3km

- Pearl River Delta Regional Air Quality Monitoring Network (PRD RAQMN) was jointly established by the Guangdong Provincial Environmental Monitoring Centre (GDEMC) and the Environmental Protection Department of the Hong Kong Special Administrative Region (HKEPD) from 2003 to 2005. The PRD RAQMN was to probe the regional air quality, assess the effectiveness of emission reduction measures and enhance the roles of monitoring networks in characterizing regional air quality and supporting air quality management (Zhong et al.,2013). So sites are rarely affected by the local emissions near them. Fig. R2 showed the Spatial distribution of average concentrations of NO₂ and O₃ in the PRD-RAQMN Network. Obviously, concentrations of pollutants are smooth. The effect of very local emissions was not seen.

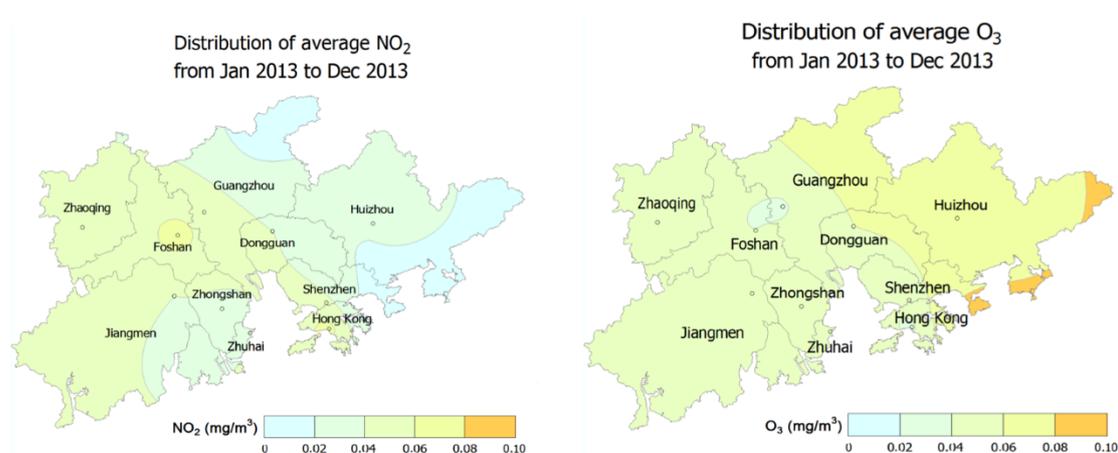


Fig.R2 Spatial distribution of average concentrations of NO₂ and O₃ in the PRD-RAQMN Network, figure is annual report of Pearl River Delta Regional Air Quality Monitoring Network in 2013 (https://www.epd.gov.hk/epd/sites/default/files//epd/english/resources_pub/publications/files/PRD_2013_report_en.pdf)

- Sites in EANET are mostly located in islands (Hedo, Ogasawara and Oki) and remote regions (Rishiri, Ochiishi, Yusuhara, Sado-seki, Happo). More information can be found in Ban et al. (2016).

As for NO₂ measurements, we agree that molybden convertors devices may cause errors. Ge et al. (2013) compared the measurements at an urban site in Beijing in summer by commercially standard chemiluminescence-based (called CL hereafter) instruments and Aerodyne Cavity Attenuated Phase Shift Spectroscopy (CAPS). The CAPS NO₂ monitor directly measures the absorption of NO₂ at the wavelength of 450 nm and requires no conversion of NO₂ to other species.

Fig. R3-R4 presents the comparison between instruments. Generally, the biggest discrepancy appeared in 12:00-16:00, with a magnitude of 10-20%. In other periods, NO₂ by CL and CAPS were similar. On average, discrepancies between CL and CAPS

were less than 10%. The linear fitting slope reached 0.999 between CL and CAPS.

As shown in Fig. R4, observations between CL and CAPS agreed well with each other with hourly $\text{NO}_2 > 15$ ppbv. In low hourly $\text{NO}_2 (< 10$ ppbv), CL NO_2 overestimated CAPS by 10-30%. This is consistent with the statement by the reviewers, which reported NO_2 exist sometimes biases for stations far from sources in the measurements.

In this study, we compared observed monthly mean NO_2 with models, instead of daytime NO_2 . This partly decreased the impact of errors from CL instrument. What's more, the observed NO_2 in EA1 and EA3 were 20 ppbv or more. In these high NO_x emission regions, biases from CL instruments may not bring too much impact on model validation. In EA4, most stations are located in islands or remote regions, with ~ 2 ppbv NO_2 . The CL NO_2 will overestimated NO_2 concentrations.

In the revised manuscript, we added a discussion on observation sites and instruments in section 2.3.

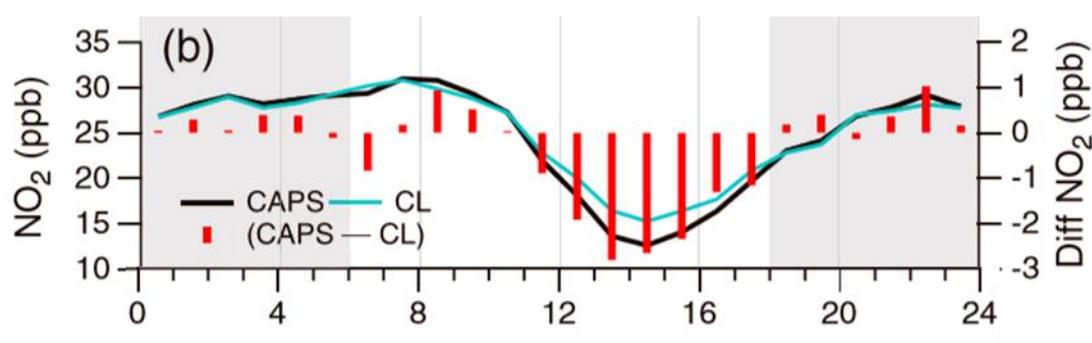


Fig. R3 Observed mean diurnal variation of NO_2 in summer in Beijing by chemiluminescence-based (CL) instruments and CAPS in Beijing. Also shown is the difference of two instruments.

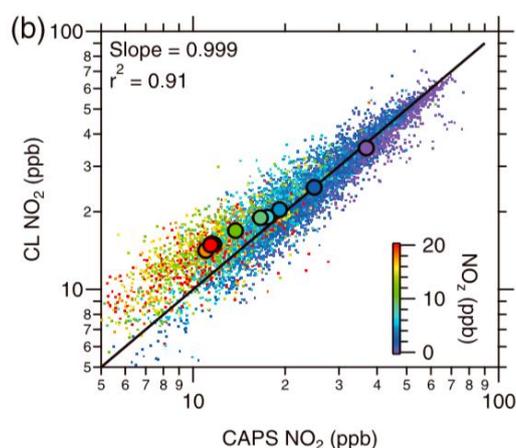


Fig. R4 Comparison of NO_2 measured by the CL NO_x analyzer and CAPS.

Changes in the revised manuscript: Page 7 Line 20-25; Page 8 Line 1-5, Line 11;

Page 8 Line 15-17.

Comment 4: I have the impression that authors do not need to include the EA2 region in the paper, you never use it in your discussions.

Reply: We agree. In the revised manuscript, we corrected it (EA1->EA1; EA3->EA2; EA4->EA3).

In this reply, we used EA1, EA3 and EA4 to give a clear comparison with the previous manuscript.

Changes in the revised manuscript: we corrected it (EA1->EA1; EA3->EA2; EA4->EA3).

Comment 5: Authors do evaluate several parameters relevant for model evaluation. It would have been better to have observations to put against models. It is often complicated to get all needed observations but maybe you can at least mention that in the perspectives. It becomes possible to have network ceilometers for PBLH evaluation. A lot of satellite observations are available to evaluate NO_x or ozone at larger scales. What about vertical profiles?

Reply: We totally agree. In the revised manuscript, we collected observation data as much as possible. The new observation data includes: 1) vertical profiles of O₃ in EA3 and EA4 from by World Ozone and Ultraviolet Radiation Data Centre (WOUDC); 2) PBLH in EA1 and EA3; 3) dry deposition velocities in EA1 and EA4. We also discussed the model performance against these observations.

Fig. R5 presents the simulated and observed O₃ profiles in subregions. Because there was a lack of O₃ sounding in EA1 in 2010, only observations in EA3 and EA4 are shown. In general, ensemble means (Ense) presented an underestimation and overestimation for EA3 O₃ in middle (500-800 hpa) and lower (below 900 hpa) troposphere, respectively. In winter, the underestimation even extended to 200hpa in winter. The magnitudes of underestimation and overestimation reached 10-40 ppbv and 10-20 ppbv. In EA4, Ense reproduced the vertical structure of ozone in both summer and winter. An overestimation existed below 800 hpa in summer, with a magnitude of 10-20 ppbv.

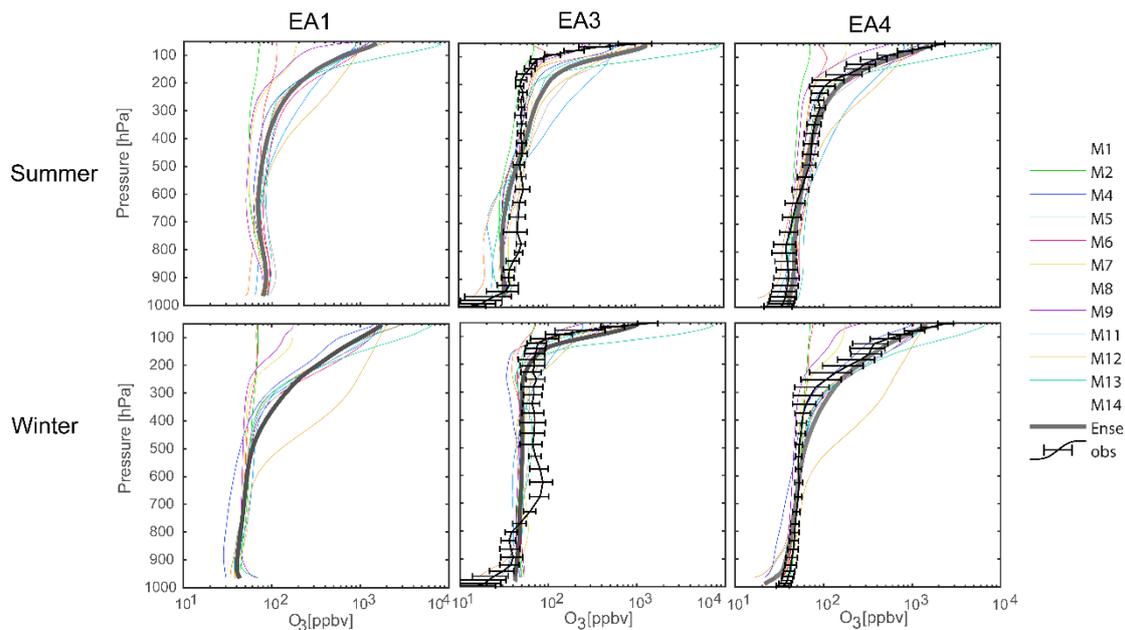


Fig. R5 Simulated and observed O_3 profiles in summer and winter of 2010, averaged over all observed stations in three subregions over East Asia (EA1: left column, EA3: middle column, EA4: bottom column). The ozonesonde data observe in 2010 was taken from the data base stored by World Ozone and Ultraviolet Radiation Data Centre (WOUDC).

On dry depositions, most models underestimated dry deposition velocities of O_3 (v_d) in August-September, but still fell into the range of observed standard deviation. This partly explained the overestimation of O_3 concentrations in summer discussed in section 3.2. In October-November, simulated v_d apparently overestimated observations by 30-50%.

In EA4, most stations were remote oceanic sites, and few dry deposition observations were conducted. So, we collected observations in other oceanic sites to evaluate model performance (Helmig et al., 2012). Tex, STR, GGSEX and AMMA represents observed ozone v_d in (1) TexAQS06 (7 July–12 September 2006; north-western Gulf of Mexico), (2) STRATUS06 (9–27 October 2006; the persistent stratus cloud region off Chile in the eastern Pacific Ocean), (3) GasEx08 (29 February– 11 April 2008; the Southern Ocean), and (4) AMMA08 (27 April–18 May 2008; the southern and northern Atlantic Ocean). Because M11 v_d were much higher than other models, we exclude M11 in calculating the Ense for v_d . As shown in Fig. R6, Ense of v_d agreed with observations, reasonably. Both and simulated v_d showed a July-September maximum.

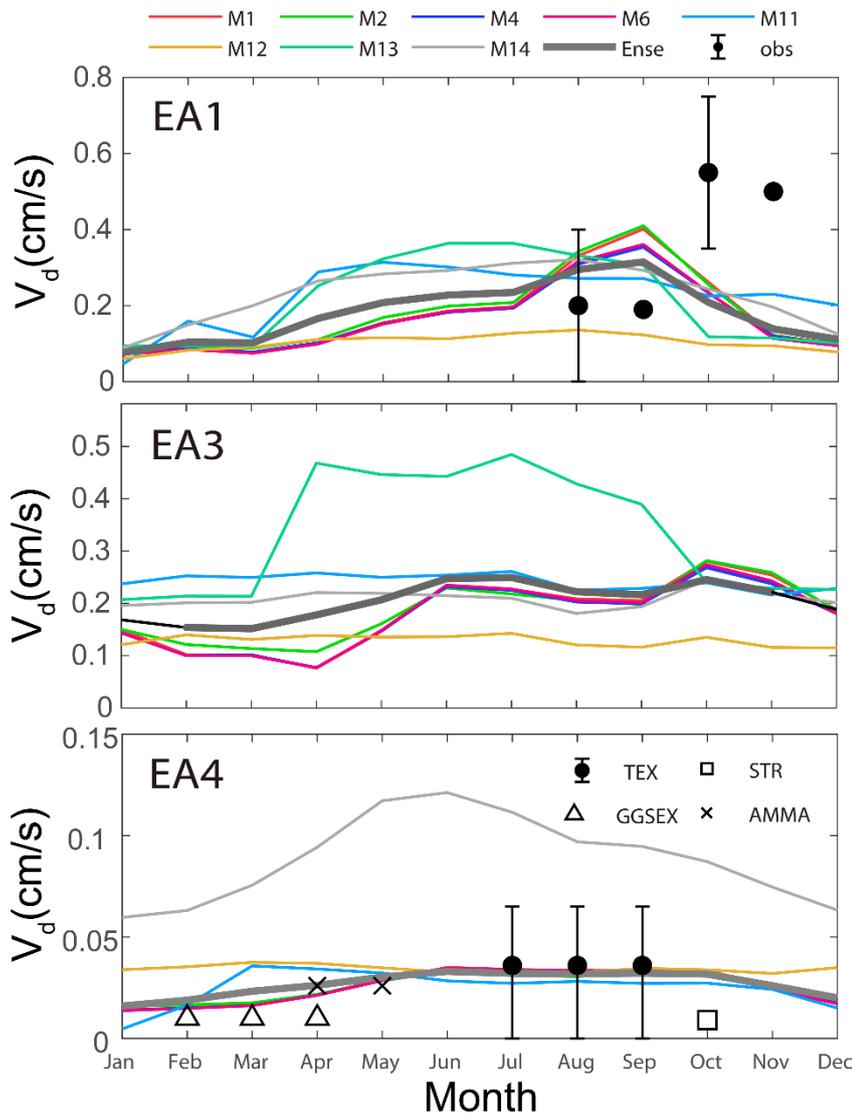


Fig. R6 simulated and observed monthly O₃ dry deposition velocities. Observations in EA1 were from Sorimachi et al. (2003) and Pan et al. (2010). Observations in EA4 were from Luhar et al. (2017).

Fig. R7 shows the comparison of simulated daytime PBL height with observations. In EA1, all the selected models exhibited the spring-maximum and winter-minimum season cycle, which captured the major pattern of climatology of PBLH observations (Guo et al.,2016). The Ense on PBLH was 100-200 m higher than radiosonde measurements. This is likely caused by the inconsistency of samples between models and measurements. The simulation was the mean value of 12 hours (08:00-20:00), while the average of measurements was calculated based on 3 hours (08:00, 14:00 and 20:00).

In EA3, observed PBLH did not varied as that in EA1, and differences between seasons were within 100 m. This pattern was captured by models. Similar as EA1, the simulated PBLH in EA3 was 100-200m higher than measurements.

Few measurements on remote oceanic site were conducted in East Asia. So, we compared simulations with European Centre for Medium-Range Weather Forecasts Reanalysis Data (von Engel et al., 2013). Both showed a winter-maximum pattern of PBLH.

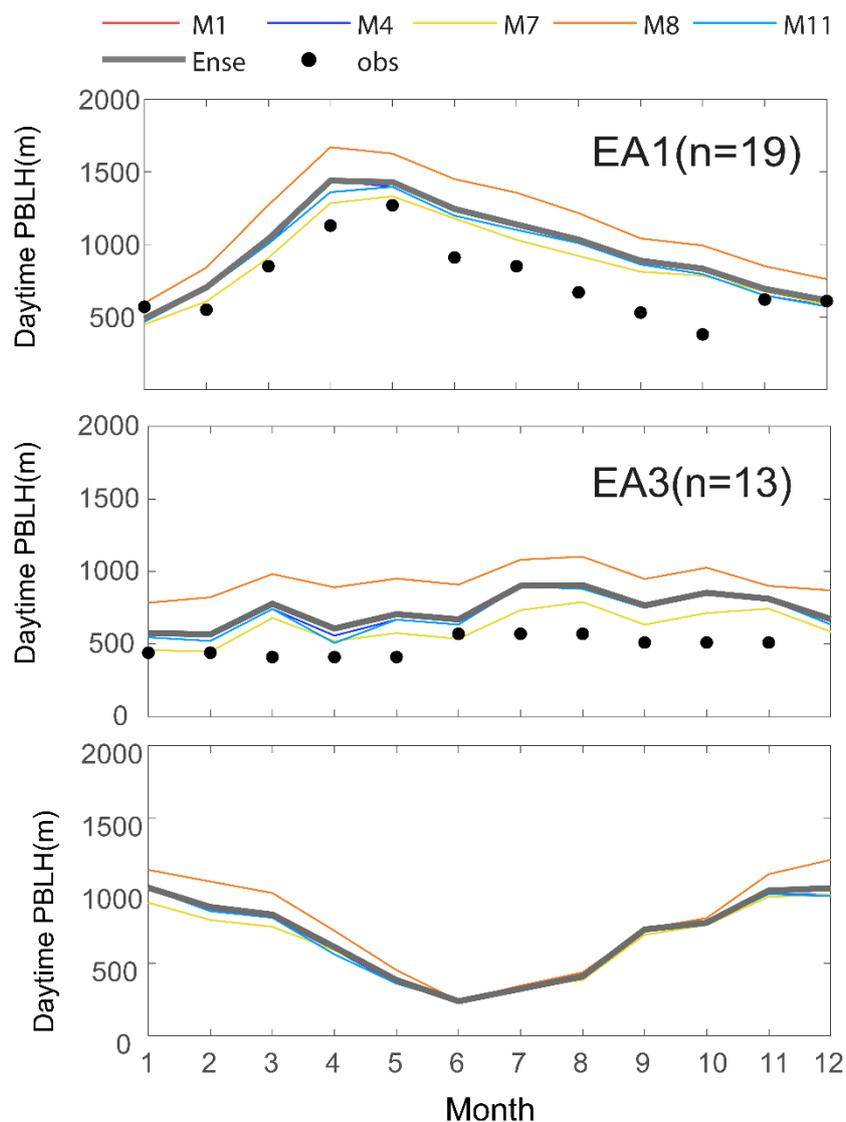


Fig. R7 Simulated daytime (08:00-20:00 LST) PBL height (m). Also shown are observed mean PBL height (m) at 08:00, 14:00 and 20:00 LST from Guo et al. (2016).

We totally agree with the reviewer that satellite observations evaluate NO_x or ozone at larger scales. Sometimes satellite data is lack in cloudy or heavy haze days. So, the monthly values of satellite could not be averages of all days. Unfortunately, only monthly data of models (all days in one month) was submitted in MICS-Asia III. This inconsistency of samples between models and satellite would bring bias for model validation. So, we will conduct the model validation using satellite data in MCIS-Asia IV by collecting daily data.

Changes in the revised manuscript: Page 13 Line 1-23; Page 16 Line 23-25; Page

17 Line 12-14.

Other comments etc...

Comment 6: Page 3-Line 7 – Please remind the value of the threshold

Reply: We added it ($100 \mu\text{g}/\text{m}^3$).

Changes in the revised manuscript: Page 3 Line 7.

Comment 7: Page 10 - Line 4 – Please suppress “4)”

Reply: We deleted it.

Changes in the revised manuscript: We deleted it.

Comment 8: Page 10 – Line 18 – A good example where using the ensemble average allows to better structure the discussion and to be more precise on the model skills.

Reply: We added a discussion on the using the ensemble average.

“In general, model results for three sub-regions exhibited a larger spread with a magnitude of 10-50 ppbv throughout the diurnal cycle than that in Europe and North America (Solazzo et al., 2012). The Ense O₃ in summer exhibited a systematic overestimation (20 ppbv) throughout the diurnal cycle in EA1. This indicated that models had difficulty dealing with O₃ in North China Plain. Compared with summer, there was only a slight systematic overestimation of Ense O₃ in other seasons (3-5 ppbv)”

Changes in the revised manuscript: Page 10 Line 20-27.

Comment 9: Page 10– Line 24-25 – “...due to difficulties in dealing with vertical mixing”: how do we know that?

Reply: In M11, the minimum of vertical diffusivity was set to be $0.5 \text{ m}^2 \text{ s}^{-1}$. This value is a little higher than other models (e.g. CAMx: $0.1 \text{ m}^2 \text{ s}^{-1}$). In the stable boundary layer on nighttime, the higher vertical diffusivity may transport high ozone in upper layer to the surface, and also uplifted surface NO. The lower NO weakens the ozone titration.

We realized that vertical mixing is not the only reason of nighttime ozone overestimation in M11. We needed more observed evidence to support our guess. So we deleted it in the revised manuscript.

Changes in the revised manuscript: We deleted it in the revised manuscript.

Comment 10: Page 12 – Line 16 – How statistics are calculated? on hourly values?

Reply: These statistics are calculated by Appendix A in the revised manuscript based on monthly values. We added descriptions in the revised manuscript.

Changes in the revised manuscript: Page 21 Appendix A. Statistical Measures

Comment 11: Page 13– Line 16 – Why choosing a sub selection of models? It would be interesting to have all models.

Reply: We agree. It’s better to present the intercomparison of PBLH from all models. Unfortunately, the other models have not outputted PBLH in this study. In MICS-Asia IV, all models will be requested to output PBLH.

Changes in the revised manuscript: None.

Comment 12: Page 14 – Line 3 – Von Engeln no ?

Reply: Yes, it is “von Engeln”.

Changes in the revised manuscript: Page 19 Line 14.

Comment 13: Page 14 – Line 7 – You do not discuss VOC emissions. Would you suggest that models have no sensitivity to these emissions?

Reply: We plotted VOCs (ethene) emissions (Fig. R8). Compared with NO, the consistency on ethene is better. Only M2 showed a small underestimation and overestimation in EA1 and EA3, respectively.

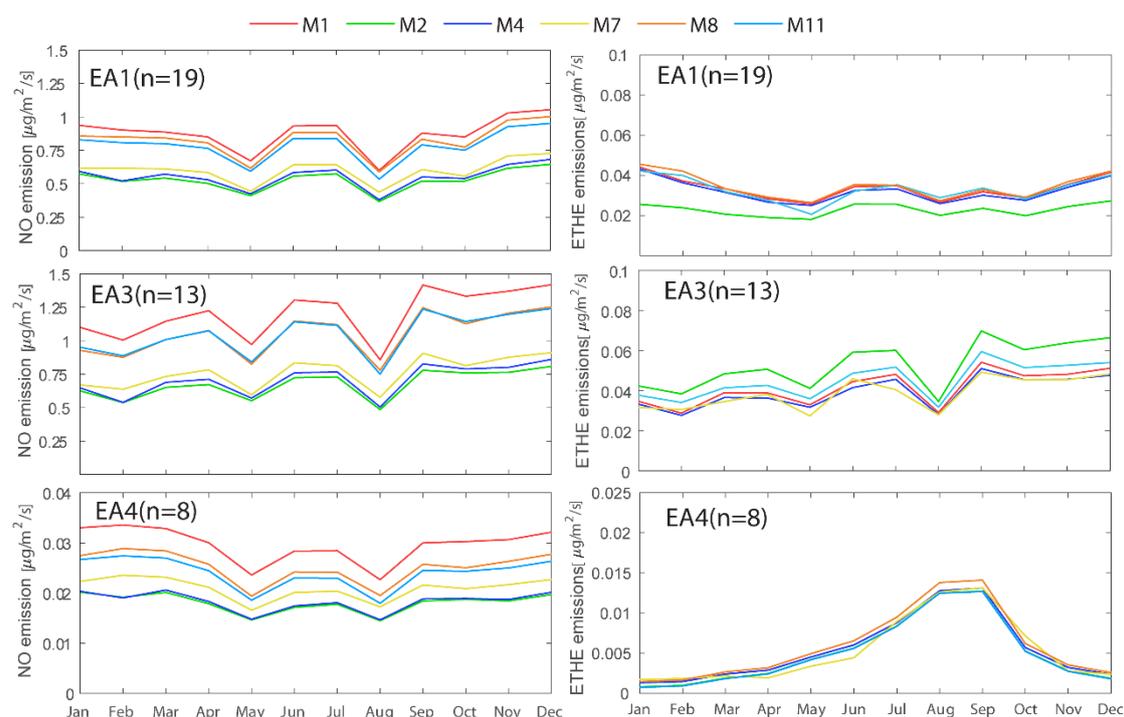


Fig.R8 NO (left) and ethene (right) emission fluxes on the first day in each month.

Comment 14: Page 14 – Line 15-20 – The discussion and the links between arguments are not that clear.

Reply: Thanks a lot.

“The difference in emissions allocations could contribute to the simulation variability. In the future, the projected gridded anthropogenic emissions should be provided to each group to eliminate the possibility that each group uses different mapping method. Interestingly, emissions in M1 and M8 exhibited similar levels, but their simulated NO₂, NO and O₃ presented a high intermodel variability in EA1 (Fig. 3 and Fig. 6). M1 simulated summer O₃ reached 80 ppbv while M8 was only 30 ppbv. This indicated that there were others causes to bring the intermodel variability on O₃.”

Changes in the revised manuscript: we delete related discussions on emissions because of the limitation of manuscript length.

Comment 15: Page 14 –Line 22 – I would say “net sink” since chemistry is a much higher absolute sink than deposition.

Reply: We agree.

Changes in the revised manuscript: Page 16 Line 14.

Comment 16: Page 16 – Line 4 to 6 – Seems contradictory to have a small sink with considerable effect on oceanic surface. I would rather say that even if dry deposition velocities are small over oceanic surfaces, the impact of dry deposition over ocean is globally important because of the large surface ocean are representing.

Reply: We agree. In the revised manuscript, we reworded this sentence. “Compared to other regions, surface O₃ in EA4 were more sensitive to dry deposition parameterization schemes in CTMs (Park et al.,2014). Park et al. (2014) revealed that O₃ on oceans differed by 5-15 ppbv in East Asia resulting from different dry deposition parameterization schemes”. We deleted “Ganzeveld et al. (2009) revealed that surface O₃ may differ by up to 60% when O₃ dry deposition velocity varied from 0.01 to 0.05 cm/s.”

Changes in the revised manuscript: Page 17 Line 14.

Comment 17: Page 16 – Line 6-8 – Why can we do the assumption that dry deposition is specifically important for EA4?

Reply: This assumption was taken from Park et al. (2014), in which the impact of O₃ dry deposition was examine over East Asia. They found that O₃ mixing ratios in EA4 were more sensitive to dry deposition parameterization schemes in CTMs than other regions. O₃ decrease as low as 5-15 ppbv at stations in EA4 in Wesely scheme than M3DRY scheme (1990). In EA1 and EA3, the changes of O₃ only ranged from 0-5 ppbv.

Changes in the revised manuscript: Page 17 Line 14.

Comment 18: Page 17 – Line 1 – I observe that range of concentrations for O₃ and

NO_x can be very different between models but it is not clear if slopes are that different.

Reply: We plotted the slopes between NO_x and O₃ in Fig. 8 in the revised manuscript. The slopes between NO_x and O₃ in EA1 ranged from -2.84 to -0.09 between models.

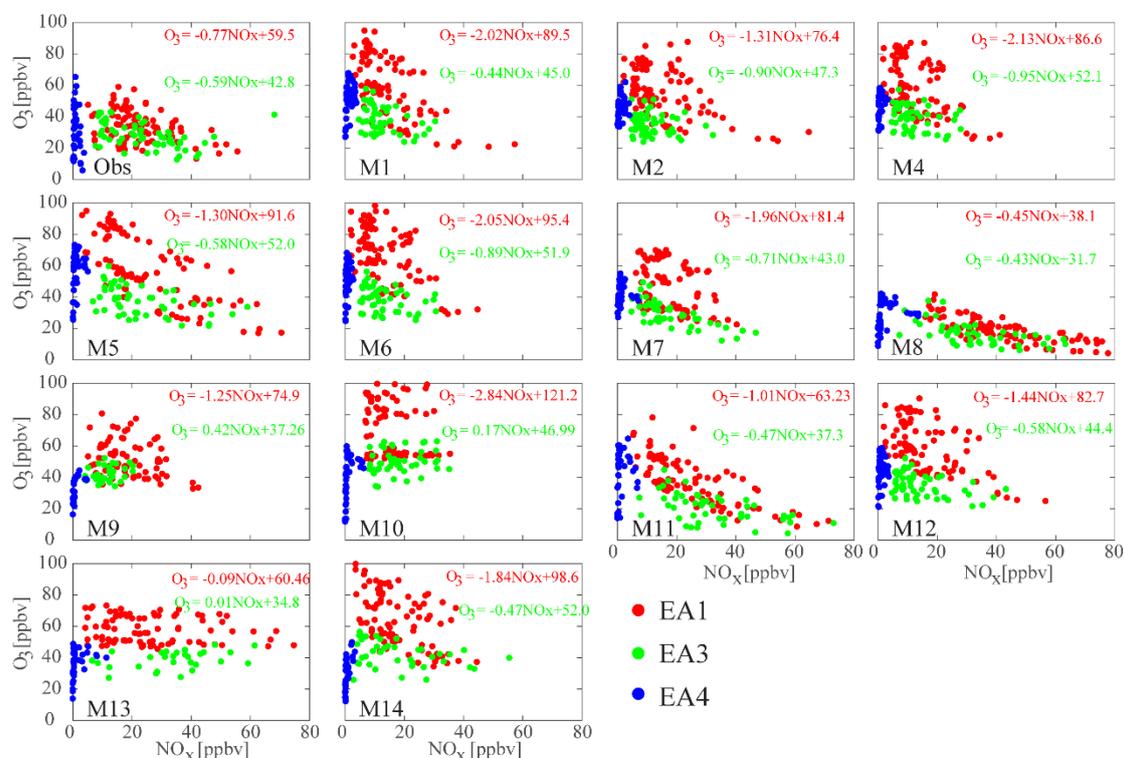


Fig. R9 Scatter plots between monthly daytime (08:00-20:00) surface NO_x and O₃ at each station over EA1 (red), EA3 (green) and EA4 (blue) in May-October, for observations (obs) and models

Changes in the revised manuscript: Page 18 Line 10-11.

Comment 19: Page 18 – Line 2 to 5 but also Line 7 to 20 – The variability authors are mentioning is not clear from figure 9. Also for differences between winter and summer, we need to have numbers to better evaluate this variability.

Reply: Thanks. Line 2-5: “A small variability in winter appeared below 900 hPa in three sub-regions, and slowly decreased with height. The mean standard deviation (σ) below 900 hpa were 7.6 ppbv, 6.9 ppbv and 6.0 ppbv in EA1, EA3 and EA4, which covered 18.3%, 15.0% and 15.4% of mean O₃ concentrations. In 700-900 hpa, σ decreased to 5.4 ppbv, 4.4 ppbv and 4.8 ppbv in EA1, EA3 and EA4, 12.2%, 9.4% and 10.8% of mean O₃ concentrations”.

Line 7-20: “With the increase of solar radiation and air temperature, vertical profiles were more scattered in the lower troposphere in summer. In polluted regions (EA1), various vertical structures of O_x were found below 700 hPa. σ reached 16.3 ppbv, 20.8 % of mean concentrations, which was higher than winter (6.2 ppbv, 15.2%). ... In EA3, vertical structures of O_x among models were consistent, but concentrations differed more than those in EA1. The mean standard deviation of models covered 22%

of mean concentrations”.

Table R3 Ensemble mean simulated ozone (Ense) and its standard deviation(std) in EA1

	Winter			Summer		
	Ense/ppbv	Std/ppbv	Std/Ense (%)	Ense/ppbv	Std/ppbv	Std/Ense(%)
1000-900 hpa	41.4	7.6	18.3	82.1	17.7	21.6
900-700 hpa	44.3	5.4	12.2	78.4	14.2	18.1
700-550 hpa	51.3	7.0	13.5	70.1	11.7	16.7
550-300 hpa	87.0	82.8	95.2	89.4	30.6	34.2

Changes in the revised manuscript: Page 13 Line 12-15, 17-18 and 23.

Comment 20: Page 18 – Line 5-6 – Authors do have this information, it should more than an suggestion, no?

Reply: Thanks a lot. This sentence is our guessed possible causes and we have not more evidences on the impact of convection and turbulent mixing on vertical profiles. So we deleted this sentence in the revised manuscript. In the MICS-Asia IV, we will directly output the impact of each process (convection, turbulent) from all models.

Changes in the revised manuscript: We deleted it.

Comment 21: Page 19 – Line 8 – 9 – Itis mention that dispersion between models is higher here than for the European case and authors suggest the models do not represent uncertainties, could you develop? Also authors mention that key processes could miss, what kind of processes are they thinking to?

Reply: Thanks a lot. We totally agree that an ensemble averages representing the uncertainty of O₃ is helpful. In MICS-ASIA III, the arithmetic means of all models is difficult meet this criteria, although it has been successfully in other regions. Potempski and Galmarini (2009) did some basic theoretical to find optimal linear combination of model results with the help of complex mathematical tools. Solazzo et al. (2012) used this method for O₃ ensemble in Europe and North America. They found that the most skillful ensemble is not necessarily generated by including all available models, and suggested that the clustering technique could generate a better ensemble average, but needs further refinement. This is beyond the scope of this manuscript and will be the major topic of our next manuscript

We mentioned that most models did not reflect this uncertainty or missed key processes in MICS-Asia III. The parameterization of heterogeneous chemistry in models is possibly a key process. The manuscript by Akimoto et al. (2019) in this special issue found that the missing heterogeneous “renoxification” reaction of HNO₃

on soot in most models except NAQPMS would partly explain the overestimation of simulated O₃ mixing ratios. The treatment of O₃ vertical transport in models also affect the simulated results significantly in Akimoto et al. (2019).

Changes in the revised manuscript: Page 15 Line 1-5.

Comment 22: Page 20 – Line 11 to 15 – Do we observe same differences for higher levels? Maybe in some models plumes are also present but at different altitudes.

Reply: We also compared simulated O₃ in upper boundary layer (Fig. R10). The results were similar as surface ozone.

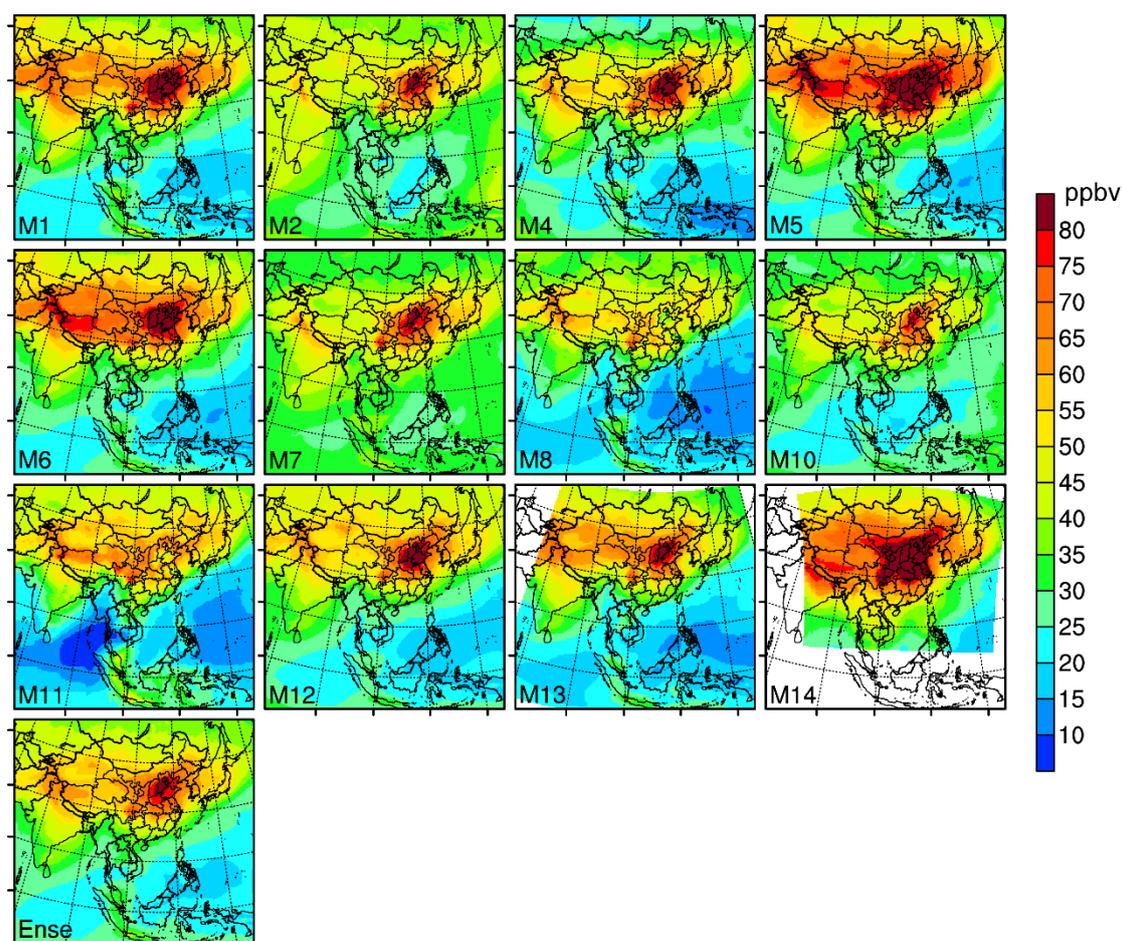


Fig. R10 500m O₃ spatial distribution from 13 models for summer 2010

Changes in the revised manuscript: Fig. 7.

Comment 23: Page21 – Line 2 – I'm not sure that author do define mathematically the coefficient of variation.

Reply: The CV is defined as the standard deviation of the modeled fields divided by the average. The larger the value of CV, The lower the consistency among the models.

Changes in the revised manuscript: Page 13 Line 27.

Comment 24: Page 21 – Line 13 – Like in table1 authors do mention that “default” is used as boundary conditions. Default values should be more clearly defined? climatology? from where?

Reply: In MICS-ASIA III, M2 and M7 made boundary conditions depending on their own previous experience denoted by "default" in Table 1.

In M2, the default initial condition and boundary conditions were based on Gipson (1999) to represent the clean air concentrations, and have been formulated from available measurements and results obtained from modeling studies.

In M7, the default initial condition and boundary conditions were derived from the idealized profile based upon northern hemispheric, mid-latitude, clean environment conditions from a NOAA-Aeronomy Laboratory Regional Oxidation Model (NALROM) (Liu et al.,1996).

Changes in the revised manuscript: Table 1.

Comment 25: Page 22 –Line 7 – “ .. its relevant species ..” I also see VOC or even radicals as relevant species for the tropospheric ozone cycle then it is better to mention O₃ and NO_x instead.

Reply: We agree and revised it.

Changes in the revised manuscript: Page 20 Line 4.

Comment 26: About Table and Figures Table2 – Maybe it is mandatory to mention how statistic alindicator are calculated (i.e formula). Be careful “suqare” in the title instead of square. RMSE do have units, please mention it. Figure 1 – as mention earlier I would have removed EA2 that is not discussed.

Reply: We agree. We listed the formula in the Appendix A in the revised manuscript. And also added RMSE units and corrected “suqare” to “square”. In the revised manuscript, we removed EA2.

Changes in the revised manuscript: Table 2 and Figure 2.

Comment 27: Figure 2 – probably too small as it is. The full blackline does not seems necessary.

Reply: We revised it.

Changes in the revised manuscript: Figure 2.

Comment 28: Figure 9 – Maybe it is possible to reduce horizontal scale down to 10 ppb to have more space on the right and to better evaluate the ensemble dispersion.

Reply: We revised it.

Changes in the revised manuscript: Figure 5.

Comment 29: Figure 10 – Maybe too small also Figure 11 – Same as Figure10

Reply: We revised it.

Changes in the revised manuscript: Figure 7.

References:

Akimoto, H., Nagashima, T., Li, J., Fu, J. S., Ji, D., Tan, J., and Wang, Z.: Comparison of surface ozone simulation among selected regional models in MICS-Asia III – effects of chemistry and vertical transport for the causes of difference, *Atmos. Chem. Phys.*, 19, 603-615, <https://doi.org/10.5194/acp-19-603-2019>, 2019.

Ban, S. , Matsuda, K. , Sato, K. , Ohizumi, T. : Long-term assessment of nitrogen deposition at remote EANET sites in japan. *Atmos. Environ*, 146, 70-78, 2016.

Ge, B., Sun, Y., Liu, Y., Dong, H., Ji, D., Jiang, Q., Li, J., and Wang, Z.: Nitrogen dioxide measurement by cavity attenuated phase shift spectroscopy (CAPS) and implications in ozone production efficiency and nitrate formation in Beijing, China, *J. Geophys. Res. Atmos.*, 118, doi:10.1002/jgrd.50757, 2013.

Gipson, G. L.: The Initial Concentration and Boundary Condition Processors. In Science algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modeling System, US Environmental Protection Agency Report, EPA-600/R-99/030, 12-1–12-91, 1999.

Guo, J., Miao, Y., Zhang, Y., Liu, H., Li, Z., Zhang, W., He, J., Lou, M., Yan, Y., Bian, L., and Zhai, P.: The climatology of planetary boundary layer height in China derived from radiosonde and reanalysis data, *Atmos. Chem. Phys.*, 16, 13309-13319, <https://doi.org/10.5194/acp-16-13309-2016>, 2016

Helmig, D., Lang, E. K., Bariteau, L., Boylan, P., Fairall, C. W., Ganzeveld, L., Hare, J. E., Hueber, J., and Pallandt, M.: Atmosphere-ocean ozone fluxes during the TexAQS 2006, STRATUS 2006, GOMECC 2007, GasEx 2008, and AMMA 2008 cruises, *J. Geophys. Res.*, 117, D04305, doi:10.1029/2011JD015955, 2012.

Li, J., Yang, W., Wang, Z., Chen, H., Hu, B., Li, J. J., Sun, Y., Fu, P., Zhang, Y.: Modeling study of surface ozone source-receptor relationships in East Asia. *Atmos. Res.*, S0169809515002227, 2015.

Liu, S. C., McKeen, S. A., Hsie, E-Y., Lin, X., Kelly, K. K., Bradshaw, J. D., Sandholm, S. T., Browell, E. V., Gregory, G. L., Sachse, G. W., Bandy, A. R., Thornton, D. C., Blake, D. R., Rowland, F. S., Newell, R., Heikes, B. G., Singh, H., and Talbot, R. W. : Model study of tropospheric trace species distributions during PEM-West A, *J.*

Geophys. Res., 101, 2073-2085,1996.

Zhong, L., Louie, P., Zheng, J., Wai, K. M., Josephine W.K. Ho, Yuan, Z., Lau, A. K. H., Yue, D., Zhou, Y.: The Pearl River Delta Regional Air Quality Monitoring Network – Regional Collaborative Efforts on Joint Air Quality Management, *Aerosol and Air Quality Research*, 13: 1582–1597, 2013

Pan, X., Wang Z., Wang X., Dong H., Xie, F., Guo, Y.: An observation study of ozone dry deposition over grassland in the suburban area of Beijing. *Chinese Journal of Atmospheric Sciences (in Chinese)*, 34(1), 120-130, 2010.

Park, R. J., et al. : An evaluation of ozone dry deposition simulations in East Asia. *Atmospheric Chemistry and Physics* 14(15): 7929-7940,2014.

Potempski, S., Galmarini, S.: Est Modus in Rebus: analytical properties of multi-model ensembles. *Atmospheric Chemistry and Physics* 9, 9471-9489,2009.

Solazzo, E., Bianconi, R., Pirovano, G., Matthias, V., Vautard, R., Appel, K. W., Bessagnet, B., Brandt, J., Christensen, J. H., Chemel, C., Coll, I., Ferreira, J., Forkel, R., Francis, X. V., Grell, G., Grossi, P., Hansen, A., Miranda, A. I., Moran, M. D., Nopmongco, U., Parnk, M., Sartelet, K. N., Schaap, M., D. Silver, J., Sokhi, R. S., Vira, J., Werhahn, J., Wolke, R., Yarwood, G., Zhang, J., Rao, S. T., Galmarin, S.: Model evaluation and ensemble modelling of surface-level ozone in Europe and north America in the context of AQMEII, *Atmos. Environ.*, 53(6), 60-74,2012.

Sorimachi, A, Sakamoto, K, Ishihara H, Fukuyama, T, Utiyama, M., Liu, H., Wang, W., Tang, D., Dong, X., Quan, H.: Measurements of sulfur dioxide and ozone dry deposition over short vegetation in northern China-A preliminary study. *Atmos. Environ.*, 37(22), 3157-3166, 2003.

Zheng, J., Zhong, L. , Wang, T. , Louie, P. K. K. , Li, Z.: Ground-level ozone in the pearl river delta region: analysis of data from a recently established regional air quality monitoring network. *Atmos. Environ.*, 44(6), 814-823,2010.

Model evaluation and inter-comparison of surface-level ozone and relevant species in East Asia in the context of MICS-Asia phase III Part I: overview

Jie Li^{1,2,3}, Tatsuya Nagashima⁴, Lei Kong^{1,2}, Baozhu Ge^{1,2,3}, Kazuyo Yamaji⁵, Joshua S. Fu⁶, Xuemei Wang⁷, Qi Fan⁸, Syuichi Itahashi⁹, Hyo-Jung Lee¹⁰, Cheol-Hee Kim¹⁰, Chuan-Yao Lin¹¹, Meigen Zhang^{1,2,3}, Zhining Tao¹², Mizuo Kajino^{13,14}, Hong Liao¹⁵, Meng Li¹⁶, Jung-Hun Woo¹⁰, Jun-ichi Kurokawa¹⁷, Qizhong Wu¹⁸, Hajime Akimoto⁴, Gregory R. Carmichael¹⁹ and Zifa Wang^{1,2,3}

¹LAPC, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China

²College of Earth Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China

10 ³Center for Excellence in Urban Atmospheric Environment, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen, 361021, China

⁴National Institute for Environmental Studies, Onogawa, Tsukuba, 305-8506, Japan

⁵Graduate School of Maritime Sciences, Kobe University, Kobe, 657-8501, Japan

15 ⁶Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN, 37996, USA

⁷Institute for Environment and Climate Research, Jinan University, Guangzhou, 510632, China

⁸School of Atmospheric Sciences, Sun Yat-sen University, Guangzhou, 510275, China

⁹Central Research Institute of Electric Power Industry, Tokyo, 100-8126, Japan

¹⁰Department of Atmospheric Sciences, Pusan National University, Pusan, 46241, South Korea

20 ¹¹Research Center for Environmental Changes/Academia Sinica, 11529, Taipei

¹²Universities Space Research Association, Columbia, MD, 21046, USA

¹³Meteorological Research Institute, Tsukuba, 305-8506, Japan

¹⁴Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, 305-8506, Japan

25 ~~¹⁵Jiangsu Key Laboratory of Atmospheric Environment Monitoring and Pollution Control, Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, School of Environmental Science and Engineering, Nanjing University of Information Science & Technology, Nanjing, 210044, China~~¹⁵Jiangsu Key Laboratory of Atmospheric Environment Monitoring and Pollution Control, Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, School of Environmental Science and Engineering, Nanjing University of Information
30 Science & Technology, Nanjing, 210044, China

¹⁶Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, 100084, China

35 ¹⁷Japan Environmental Sanitation Center, Asia Center for Air Pollution Research, Niigata, 950-2144, Japan

¹⁸Beijing Normal University, Beijing, 100875, China

¹⁹Center for Global and Regional Environmental Research, University of Iowa, Iowa City, IA, 52242, USA

Correspondence to: Jie Li (lijie8074@mail.iap.ac.cn)

Abstract: Long-term ozone (O_3) and nitrogen oxide (NO_x) from fourteen state-of-the-art chemical transport models (CTMs) are evaluated and intercompared to O_3 observations in East Asia, within the framework of the Model Inter-Comparison Study for Asia phase III (MICS-ASIA III), designed to evaluate the capabilities and uncertainties of current CTMs simulations in Asia and provide multi-model estimates of pollutant distributions. These models were run by fourteen independent groups working in China, Japan, South Korea, the United States and other countries/regions. Compared with MICS-Asia II, the evaluation against observations was extended to be one-full year in China and the western Pacific Rim from four months and the western Pacific Rim. ~~Potential causes of discrepancies between model and observation were investigated by assessing the planetary boundary layer heights, emission fluxes, dry deposition, O_3 - NO_x relationships and vertical profiles as determined by the models in this study.~~ In general, the model performance levels for O_3 varied widely, depending on region and seasons. Most models captured the key pattern of monthly and diurnal variation of surface O_3 and its precursors in North China Plain, ~~Yangtze River Delta~~ and western Pacific Rim, but failed in Pearl River Delta. A significant overestimation of surface O_3 was evident in May-September/October and January-May over the North China Plain, western Pacific Rim and Pearl River Delta. Comparison with observations revealed that underestimation on dry deposition velocities and large diversity of photochemical production partly contributed to this overestimation and large intermodel variability on O_3 in North China. In term of O_3 soundings, the ensemble average of models reproduced the vertical structure in western Pacific, but overestimated O_3 below 800 hpa in summer. In industrialized Pearl River Delta, the ensemble average presented an overestimation in the lower troposphere and underestimation in the middle troposphere. ~~A large intermodel variability of O_3 existed in all subregions over East Asia in this study, which was caused by the internal parameterizations of chemistry, dry deposition and vertical mixing of models, even though the native schemes in the models were similar.~~ The ensemble average of 13 models for O_3 did not always exhibit a superior performance compared to certain individual model, in contrast to its superiority in Europe. This suggested that the spread of ensemble-model values had not represented all uncertainties of O_3 or most models in MICS-~~ASIA-Asia~~ III missed key processes. ~~Compared with the previous phase of MICS-Asia (MICS-Asia II),~~ This study improved the performance of modeling O_3 in March at Japanese sites than the previous phase of MICS-Asia (MICS-Asia II). However, it overpredicted surface

O₃ concentrations in western Japan in July, which has not been found in MICS-Asia II. Major challenges still remain in regard to identifying the sources of bias in surface O₃ over East Asia in CTMs.

~~Potential causes of discrepancies between model and observation were investigated by assessing the planetary boundary layer heights, emission fluxes, dry deposition, O₃-NO_x relationships and vertical profiles as determined by the models in this study. A large intermodel variability of O₃ existed in all subregions over East Asia in this study, which was caused by the internal parameterizations of chemistry, dry deposition and vertical mixing of models, even though the native schemes in the models were similar.~~

1. Introduction:

Tropospheric ozone (O₃) is a significant secondary air pollutant produced through thousands of photochemical reactions and detrimental to human health, ecosystems, and climate change as a strong oxidant (WHO, 2005; The Royal Society, 2008). With the fast industrialization and urbanization in the last two decades, O₃ concentration is rising at a higher rate in East Asia than other regions and 30% of the days in megacities (e.g. Beijing, Shanghai Guangzhou in China) exceeds air quality standard of World Health Organization ([100 µg/m³](#)) ([Wang et al.,2017](#)) for 8-hour average surface O₃ concentration ([Wang et al.,2017](#)). The high O₃ concentrations received more attention from the public and from policy-makers in East Asia. The Ministry of Environment Japan has imposed stringent measures to reduce traffic emissions since 1990s, and non-methane volatile organic compounds (NMVOCs) and NO_x mixing ratios have decreased by 40-50 % and 51-54 %, respectively (Akimoto et al.,2015). In 2012, China released a new ambient air quality standard in which the 8-hour O₃ maximum was set limits for the first time. However, these measures don't prevent the persistent increase of the ground-level O₃ in East Asia. The average mixing ratio of O₃ increased 20-30% in Japan over the last 20 years (Akimoto et al.,2015). In Chinese megacities, 8-hr O₃ concentrations have increased 10-30 % since 2013 ([Wang et al.,2017](#)).

The primary method for detailed evaluation of the effect of air quality policies at the scale of East Asia is numerical air quality modeling. Several global and regional scale CTMs (e.g. GEOS-Chem, CHASER, CMAQ, CAMx, WRF-Chem and NAQPMS) over the past few decades have been developed and widely used to simulate the O₃ formation process and evaluate its control strategies (Streets et al.,

2008; Li et al., 2007; 2008; Yamaji et al., 2006; Zhang et al., 2008; Liu et al., 2010; Wang et al., 2013; He et al., 2017; Nagashima et al., 2017). These simulations have identified the key precursors of O₃ formation in East Asia (Zhang et al., 2008; Liu et al., 2010; Tang et al., 2011; He et al., 2017), assessed the contributions of international and regional transport (Streets et al., 2008; Li et al., 2008), and predicted the O₃ mixing ratios in different future emission scenarios (Wang et al., 2013). However, discrepancies remain between models and observations, indicating that model simulations of O₃ in East Asia still need to be improved (Han et al., 2008). Modeling uncertainties related to the emissions, chemistry, wet and dry deposition, and transport can hardly be handled using one single model. Model inter-comparison has thus been recognized as an effective way to address problems and has been successfully applied in Europe and North America in the phase 2 of the Air Quality Model Evaluation International Initiative (AQME II; Rao et al., 2011). Limited model inter-comparison related to air quality in East Asia has been conducted. Phases I and II of the Model Inter-Comparison Study for Asia (MICS-Asia) were initiated in 1998 and 2003, and to explore the potential sources of model uncertainties regarding sulfur, O₃, nitrogen compounds and aerosols (Carmichael et al., 2002, 2008). They found that the predicted temporal variations of surface O₃ in eight regional CTMs generally tended to be lower than that observed in 2001 with poor correlations in the western Pacific in March and December (Han et al., 2008). Model performance levels for O₃ varied largely in southern China. The inconsistency of horizontal grids, emissions and meteorological inputs among models increased the difficulty of explaining intermodel variability in the MICS-Asia II. More importantly, model evaluation in industrialized China has not been conducted because of few observations, which has been detrimental to efforts to improve model performance levels on O₃.

Recently, regional CTMs have been greatly improved by coupling more mechanisms (e.g. heterogeneous chemistry and on-line calculation of photolysis rates) and accurate chemical reaction rates. For example, the gas-phase chemistry mechanisms in Models 3-Community Multiscale Air Quality (CMAQ) have been developed into CBM05 and SAPRC07 from CB04 and SAPRC99. It is critical to evaluate the updated models' abilities for simulating current air quality over East Asia. In 2010, MICS-Asia was expanded to Phase III, in which 13 regional CTMs and 1 global CTM are run over one-full year by 14 independent groups from East Asia and North America, under a common reference model input

data set (namely, the emission inventory, meteorological fields and horizontal grids). In addition to observations made in Japan by the Acid Deposition Monitoring Network in East Asia (EANET) that were used in MICS-Asia II, new observational data from China were made available for MICS-Asia III, which were obtained from the Chinese Ecosystem Research Network (CERN) and the Pearl River Delta Regional Air Quality Monitoring Network (PRD RAQMN). An intercomparison of CTMs in China, Japan and western Pacific over one full year has never been performed, which provided a wider database to use in the comparisons. The completeness of MICS-Asia III is therefore unique.

In this paper, we mainly evaluate the abilities of participating models in MICS-Asia III for simulating the concentration of O₃ and its related species in the framework of MICS-Asia III. Several questions are addressed: (1) What is the performance level of various air quality models for simulating O₃ in East Asia? (2) How consistent or discrepant are the models? ~~(3) What are the potential factors responsible for differences and deviations between model results and observations?~~ (4) How do multi-model ensembles improve the simulation accuracy for O₃? This paper is expected to provide valuable insights into the abilities and limitations of CTMs in East Asia.

2. Models and data

2.1 Experimental set up

In this study, all participating models were run for the year 2010 and provide gridded monthly mean diurnal O₃ and its precursors mixing ratios in the lowest model layer. For O₃, monthly three-dimensional data were also submitted. If two or more observation sites were in the same grid of model, their mean values will be used to evaluate model performance.

~~Surface concentrations were interpolated to the monitoring locations for the model evaluation.~~

2.2 Participating models and input data

Table 1 summarizes the specifications of participating CTMs. These models include two versions of CMAQ (v4.7.1 and 5.0.2; Byun and Schere, 2006), the Weather Research and Forecasting model coupled with Chemistry (WRF-Chem; <http://www.acd.ucar.edu/wrf-chem>), Nested Air Quality Prediction Modeling System (NAQPMS; Li et al., 2007), the Japan Meteorological Agency (JMA)'s non-

hydrostatic meteorology-chemistry model (NHM-Chem; Kajino et al., 2012), the NASA-Unified Weather Research and Forecasting (NU-WRF; Tao et al., 2013) and GEOS-Chem (<http://acmg.seas.harvard.edu/geos/>). They have been documented in the scientific literature and widely applied in modeling studies over East Asia. Table 1 did not list model names to maintain model
5 anonymity for each participating model. Similar behavior was also found in MICS-Asia II and other model intercomparison projects (e.g. AQME II).

MICS-Asia III participants were provided with a reference meteorological field for the year 2010, generated with the Weather Research and Forecasting Model (WRF) version 3.4.1 model. The domain of meteorological fields is shown in (Fig. 1). WRF v3.4.1 are driven by the final analyses dataset
10 (ds083.2) from the National Centers for Environmental Prediction (NCEP), with $1^\circ \times 1^\circ$ resolution and a temporal resolution of 6 h. A four-dimensional data assimilation nudging toward the NCEP dataset was performed to increase the accuracy of WRF. The horizontal model domain, which is 182×172 grids on a Lambert conformal map projection with 45-km horizontal resolution, is shown in Fig. 1. Vertically, the WRF grid structure consists of 40 layers from the surface to the model top (10 hPa.). The standard
15 meteorological fields were applied by the majority of groups. Several other models performed simulations using their own meteorological models (e.g., RAMS-CMAQ and GEOS-Chem). The WRF-Chem utilized the same model (WRF) as the standard meteorological simulation, but they considered the feedback of pollutants to the meteorological fields. Consequently, their meteorological fields are possible slightly different from the standard. GEOS-Chem is driven by the GEOS-5 assimilated meteorological
20 fields from the Goddard Earth Observing System of the NASA Global Modeling Assimilation Office. The couples of meteorological and CTMs vary for each group, likely resulting in a diversified set of model output.

MICS-Asia III provided a set of monthly anthropogenic emission inventory for the year 2010, which is called as MIX (Li et al., 2016). MIX is a mosaic of up-to-date regional and national emission inventories
25 that include Regional Emission inventory in ASia (REAS) version 2.1 for the whole of Asia (Kurokawa et al., 2013), the Multi-resolution Emission Inventory for China (MEIC) developed by Tsinghua University, a high-resolution NH_3 emission inventory by Peking University (Huang et al., 2012), an Indian emission inventory developed by Argonne National Laboratory (ANL-India, Lu et al., 2011; Lu

and Streets, 2012), and the official Korean emission inventory from the Clean Air Policy Support System (CAPSS; Lee et al., 2011). The biogenic emissions are taken from the Model of Emissions of Gases and Aerosols from Nature (MEGAN). Hourly biogenic emissions were obtained for the entire year of 2010 using version 2.04 (Guenther et al., 2006). Biomass burning emissions were processed by re-gridding the
5 Global Fire Emissions Database version 3 (GFEDv3) (0.5 by 0.5 degree). Volcano SO₂ emissions were provided, with a daily temporal resolution by the Asia Center for Air Pollution Research (ACAP). The emission group in MICS-ASIA III directly prepared a gridded inventory according to the configuration of each CTM. NMVOC emissions are spectated into model-ready inputs for three chemical mechanisms: CBMZ, CB05 and SAPRC-99. Weekly and diurnal profiles were also provided. The standard emission
10 inventory was applied by all models. The majority of models employed official suggested vertical and time profiles of pollutants from each sector by the emission group. ~~Several other models M13 and M14~~ made the projection by themselves. More information can be found in the paper of Li et al. (2017) and Gao et al. (2017).

MICS-Asia III also provided two sets of chemical concentrations at the top and lateral boundaries
15 of the model domain, which were derived from the 3-hourly global model outputs for the year 2010. ~~(<http://acmg.seas.harvard.edu/geos/>; Sudo et al., 2002, respectively).~~ The global models were run by University of Tennessee (<http://acmg.seas.harvard.edu/geos/USA>) and Nagoya University ([Sudo et al., 2002](http://acmg.seas.harvard.edu/geos/) ~~Japan~~), ~~respectively~~ (~~<http://acmg.seas.harvard.edu/geos/>; Sudo et al., 2002, respectively~~). GEOS-Chem was run with a 2.5°×2° horizontal resolution and 47 vertical layers and Chemical AGCM for Study
20 of Atmospheric Environment and Radiative Forcing (CHASER) was run with a 2.8°× 2.8° horizontal resolution and 32 vertical layers. Some models made boundary conditions depending on their own previous experience.

2.3 Observational data for O₃

In this study, East Asia has been divided into ~~four~~ three sub-regions as shown in Fig. 1. The selection
25 of the sub-regions is based on emissions, climate and observation data coverage. The North China Plain (EA1); ~~Yangtze River Delta (EA2)~~, and Pearl River Delta (~~EA3~~EA2) represent the highly industrialized regions in the mid-latitudes. EA1 ~~and EA2~~ have a temperate and tropical continental monsoon climate with marked seasonality, respectively. ~~EA3~~EA2 is located in the south of China, and is less affected by

the continental air masses. EA4EA3 consists of the northwest Pacific and Sea of Japan, and represents the downwind regions of Asian continent with a marine climate.

Hourly O₃ and NO_x observations in the year 2010 in East Asia were obtained from CERN, PRD-RAQMN), and EANET. The CERN was built by the Institute of Atmospheric Physics, Chinese Academy of Sciences and consists of 24-19 surface stations within an area of 500 × 500 km² in North China Plain (EA1 sub-region; Ji et al., 2012). These stations were set up according to the United States Environmental Protection Agency method designations. Half of them were remote, rural or suburban and clear urban sites. 9 sites were located in the meteorological stations or campuses of universities in urban regions, with little influence from local sources and sinks. The comparison of NO emission rates at these sites in 45km and 3km resolution emission inventories showed that observation generally represented the ~45 km averages of pollutants. The PRD RAQMN was jointly established by the governments of the Guangdong Province and the Hong Kong Special Administrative Region and consists of 16 automatic air quality monitoring stations across the EA~~23~~ sub-region (Zhong et al., 2013). Thirteen of these stations are operated by the Environmental Monitoring Centers in Guangdong Province and the other three are located in Hong Kong and are managed by the Hong Kong Environmental Pollution Department. The PRD RAQMN was to probe the regional air quality, assess the effectiveness of emission reduction measures and enhance the roles of monitoring networks in characterizing regional air quality and supporting air quality management. So, sites are rarely influenced by local sources and sinks. The EANET was launched in 1998 to address acid deposition problems in East Asia, following the model of the Cooperative Program for Monitoring and Evaluation of the Long-Range Transmission of Air Pollutants in Europe. In this study, ~~seven~~ eight remote stations in the northwest Pacific and Japan (EA4EA3 sub-region) were selected for use in evaluating the model performance level in the downwind regions of the Asian continent (Ban et al., 2016). More information on the EANET can be found at <http://www.eanet.asia/>. Note that only stations with at least 75% data validity were chosen. Table S1 in the supplements lists detailed site description.

The O₃ and NO_x instruments were an ultraviolet photometric analyzer (model49i, Thermo Fisher Scientific (Thermo), USA) and a chemiluminescence analyzer (model42iTL, Thermo, USA), respectively. NO_x measurement existed sometimes biases (especially for stations far from sources) when

5 using molybden convertors devices since all nitrogen oxydes are measured. A one-month continuous measurement in August by a chemiluminescence analyzer and Aerodyne Cavity Attenuated Phase Shift Spectroscopy (CAPS) showed that this bias from a chemiluminescence analyzer was small when NO₂ concentrations were more than 10-15 ppbv, and ranged from 10% to 30% under low NO₂ (<10 ppbv) (Ge et al., 2013).

3. Model validation and general statistics

3.1 Annual concentrations of surface O₃, nitric oxide (NO) and nitrogen dioxide (NO₂)

10 Fig. 2 provides a concise comparison of model performance on annual O₃, NO and NO₂ in three sub regions in East Asia. A box-and-whisker representation was used to show the frequency distribution of monthly concentrations at stations in each sub-region. The O₃ normalized mean bias (NMB) and root mean square error (RMSE) of ensemble mean were significantly less than the ensemble median in most situations (Table 1). Therefore, we only presented the results of multi-model mean ensemble (Ense). ~~Note that because the Yangtze River Delta has only one station, a comparison of stations in Yangtze River Delta (EA2) is not shown in Fig. 2.~~ In general, the majority of models significantly overestimated annual surface O₃ compared with the observations in EA1, EA3~~2~~ and ~~EA4~~EA3. Ense overestimated surface O₃ by 10-30 parts per billion volume (ppbv) in these subregions. Ense NO₂ was close to the observations to within ±20% in all subregions. In EA1 and EA2, Ense NO was 5-10 ppbv lower than observation, while it showed a reasonable performance in EA3.

20 Among all models, M11 in sub-regions EA1 and EA3~~2~~, M7 in EA3~~2~~ and ~~EA4~~EA3 and M8 in all sub regions were closer O₃ observations. M11 simulated O₃ achieved a RMSE of 9.5 ppbv and 13.3 ppbv in EA1 and EA2, respectively (Table 2). were exceptions. M11 simulated O₃ in EA1 and EA3 agreed with observations, achieving a root mean square error (RMSE) of 9.5 parts per billion volume (ppbv) and 13.3 ppbv (Table 2). The simulation of O₃ in M7 was close to observations in EA3 and EA4. Interestingly, M8 underestimated O₃ in all regions, which was opposite to other models.

25 The performance levels of models for simulating O₃ were closely related to their performances for NO₂ and NO. In highly polluted regions (EA1 and ~~EA3~~EA2), a persistent underestimation of NO was evident across most models. ~~An exception was M8, which overestimated NO mixing ratios in all sub-~~

~~regions by 40-50%. This indicated M8 had the strongest O₃ titration that resulted in lower O₃ than other models and observations. An interesting phenomenon was that models' performance regarding O₃ varied greatly in EA3, although they M7 performed better at simulating O₃ than most other models did, although its performance at modeling NO was comparable to other models in EA1 and EA3. Therefore, the intercomparison of NO_x-O₃ chemistry between M7 and other models is needed in the next work. In EA4, all models but M8 showed a consistent performance with respect to NO and NO₂, although their performance regarding O₃ varied greatly. This suggests that O₃ was significantly affected by other factors in addition to local chemistry in EA4EA3. Interestingly, M8 underestimated O₃ and overestimated NO in all sub-regions by 40-50%, which was opposite to other models. This indicated M8 had the strongest O₃ titration in M8 may that result ed in lower O₃ than other models and observations. An exception was M8, which overestimated NO mixing ratios in all sub-regions by 40-50%.~~

3.2 Monthly variation of surface O₃, NO and NO₂

Fig. 3 presents the monthly mean concentrations of O₃, NO and NO₂ in ~~four~~ three sub-regions over East Asia. All models captured the observed seasonal cycles of O₃, NO and NO₂ in EA1. In May-September, Ense O₃ was 10-30 ppbv higher than observations, 30-70% of observed values, while Ense NO and NO₂ appeared to be consistent with observations, attaining mean biases of < 3 ppbv. This suggests that the intercomparison on O₃ production efficiency per NO_x with observations is needed. In EA2, Ense O₃ agreed well with observed high autumn O₃, but overestimated from January to September by 5-15 ppbv (15-60% of observations). This overestimation reached the highest in March-April (15ppbv) and led to a spring peak in simulated O₃ which was not found in observations. This overestimation was partly related to the underestimation of NO in the same months, which decreased the titration effect. For NO₂, Ense agreed well with observed values in June-December, and slightly underestimated observations in January-May. In EA3, the ensemble NO₂ was generally close to the observations to within ±0.5 ppbv. A significant overestimation of O₃ and underestimation of NO existed in June-October. Similar results have been found in MICS-Asia II and other model inter-comparison project under the Task Force on Hemispheric Transport of Air Pollution (TF HTAP), which suggested that such results may stem from the difference in the representation of dispersion by southwesterly clean marine air masses in different metrological fields used in CTMs (Han et al., 2008; Fiore et al., 2009).

~~For individual models, Overestimates of O₃ of 30-60 ppbv (out of total observed values of 20-40 ppbv) in May-September were found in most models except M11. In the same period (May-September), simulations of NO and NO₂ by these models appeared to be consistent with observations, attaining mean biases of < 10 ppbv. This suggests that the intercomparison on O₃ production efficiency per NO_x in these models is needed. The M11 achieved the best model reproductivity of monthly mean O₃ in EA1 among models. Other most models overestimated O₃ by 100-200% in May-October. The largest model bias and intermodel variability for NO and NO₂ appeared in winter, which. These model biases likely came from the NO_x surface emissions, dry deposition, vertical diffusion and heterogeneous chemistry (Akimoto et al., 2019), which will further be discussed in next section. In EA23, M7 seems to have achieved the best reproducibility for O₃. Most models (except M7, M118 and M124) exhibited high O₃ concentrations in March-May and September-November. Observed O₃ showed that the highest concentrations appeared in October-November a two-peak seasonal cycle but the observation exhibited a one-peak seasonal cycle. The O₃ concentration in January-May was significantly overestimated by 15-35 ppbv (out of observed values of 20-30 ppbv). In other months, O₃ was slightly overestimated by these models (-10 ppbv). The underestimation of NO titration strength partly explained the overestimation of O₃. Simulation results for NO fell in the range of 1-5 ppbv in most models, which was much less than observed concentration (10-25 ppbv). M11 captured the observed January-May O₃ because of relatively high NO concentrations. However, NO was overestimated by M11 in May-September, which led to the underestimation of O₃. M7 seems to have achieved the best reproducibility for O₃, but its simulated values of NO were only 10-30% of observations. In EA4EA3, spatially averaged O₃ concentrations often differ by more than 20 ppbv in the individual models. The highest intermodel variability on O₃ appeared in May-October, which overestimated O₃ in comparison to observations by 10-40 ppbv. Similar results have been found in MICS-Asia II and other model inter-comparison project under the Task Force on Hemispheric Transport of Air Pollution (TF-HTAP), which suggested that such results may stem from the difference in the representation of dispersion by southwesterly clean marine air masses in different metrological fields used in CTMs (Han et al., 2008; Fiore et al., 2009). In this study, however, most model employed the common reference meteorological fields. This indicated the representation of regional photochemical chemistry seems to be responsible for the model intermodel variability and overestimation~~

~~rather than the representation of southwesterly winds.~~ Interestingly, although M8, M9 and M14 exhibited a similar magnitude with observations in June-September, they significantly underestimated observations in other months by 200-300%. A detailed investigation is required in future studies.

~~In contrast to O₃, the simulated NO₂ results in the models exhibited a satisfactory consistency, and agreement with observations. 4) As shown in Fig.1, only 1 station exists in EA2. Thus, the model validation with observations was likely beset by large uncertainties in EA2. Most models reproduced the O₃ general seasonal cycle with May-October maximum-winter minimum (Figure not shown). In May-October, only M7 and M11 estimated O₃ concentrations at the same magnitude as measurements, and other models overestimated O₃ by 100-200%. In winter and spring (November-April), half participant models (M1, M4, M7, M11, M12 and M14) agreed well with observation well, and others overestimated observations by 50%-100%. As other sub-regions, M8 underestimated O₃ for the whole year in EA2 because it overestimated NO by 300%. Most models appeared to have difficulties in capturing the observed NO concentrations and exhibited large scatter effects in winter. Two exceptions were M5 and M11. M5 and M11 achieved satisfactory performances in summer and other seasons. For NO₂, most models (except M13 and M5) showed a good consistency with observations, and a lower bias than for O₃ and NO.~~

3.3 Diurnal concentrations of surface O₃

Sub-regional O₃ diurnal variations are shown in Fig.4. In general, model results for ~~four~~three sub-regions exhibited a larger spread with a magnitude of 10-50 ppbv throughout the diurnal cycle than that in Europe and North America (Solazzo et al., 2012). The Ense O₃ in summer exhibited a systematic overestimation (20 ppbv) throughout the diurnal cycle in EA1. This indicated that models had difficulty dealing with summer O₃ in East AsiaNorth China Plain. Compared with summer, there was only a slight systematic overestimation of Ense O₃ in other seasons (3-5 ppbv). In EA2, Ense O₃ generally agreed with the observations in summer, autumn and winter. In particular, the O₃ maximum around noon was reproduced, reasonably. There was only a 3-5 ppbv overestimation during 16:00-23:00 and early morning (6:00-10:00). In spring, a systematic overestimation of Ense O₃ existed in the whole diurnal cycle (5-10 ppbv). In EA3, Ense captured the small diurnal variation of O₃ in four seasons, but significantly

~~overestimated observations in summer and autumn (5-20 ppbv). In spring and winter, differences between Ense and observations were within 5 ppbv.~~

~~Among all models, In EA1, the deepest diurnal variation of observed O₃ appeared in summer. The majority of the models exhibited a consistent overestimation (20-60 ppbv) throughout the diurnal cycle to varying degrees, with the exception of one outlying model (M8), which systematically underestimated O₃ concentration. Among models, M11 exhibited the best model performance level on peak daily O₃ concentrations of 60 ppbv in 14:00-16:00 in EA1, but still overestimated. On nighttime O₃, M11 had a slight overestimation of by 10 ppbv, due to difficulties in dealing with vertical mixing.~~ Compared with summer, models' performances had a significant improvement in winter because of the weak intensity of photochemical reactions, except M2, M10 and M8. Differences between observations and most simulations in both nighttime and daytime were within 5 ppbv. The contrast of the models' performances between summer and winter implied that the variety of parametrizations on chemistry in different models partly explained the intermodel variability of simulated O₃ in EA1 (North China Plain).

In EA~~23~~ (Pearl River Delta, China), the majority of models agreed well with the diurnal variation in summer and autumn. ~~However, In other seasons,~~ most models had a tendency to overestimate the O₃ concentrations in both daytime and nighttime in spring. ~~In particular, t~~The overestimated magnitude exceeded 10 ppbv and 25 ppbv (out of observed values of 20-35 ppbv) in nighttime and daytime, respectively. M11 reproduced the observed O₃ in spring, but underestimated O₃ in summer and autumn.

In EA~~4~~EA3, ~~the all models captured the small diurnal variation of O₃ in four seasons. However,~~ significant intermodel variability still existed throughout the year. ~~As shown in Fig.4, t~~The amplitude of intermodel variability except M8 and M14 reached approximately 20 ppbv and approximately 10 ppbv in spring-summer and autumn-winter, respectively. ~~Compared with the observations, the majority of models except for that of M8 and M14 generally reproduced the magnitude of observed values in spring and winter. Observations lay in the middle of simulated values. In summer, the majority of models overestimated observed both daytime and nighttime O₃. As discussed in section 3.2, M8 and M14 exhibited the lowest O₃ among models in the whole year.~~

3.4 Error statistics on surface concentrations

In this section, we present statistics concerning the performance levels of the models based on monthly values. They are calculated by equations in Appendix A. On a yearly basis, all models showed the highest (0.8-0.9) and lowest (0.1-0.6) correlation coefficients for O₃ in EA1 and EA3₂, respectively (Table 2). The high correlations in EA1 were mainly because the summer-maximum and winter-minimum seasonal cycle is the typical pattern in polluted regions that were well represented in all the participating models. In general, Ense performed a better performance level than individual models for representing NO₂ in East Asia, reproducing the observed seasonal cycle and magnitudes. However, Ense did not always exhibited a superior performance for O₃ over certain individual model in East Asia, which was in contrast to its performance in Europe (Table 1). M7 and M11 agreed well with observations in EA1 and EA2, while ENSE tended to overestimate O₃ concentrations in May-September in EA1 and January-September in EA2. Loon et al. (2007) indicated that ENSE exhibited a superior performance level only when the spread of ensemble-model values was representative of the uncertainty of O₃. This indicated that most models did not reflect this uncertainty or missed key processes in MICS-Asia III.

The large overestimation of most models in May-September led to high ~~normalized mean bias~~ (NMB \neq 0.25-1.25) and RMSE (10-33 ppbv) in EA1. M11 had the lowest NMB (0.09) and RMSE (9.46 ppbv) among models. ~~So, the model intercomparison between M11 and other models is helpful for improving the model performance level in in EA1.~~ In EA2₃, M9 and M10 had larger correlations than the other models. However, their NMB and RMSE were also the highest. This implied that systematic model biases existed in these two models. ~~The positive bias in the majority of models except M7 and M11 was mainly caused by a large overestimation during the winter and spring seasons.~~ M7 exhibited a lower NMB and RMSE than other models, but its correlation was only 0.29. ~~M11 underestimated O₃ concentrations by 25%. Investigating differences of model parameterization between M7, M11 and others is a good way to improve the model performance level in this region.~~ In EA4EA3, the correlations exhibited the largest intermodel variability among all sub-regions, ranging from -0.13-0.65. M7 showed the lowest NMB and RMSE. This is likely caused by the cancelling effect of its overestimation in summer and underestimation in other seasons (Fig. 3).

For NO, correlations of models in EA1 ranged from 0.57-0.68, which indicated all models did a good job in reproducing the spatial variability of NO in this sub-region (Table 3). The NMBs indicated underestimation by models except M8 which mostly occurred in winter (Fig. 3). This underestimation partly was attributed to the coarse model horizontal resolution (45km) used in the MICS-Asia III, which hardly reproduced concentrations of short-lived species. Although most of the models employed the same emission inventory and meteorological field, EA1 still had a high model intermodel variability (scattered NMBs). This implied that the treatment of models on chemistry, vertical diffusion and dry deposition may have contributed to this underestimation of NO. In contrast to most models, M8 overestimated NO concentrations in all three sub-regions. It is noted that observations of NO were too low (<0.3 ppbv) in EA4EA3 to be discussed in this study.

Table 4 shows the statistics of models' performance levels for NO₂. In general, most models exhibited a better performance levels for representing NO₂ than O₃ and NO in EA1. The NMBs ranged from -0.28-0.32, which were much lower than O₃ (0.48-1.25). The correlations were 0.54-0.66, implying the reliable model performance levels for reproducing the spatial and month-to-month variability of NO₂ in EA1. Similar to O₃ and NO, the correlation coefficients of NO₂ in EA23 remained low. Thus, a dedicated investigation on O₃, NO and NO₂ in EA32 is urgent, but beyond the scope of this study. In EA4EA3, correlation coefficients ranged from 0.5-0.72. The NMBs and RMSEs except M8 ranged from -0.42-0.46 and 0.91-1.79 ppbv, respectively.

~~3.5 Last but not least, large intermodel variability for O₃, NO and NO₂ occurred in EA1, EA3 and EA4, higher than for Europe and North America despite, using the same meteorological fields and emissions data (Solazzo et al., 2012). This indicated that treatment of models' parameterizations on physical and chemical processes contains nonnegligible uncertainties in East Asia. We must thus investigate the possible causes and improve the model's performance levels for O₃ in East Asia. Vertical profiles of O₃~~

~~Fig. 5 the vertical profiles of observed and simulated O₃ in East Asia in summer and winter. Ensemble means (Ense) presented an underestimation and overestimation for EA2 O₃ in middle (500-800 hpa) and lower (below 900 hpa) troposphere, respectively. In winter, the underestimation was even extended to 200hpa. The magnitudes of underestimation and overestimation reached 10-40 ppbv and 10-~~

20 ppbv, respectively. In EA3, Ense reproduced the vertical structure of ozone in both summer and winter. An overestimation existed below 800 hpa in summer, with a magnitude of 10-20 ppbv.

A large intermodel variability of O₃ above 300 hPa is evident in all sub-regions, which is attributable to the various different top boundary conditions among models. However, this large variability was not transmitted to middle troposphere (400-600 hPa), in which O₃ concentrations were consistent among models. In the lower troposphere, a small intermodel variability in winter appeared below 900 hPa in three sub-regions, and slowly decreased with height. The mean standard deviations of models (σ) below 900 hPa were 7.6 ppbv, 6.9 ppbv and 6.0 ppbv in EA1, EA2 and EA3, which covered 18.3%, 15.0% and 15.4% of mean O₃ concentrations. In 700-900 hpa, σ decreased to 5.4 ppbv, 4.4 ppbv and 4.8 ppbv in EA1, EA2 and EA3, 12.2%, 9.4% and 10.8% of mean O₃ concentrations.

In the lower troposphere, the intermodel variability in summer were generally higher than those in winter. In polluted regions (EA1), σ reached 16.3 ppbv (20.8 % of mean concentrations) in summer, greatly exceeding those in winter (6.2 ppbv, 15.2%). Various vertical structures of O₃ were found below 700 hPa in summer. O₃ concentrations slowly increased with height in M8 and M11, but they mixed well in the PBL and decreased from 800 hPa to 700 hPa in the other models. Akimoto et al. (2019) found that the parameterization on downward O₃ transport from the upper boundary layer contributed a lot to the discrepancy between M1, M6 and M11. In EA2, vertical structures of O_x among models were consistent, but concentrations differed more than those in EA1. σ covered 22% of mean concentrations.

4. Investigation of intermodel variability on O₃

In MICS-Asia II, Han et al. (2008) briefly attributed the intermodel variability to the diversity of meteorological fields, emissions, boundary conditions, model treatment of chemistry, vertical diffusion and dry deposition. Because every model in MICS-Asia II employed their own input data, these potential reasons were not carefully examined. In MICS-Asia III, the postprocess to the common reference input data set maybe caused some discrepancies between models, because they have their own vertical structures. In addition, three models applied their own meteorological fields, which were different from the meteorology employed by other models. Thus, we compared the PBLH, emissions fluxes, dry deposition velocities and relationships between NO_x and O₃ in the sub-regions, as well as the vertical profiles of O₃ and its precursors among models.

4.1 Daytime PBLH

The evolution of the PBLH plays a major role for the O_3 and its precursors. In general, O_3 precursors are mostly constrained within the boundary layer (Quan et al., 2013). A better understanding of the evolution of PBL is essential for the interpretation of model biases and intermodel variability. Fig.5 presents the monthly variations of spatial mean daytime PBLH (08:00–18:00 LST) in M1, M4, M7, M8 and M11 at observed stations in EA1, EA3 and EA4. These models were selected because their simulations are largely scattered on O_3 and its precursors and covered the overall variability of all the models in this investigation. In EA1, all the selected models exhibited the spring maximum and winter minimum season cycle, which captured the major pattern of observations (Guo et al., 2016). In the climatology of PBL derived from the radiosonde by Guo et al. (2016), daytime PBLH in EA1 (North China Plain) ranged from 0.5 km in winter and 1.5 km in spring. The magnitudes of simulated PBLHs were also consistent with observations. Among models, the simulated PBLHs were very close between M1, M4, M7 and M11. The simulated PBLH by M8 was systematically higher than those by other models, but the positive bias was less than 100–150 m (<10–20%). In EA3, larger scatters of PBLH appeared than in EA1, which were almost exclusively caused by the difference between M8 and other models. The variability between M1, M4, M7 and M11 was only approximately 50 m. Compared with the climatological observations, the simulations remained at a similar magnitude with the radiosonde data (Guo et al., 2016). In EA4, all models exhibited the winter maximum pattern of PBLH, which was consistent with those derived from European Centre for Medium Range Weather Forecasts Reanalysis Data (Engeln et al., 2013). In particular, PBL in May–October, the season with the highest intermodel variability of O_3 , was quite consistent between models (<50 m). This consistency of models on PBLH in these sub-regions implied that PBL hardly explained the large intermodel variability and model biases in East Asia.

4.2 Emissions

In the Phase III of MICS-Asia project, the anthropogenic emission inventory in all models basically came from the monthly gridded MIX inventory at $0.25^\circ \times 0.25^\circ$ resolution (Li et al., 2016). The mapping of MIX onto the different model grids and different months could lead to some discrepancies between

models, which can cause an intermodel variability in concentrations of pollutants. Fig.6 presents the spatial averaged monthly NO emission fluxes of M1, M2, M4, M7, M8 and M11 at stations over each sub region. In general, two groups can be formed: one consisting of M1, M8 and M11, and the other consisting of M2, M4 and M7. NO emissions in the two groups were consistent in EA1, with magnitudes of around 0.8 $\mu\text{g}/\text{m}^2/\text{s}$ and 0.6 $\mu\text{g}/\text{m}^2/\text{s}$, respectively. Interestingly, the simulated NO_2 , NO and O_3 evenly presented a high intermodel variability in the same group. For example, the highest (M1) and lowest (M8) values of simulated summer O_3 in the first group were 80 ppbv and 30 ppbv, respectively, with the same vertical structure (Fig.3). Fig. 6 clearly indicates that the difference in emissions allocations contributed to the simulation variability. In the future, the projected gridded anthropogenic emissions should be provided to each group to eliminate the possibility that each group uses different mapping method.

4.3 Dry depositions

Previous studies revealed that dry deposition processes are the key sink of O_3 , accounting for about 25% of total removed from the troposphere (Lelieveld and Dentener, 2000). The uncertainty of dry deposition in CTMs is still high because many processes are heavily parameterized in models (Hardacre et al.,2015). In East Asia, the land cover is highly heterogeneous, which brings additional difficulties to the simulation of dry deposition. The surface cover class in EA1 is the most complex, and includes deciduous broad leaf forest, urban and cropland areas. EA3 and EA4 consist of urban, ocean and islands. In this study, the simulated dry deposition velocities of O_3 were compared. Simulated deposition velocities were calculated from Eq. (1):-

$$V_d = F/C \quad (1)$$

Where F and C represent the simulated dry deposition flux and surface O_3 concentrations, respectively. We determined the spatial mean dry deposition velocities at stations in each sub region.

Fig. 7 presents the monthly spatial mean dry deposition velocities of O_3 in eight models over EA1, EA3 and EA4. In EA1, O_3 dry deposition velocities in M1, M2, M4 and M6 presented a sharp increase from July to September, ranging from 0.2 cm/s to 0.4 cm/s. The peaks of dry deposition velocities in M11, M13 and M14 were broader and extended from April to September, with a constant magnitude of 0.3-0.35 cm/s. The seasonality in M12 was small and remained at 0.1 cm/s in the whole year. The lower dry deposition velocities of O_3 from M1, M2, M4 and M6 than that of M11 partly explained higher

summer surface O_3 from those simulations than that from M11. However, M13 and M14 still produced high O_3 concentrations in May–September although their dry deposition velocities were similar to that of M11 (Fig.3). This suggested that there were other factors besides dry deposition playing important roles in the overestimation of summer O_3 in the majority of models. The intermodel variability between models were expected. In MICS Asia III, M1, M2, M4 and M6 are the same model with different versions. Hence the dry deposition velocities were consistent among these models. M11, M12, M13 and M14 employed their own vertical structures or meteorological drivers, which partly contributed to differences in O_3 dry deposition velocities as compared with M1, M2, M4 and M6. Interestingly, deposition velocities simulated in M11, M13 and M14 were quite similar. As for M12, the unique dry deposition parameterization (Zhang et al., 2001) was believed to contribute to the intermodel variability.

In EA3, similar features with EA1 are found. M1, M2, M4 and M6 were quite consistent with each other, with a seasonal cycle of spring minimum. M11, M12 and M14 had no obvious seasonal variability, with a magnitude of 0.1–0.2 cm/s. The seasonal pattern in M13 was considerably different from the other models, exhibiting a maximum in April–September with higher dry deposition velocities (0.5 cm/s). The performance of the models for dry deposition velocities was not always consistent with O_3 concentrations. For example, O_3 concentrations in M13 still remained high levels under higher dry deposition velocities conditions.

In EA4, all but M12 simulated small dry deposition velocities of 0.02–0.04 cm/s. This was expected because stations in this region are mostly located in coastal areas and islands, and thus their results accord with the finding in Hardacre et al. (2015), who reported that the simulated O_3 dry deposition velocities in eighteen models in HTAP project were <0.1 cm/s over oceans. Dry deposition velocities have a considerable effect on concentrations of surface O_3 on oceans, although the effect in absolute terms is small. Ganzeveld et al. (2009) revealed that surface O_3 may differ by up to 60% when O_3 dry deposition velocity varied from 0.01 to 0.05 cm/s. The uncertainties on dry deposition in EA4 may contribute to the overestimation of surface O_3 in the majority of models, and thus more observations are needed over oceans.

4.4 Relationships between surface NO_x and O_3

In general, surface O_3 mainly comes from the photochemistry involving NO_x and VOCs in polluted regions. Theoretical and simulation results showed that O_3 production increased almost linearly with the NO_x increase under NO_x -sensitive conditions and remained relatively unchanged or even decreased in NO_x -saturated (often called “VOCs-limited”) conditions (Kirchner et al., 2001; Sillman and He et al., 2002; Tang et al., 2010). Recent observations found that regional O_3 in the North China (EA1) and Pearl River Delta (EA3) was changing from NO_x -limited to NO_x -saturated regions (Jin et al., 2015). Examining the O_3 - NO_x relationships is a good way of investigating sources of intermodel variability and model errors concerning on O_3 in East Asia. Fig. 8 presents the O_3 concentrations as a function of NO_x in May-September based on the monthly daytime (8:00-20:00) mean observed and simulated results at stations shown in Fig. 1.

In EA1 (North China Plain), observations clearly revealed that O_3 concentrations decreased with the increase in NO_x concentration. O_3 concentrations mostly remained high levels (40-60 ppbv) when NO_x was less than 20 ppbv. This implied that O_3 was under NO_x -saturated conditions in EA1 in May-September, which was consistent with Jin et al. (2015). The 13 models showed a high intermodel variability in relationships between O_3 and NO_x . Only M5, M7, M8 and M11 showed a negative slope between O_3 and NO_x . M7 and M11 were in relative agreement with observations, reasonably. M8 showed a systematic underestimation of observed O_3 in the all range of NO_x . By contrast, M5 systematically overestimated O_3 concentrations, which reached 80-100 ppbv under low NO_x conditions (10-20 ppbv). Relationships between O_3 and NO_x in M1, M2, M4, M6, M9, M10 and M14 were consistently scattered, and had no relevance to NO_x . Interestingly, M13 maintained a similar O_3 level at all NO_x levels, which was different from other models and previous theoretical results.

In EA3 (Pearl River Delta), M1, M2, M4 and M6 reproduced observed O_3 in low NO_x (< 30 ppbv) but failed to capture the low O_3 under high NO_x conditions (30-40 ppbv). This explained the overestimation of these models for O_3 in May-September. By contrast, M8 and M11 produced excessively high NO_x values, which resulted in their underestimation for O_3 . In M13 and M14, O_3 concentrations were nearly constant in all levels of NO_x . O_3 was positively correlated with NO_x in M9

and M11, which is in contrast to observations. This suggests that more attention is needed when policy-makers designate the O₃ regime (VOCs limited or NO_x limited regimes) using M9, M11, M13 and M14.

Stations in EA4 are mostly located over clean oceans or islands. NO_x concentrations were less than 3 ppbv, which indicated the local chemistry appeared to not be a key factors of O₃ formation. Thus, we did not discuss the simulated O₃-NO_x relationship further in this study.

High intermodel variability among 13 models in the O₃-NO_x relationship existed over polluted regions in the MICS-Asia III. In some cases, the O₃ regime among models was even contradictory. This suggests that more attention must be paid to the development of abatement strategies in East Asia.

4.5 Vertical profiles of O_x (O_x=NO₂+O₃)

Previous studies revealed that vertical mixing of O₃ and its precursors can influence the ground-level O₃ concentrations because of the turbulent mixing and different intensities of NO titration effects in the surface and residual layer (Zhang et al., 2009). Field campaigns showed that O_x was an ideal index to reflect the impact of physical transport, excluding the impact of local NO titration (Wang et al., 2006). Fig. 9 presents the vertical profiles of simulated O_x in East Asia in summer and winter. A large intermodel variability of O_x above 300 hPa is evident in all sub regions, which is attributable to the various different top boundary conditions among models. For example, the lower O₃ mixing ratios in 100-300 hPa in M2 came from its the default top conditions. As shown in Fig.9, this large variability was not transmitted to middle troposphere (400-600 hPa), in which O_x concentrations were consistent among models.

In the lower troposphere, differences among models in winter were generally less than those in summer. A small variability in winter appeared below 900 hPa in three sub regions, and slowly decreased with height. This was likely caused by the near-ground chemistry or long-range transport. One exception was M4, which significantly underestimated O_x from surface to 500 hPa in EA1 and EA4, compared with other models. This suggested that vertical convection and turbulent mixing in M4 were unique compared to other models.

With the increase of solar radiation and air temperature, vertical profiles were more scattered in the lower troposphere in summer. In polluted regions (EA1), various vertical structures of O_x were found below 700 hPa. O_x concentrations slowly increased with height in M8 and M11, but they mixed well in

the PBL and decreased from 800 hPa to 700 hPa in the other models. This discrepancy between M8, M11 and other models was presumably caused by a series of factors. One was associated with PBL schemes in models. Bank et al. (2016) pointed out that non-local and local schemes in models significantly affected the vertical structure of trace gases. Another could be related to the model performance levels for simulating O₃ photochemical production rates. As discussed in section 4.4, the majority of models produced more surface O₃ in the same NO_x concentration than M8 and M11, which resulted in the accumulation of high O₃ in PBL in these models. In EA3, vertical structures of O_x among models were consistent, but concentrations differed more than those in EA1. This is likely related to the treatments of convection and cloud activity among models. EA3 is located in subtropics, and frequent convective and cloud activity redistributed O_x on the vertical dimension by strong vertical transport and changing photolysis rates.

5.4. Multi-model ensemble O₃ and comparison with MICS-Asia II

5.1 Ensemble O₃ at stations

Studies have demonstrated that the ensemble model usually exhibits a superior performance on O₃ than any single model (Solazzo et al., 2012). Table 2-4 also presents the statistics of two multi-model ensembles (Mean and Median) on O₃, NO and NO₂ in EA1, EA3 and EA4. Clearly, the O₃ NMB and RMSE of ensemble mean were significantly less than the ensemble median in most situations, which indicated the ensemble mean presented a better performance level to represent the observed O₃. Therefore, we only presented the results of multi-model mean ensemble (ENSE). In general, ENSE performed a better performance level than individual models for representing NO₂ in East Asia, reproducing the observed seasonal cycle and magnitudes (Fig. 3). However, ENSE did not always exhibit a superior performance for O₃ over certain individual model in East Asia, which was in contrast to its performance in Europe (Fig. 3). M11 and M7 agreed well with observations in EA1 and EA3, while ENSE tended to overestimate O₃ concentrations in May-September in EA1 and January-September in EA3. Loon et al. (2007) indicated that ENSE exhibited a superior performance level only when the spread of ensemble-model values was representative of the uncertainty of O₃. This indicated that most models did not reflect this uncertainty or missed key processes in MICS-Asia III.

5.24.1 Spatial distribution of single model and multi-model ensemble O₃

Fig. 6 shows that the spatial distributions of MICS-Asia III ensemble mean surface O₃ (Ense) and the coefficient of variation (CV). The CV is defined as the standard deviation of the modeled O₃ divided by the average. The larger the value of CV, the lower the consistency among the models. In summer, ENSE predicted the elevated O₃ concentration belt in the middle-latitudes (30°-45°N). A region of O₃ in excess of 60 ppbv stretched across North China Plain and China East Sea, which was much higher than values in MICS-Asia II (45-50 ppbv) for the year of 2001 (Han et al.,2008). In other seasons, the O₃ distribution shows higher O₃ over ocean than in eastern China, reflecting the O₃ titration from high NO_x emissions. Due to the stratospheric injection, surface O₃ over Tibet plateau remained a high level in the whole year, ranging from 50 to 65 ppbv. The seasonal cycle of surface O₃ in Ense in MICS-Asia III agreed with that in MICS-Asia II, but O₃ levels in polluted regions were higher (Han et al., 2008).

The CV ranged from 0.1-0.6 in East Asia. The highest values were found in EA1 in winter. These high values in low-latitude western Pacific (10°S-15°N) and Indian Ocean were likely caused by the treatment of lateral boundaries in models. In MICS-Asia III, M7, M8 and M9 employed the default configurations of models, and the others employed outputs of GEOS-Chem/CHASER/MOZART-GOCART global model. Compared with MIC-Asia II, the CVs in Asian continent except winter remained a similar level in this study (0.1-0.3) (Carmichael et al.,2008).

Although all models similarly predicted the elevated summer O₃ concentration belt in the middle-latitudes (30°-45°N), Fig.10 presents the predicted spatial distribution of seasonal averaged surface O₃ concentrations in summer for individual models. All models similarly predicted the elevated O₃ concentration belt in the middle latitudes (30°-45°N). However, the magnitude of the enhanced O₃ were different among the models (Fig. 7). M5 predicted the highest O₃ concentration of 60-90 ppbv in the North China Plain (EA1) and its outflow pathways including Bohai Sea, East China Sea, Korea, Japan and the Sea of Japan (Locations are shown in Fig. S1 in the supplements) this belt, whereas M8 predicted the lowest 35-50 ppbv. Overhang of 30 ppbv contour lines extending into Northwest Pacific in the Asian continent outflow plume differed considerably among models. The plume of 30 ppbv or higher O₃ in M1-M6, M13 and M14 reached further south and east of Japan (135°E, 20°N), than M8, M10 and M11 (120°E, 30°N). In MICS-Asia II and HTAP, differences of frequency of marine air masses from the

western Pacific Ocean were thought to be possible cause of O₃ discrepancy over ocean among models because of different meteorological drivers (Han et al., 2008). In MICS-Asia III, the winds fields in models were similar because models the same or similar meteorological fields (Fig. S2 in the supplements). Hence, this inconsistency among models have resulted from the combined influence of a series of factors that included the diversity in condensed gas-chemical mechanism and heterogeneous chemistry. Li et al. (2015) found that the chemical production was the dominated controlling factor of O₃ along the outflow pathways near the North China Plain in summer, rather than lateral and top boundary conditions. Impact of aerosols on ozone in these regions were frequently reported in Olson et al. (1997) and Li et al. (2018), by altering photolysis rates and heterogeneous chemistry. The detailed comparison on parameterization of these processes in models are needed in future intermodel comparison project in Asia.

~~The models also consistently simulated low O₃ concentrations in low latitudes (0°-15°N), but varied by approximately 20 ppbv among models. M8, M10 and M11 showed the lowest value (10-15 ppbv), and M2, M7 and M9 showed the highest (30-45 ppbv). The rest are consistently in a middle range (15-25 ppbv). This discrepancy in low latitudes among models might have resulted from the diversity of boundary conditions and dry deposition velocities. The lateral boundaries in M7, M8 and M9 came from the default configurations of models, while the rest utilized the output from CHASER, GEOS-Chem, or MOZART-GOCART global models. The largest differences among models appeared in the North China Plain (EA1) and its outflow pathways including Bohai Sea, East China Sea, Korea, Japan and the Sea of Japan. Most models (M1, M2, M4, M5, M6, M10 and M14) predicted much higher surface O₃ levels (75-85 ppbv) in North China Plain (EA1), where observations of 40-50 ppbv were reported. SAPRC99 chemical mechanism used in these models except M14 partly contributed to the overestimation. Previous studies revealed that SAPRC-99 predicts higher concentrations than CB05 and CB4 used in other models (Luecken et al., 2008). O₃ concentrations decreased with the increase of outflow distances, and reached ~60 ppbv in Sea of Japan in these models. In M8 and M11, O₃ concentrations were lower (30-50 ppbv) in source regions than other models (EA1) and increased in the long range transport to Japan. This inconsistency among models have resulted from the combined influence of a series of factors that included the diversity in condensed gas-chemical mechanism and heterogeneous chemistry. Olson et al.~~

(1997) indicated that a significant difference could appear among models with respect to the concentrations of O_3 because of differences in simulated photolysis rates, specific chemical reaction rates, and various treatments of VOCs. Li et al. (2015) found that the chemical production was the dominated controlling factor of O_3 along the outflow pathways near the North China Plain in summer, rather than lateral and top boundary conditions. The heterogeneous chemistry largely reduced surface O_3 in polluted regions of China with high aerosol loadings (Li et al., 2018). Interestingly, overhang of 30 ppbv contour lines extending into Northwest Pacific in the Asian continent outflow plume differed considerably among models. The plume of 30 ppbv or higher O_3 in M1-M6, M13 and M14 reached further south and east of Japan ($135^\circ E$, $20^\circ N$), than M8, M10 and M11 ($120^\circ E$, $30^\circ N$). In MICS-ASIAII and HTAP, differences of frequency of marine air masses from the western Pacific Ocean were thought to be possible cause of O_3 discrepancy over ocean among models because of different meteorological drivers (Han et al., 2008). As discussed in section 4.1, models in this study employed the same or similar meteorological fields. This indicated that the chemistry during the long range transport of pollutants in continental outflows seems to be a key factor causing O_3 simulation discrepancies. In winter, the distribution patterns of O_3 were quite alike among models, with high concentrations over parts of western China, northeastern India and the western Pacific from the East China Sea to south of Japan (Fig. ~~S3 in the supplements~~). Considerably high consistency was found among models in winter compared with summer. All models predicted the low concentration in eastern China because of the titration effect of high NO_x concentrations. However, the magnitude of O_3 were different. M5 and M8 notably underpredicted O_3 (~~-10 ppbv) than other models (15-30 ppbv).~~). In spring and autumn (Fig. ~~S1-S4~~ and Fig. ~~S2-S5~~ in the supplements), O_3 concentrations were generally higher than in winter in the whole model domain because of the enhancement of solar radiation or stratosphere-troposphere exchanging fluxes of O_3 . A major feature consistently produced by all models was the enhancement of O_3 over southern Tibet, northeastern India and the western Pacific, which was generally similar to that in winter. The position of O_3 enhancement further north of Japan was comparable with winter.

~~The spatial distributions of MICS-Asia III ensemble mean surface O_3 (ENSE) and the coefficient of variation (CV) were presented in Fig.12. the major features in the four seasons discussed in the preceding paragraph are more clearly identified. The distribution of ENSE O_3 concentrations was much~~

smoother than any individual model, due to data averaging. In summer, a region of O₃ in excess of 60 ppbv stretched across North China Plain and China East Sea, which was much higher than values in MICS-Asia II (45-50 ppbv) for the year of 2001 (Han et al., 2008). In other seasons, the O₃ distribution shows higher O₃ over ocean than in eastern China, reflecting the O₃ titration from high NO_x emissions. Due to the stratospheric injection, surface O₃ over Tibet plateau remained a high level in the whole year, ranging from 50 to 65 ppbv. The seasonal cycle of surface O₃ in ENSE in MICS-ASIA-III agreed with that in MICS-Asia II, but O₃ levels in polluted regions were higher (Han et al., 2008).

The CV ranged from 0.1-0.6 in East Asia. The highest values were found in EA1 in winter. These high values in low latitude western Pacific (10°S-15°N) and Indian Ocean were likely caused by the treatment of lateral boundaries in models. In MICS-ASIA-III, M7, M8 and M9 employed the default configurations of models, and the others employed outputs of GEOS-Chem/CHASER/MOZART-GOCART global model. Compared with MICS-Asia II, the CVs in Asian continent except winter remained a similar level in this study (0.1-0.3) (Carmichael et al., 2008).

5.34.2 Comparison with MICS-Asia II

In MICS-Asia II, model evaluation on O₃ were conducted in only sites in the western Pacific. Fig. 843 presents the simulated and observed surface O₃ at these monitoring sites in the phase II and III of MICS-Asia project. Note that different models were employed in two phases. In general, most models captured the major distribution of O₃ at most sites in both MICS-Asia II and III. ENSE showed a good consistency in March and December of 2001 and 2010. The underestimation of O₃ in March at Japan sites (site 4: Sado-seki, site 5: Oki and site 6: Banryu) in Phase II was largely improved in Phase III. However, the surface O₃ at western Japan (site 4: Oki, site 5: Hedo and site 6: Banryu) were severely overestimated in July 2010 by 10-30 ppbv. This overestimation has not been found in Phase II, in which the difference with observations was approximately 5 ppbv. Rural sites in western Japan were located in the upwind regions of Japanese domestic emissions, and usually used to capture the impact of Asian continent outflows. The overestimated O₃ in North China Plain (EA1) in Phase III contributed a lot to the enhanced concentrations at sites of western Japanese sites in July 2010. This indicated that the transboundary transport from the Asian continent in MICS-Asia III was likely overestimated compared with that in MICS-Asia II.

5. Discussions

In MICS-Asia II, Han et al. (2008) guessed that the diversity of meteorological fields, dry deposition, PBL, model treatment of chemistry and other physical processes contributed to model biases with observations and the intermodel variability. Quantifying the contribution of these processes is one effective way to explain model biases by sensitivity. But this required a tremendous amount of computational cost for 14 models. A qualitative analysis on potential causes by comparison between models and observations on these processes is essential to narrow sensitivity simulating scenarios for next phase of MICS-Asia. In MICS-Asia III, common input data (emission and meteorology) provide a good chance for this qualitative analysis on model parameterizations. We evaluated the models on dry depositions, PBL and chemistry by collecting their observations (dry deposition velocity and PBLH) as much as possible. This work was not conducted in MICS-Asia II and is believed to be helpful for model developers to improve model performance in East Asia.

5.1 Dry depositions

Previous studies revealed that dry deposition processes are the key net sink of O₃, accounting for about 25% of total removed from the troposphere (Lelieveld and Dentener, 2000). The uncertainty of dry deposition in CTMs is still high because many processes are heavily parameterized in models (Hardacre et al., 2015). In this study, the simulated dry deposition velocities of O₃ were compared. Simulated deposition velocities were calculated from Eq. (1):

$$V_d = F/C \quad (1)$$

Where F and C represent the simulated dry deposition flux and surface O₃ concentrations, respectively. We determined the spatial mean dry deposition velocities at stations in each sub-region.

Fig. 9 presents the simulated and observed monthly spatial mean dry deposition velocities of O₃. In EA1, ensemble mean values underestimated observed dry deposition velocities of O₃ (v_d) in August-September, but still fell into the range of observed standard deviation. This underestimation could contribute to the overestimation of O₃ concentrations in summer discussed in section 3.2. The lower dry deposition velocities in May-July from M1, M2, M4 and M6 than that of M11 partly explained higher summer surface O₃ from those simulations than that from M11. However, M13 and M14 still produced

high O₃ concentrations in May-September although their dry deposition velocities were similar to that of M11(Fig. 3). This suggested that there were other factors besides dry deposition playing important roles in the overestimation of summer O₃ in EA1. In October-November, simulated v_d apparently overestimated observations by 30-50%.

5 In EA2, similar features with EA1 are found. M1, M2, M4 and M6 were quite consistent with each other, with a seasonal cycle of spring minimum. M11, M12 and M14 had no obvious seasonal variability, with a magnitude of 0.1-0.2 cm/s. The seasonal pattern in M13 was considerably different from the other models, exhibiting a maximum in April-September with higher dry deposition velocities (0.5 cm/s). The performance of the models for dry deposition velocities was not always consistent with O₃ concentrations.
10 For example, O₃ concentrations in M13 still remained high levels under higher dry deposition velocities conditions.

In EA3, most stations were remote oceanic sites, and few dry deposition observations were conducted. So, we collected observations in other oceanic sites to evaluate model performance (Helmig et al., 2012). Ense of v_d agreed with observations reasonably (Fig. 9). Both observations and simulated
15 v_d showed a July-September maximum with a magnitude of 0.02-0.03 cm/s. Park et al. (2014) revealed that surface O₃ in EA3 were more sensitive to dry deposition parameterization schemes in CTMs. O₃ on oceans differed by 5-15 ppbv in East Asia resulting from different dry deposition parameterization schemes. Thus, more observations are needed over oceans in EA3 to decrease the uncertainties on O₃ simulations.

20 **5.2 Relationships between surface NO_x and O₃**

In general, surface O₃ mainly comes from the photochemistry involving NO_x and VOCs in polluted regions. Theoretical and simulation results showed that O₃ production increased almost linearly with the NO_x increase under NO_x-sensitive conditions and remained relatively unchanged or even decreased in NO_x saturated (often called “VOCs-limited”) conditions (Kirchner et al.,2001; Sillman and He et al.,
25 2002; Tang et al., 2010). Recent observations found that regional O₃ in the North China (EA1) and Pearl River Delta (EA2) was changing from NO_x-limited to NO_x-saturated regions (Jin et al., 2015). Examining the O₃-NO_x relationships is a good way of investigating sources of intermodel variability and model

errors concerning on O₃ chemistry in East Asia. Fig. 10 presents the O₃ concentrations as a function of NO_x in May-September based on the monthly daytime (8:00-20:00) mean observed and simulated results at stations shown in Fig. 1.

In EA1 (North China Plain), observations clearly revealed that O₃ concentrations decreased with the increase in NO_x concentration. O₃ concentrations mostly remained high levels (40-60 ppbv) when NO_x was less than 20 ppbv. This implied that O₃ was under NO_x-saturated conditions in EA1 in May-September. The slope and intercept of regression line between O₃ and NO_x were -0.77 ppbv/ppbv and 59.5 ppbv, respectively. Among models, M11 were in relative agreement with observations, reasonably. The slope and intercept (-1.01 ppbv/ppbv, 63.23 ppbv) were close to observations. Other models showed a higher model bias and intermodel variability on relationships between O₃ and NO_x. Their slopes mostly ranged from -1.25 ppbv/ppbv to -2.13 ppbv/ppbv, 1.3-2.8 times of observed slope. Their intercepts were 74.9 -121.2 ppbv, much higher than observation (59.5 ppbv). Akimoto et al. (2019) calculated the net photochemical production of M1, M6 and M11, and found that weak net chemical production in M11 were mostly responsible for low O₃ than M1 and M6. This is consistent with the low slope in M11. Interestingly, M13 maintained a similar O₃ level at all NO_x levels (Slope: -0.09), which was different from other models and previous theoretical results.

In EA2, M1, M2, M4 and M6 reproduced observed O₃ in low NO_x (< 30 ppbv) but failed to capture the low O₃ under high NO_x conditions (30~40 ppbv). This explained the overestimation of these models for O₃ in May-September. By contrast, M8 and M11 produced excessively high NO_x values, which resulted in their underestimation for O₃. In M13 and M14, O₃ concentrations were nearly constant in all levels of NO_x. O₃ was positively correlated with NO_x in M9 and M10, which is in contrast to observations. This suggests that more attention is needed when policy-makers designate the O₃ regime (VOCs-limited or NO_x-limited regimes) using M9, M10, M13 and M14.

Stations in EA3 are mostly located over clean oceans or islands. NO_x concentrations were less than 3 ppbv, which indicated the local chemistry appeared to not be a key factor of O₃ formation. Thus, we did not discuss the simulated O₃-NO_x relationship further in this study.

5.3 Other factors

Previous studies revealed that O₃ precursors are mostly constrained within the boundary layer (Quan et al., 2013). The model evaluation on PBLH is essential for the interpretation of model biases with observations. Unfortunately, this evaluation was not conducted in MICS-Asia II. In MICS-Asia III, all selected models exhibited the spring-maximum and winter-minimum season cycle in EA1 (Fig. S6 in the supplements), which captured the major pattern of climatology of PBLH observations (Guo et al., 2016). The Ense on PBLH only overestimated radiosonde measurements by 100-200 m (~10-15%). This is likely caused by the inconsistency of samples between models and measurements. The simulation was the mean value of 12 hours (08:00-20:00), while the average of measurements was calculated based on 3 hours (08:00, 14:00 and 20:00). In EA2, observed PBLH did not varied as that in EA1, and differences between seasons were within 100 m. This pattern was captured by models. Similar as EA1, the simulated PBLH in EA2 was 100-200m higher than measurements. Few measurements on remote oceanic site were conducted in East Asia. So, we compared simulations with European Centre for Medium-Range Weather Forecasts Reanalysis Data (von Engel et al., 2013). Both showed a winter-maximum pattern of PBLH.

The East Asia monsoon played an important role in seasonal cycle of O₃ in subregions by the long-range transport. Besides local intensive photochemical productions, the O₃ summer maxima in EA1 were also affected by regional transport from Yangtze River Delta under prevailed summer southern monsoon (~20%) (Li et al., 2016). In EA2, a late maximum of O₃ in September-November was quite different from EA1 and EA3. This is largely attributed to the long-range transport of O₃ and its precursors in the polluted continental air masses from northern China and photochemical formation under dry and sunny weather conditions in autumn (Zheng et al., 2010). In EA3, the seasonal change of O₃ concentrations was characterized by two peaks in spring and autumn. The first and second peak in Mar-Apr and May and June were mainly influenced by the inflow from outside of East Asia and chemically produced O₃ by regional emissions, respectively. In the next studies, we will conduct the intermodel comparison on transport fluxes of O₃ between sub-regions over East Asia.

6. Summary

In the MICS-Asia III framework, the evaluation and intercomparison of 13 CTMs were conducted with a wide variety of observations covering ~~three~~two Chinese industrialized regions and western Pacific, using long-term simulations for the year 2010. This study has focused on ~~surface-O₃ and, NO and NO₂ its relevant species~~. In particular, surface O₃ in China was evaluated, which was absent in the previous model-intercomparison projects. Large intermodel variability of O₃ existed in all subregions over East Asia in this study, with model concentrations varying by a factor of 2 to 3 between different models.

A model ensemble was conducted and evaluated. In general, the model ensemble ~~Causes responsible for discrepancy of models with observations and intermodel variability were investigated. Finally, a model ensemble was conducted and evaluated. Most models~~ captured the key pattern of monthly and diurnal O₃ ~~and its precursors (NO and NO₂)~~ in the North China Plain, ~~the Yangtze River Delta~~ and the western Pacific Rim. ~~However, the majority of models~~ It failed to capture the observed ~~single peak (autumn maximum)~~ seasonal cycle of O₃ in Pearl River Delta of China. In North China Plain and western Pacific rim. The model ensemble, which exhibited a two peak seasonal cycle.

~~Considerable difference between simulated and observed O₃ concentrations were found in all four subregions in East Asia. In North China Plain, the majority of models severely overestimated surface O₃ in May-September by 120-3040 ppbv. The only exception was M8, which underestimated surface O₃ by 10-15 ppbv.~~ This overestimation systematically appeared in both daytime and nighttime. Similarly, ~~most model~~ the model ensemble had a predominate tendency to overestimate the daytime and nighttime O₃ concentrations in ~~January Mayspring and May October~~ in Peral River Delta ~~(EA3) and western Pacific rim (EA4), respectively. Compared to MICS-Asia II, MICS-Asia III was less prone to underestimation of surface O₃ in March at Japanese sites. However, it predicted too enhanced surface O₃ concentrations at western Japan in July, which was not the case in MICS-Asia II. In term of O₃ soundings, the ensemble model in this study reproduced the vertical structure in western Pacific, but overestimated O₃ below 800 hpa in summer. In industrialized Pearl River Delta, the ensemble average presented an overestimation for O₃ in the lower troposphere and underestimation in the middle troposphere. This study revealed that ensemble average of 13 models on O₃ (ENSE) did not always exhibit a superior performance~~

to certain individual models in East Asia, which contrasted with its performance in Europe. This suggested that the spread of ensemble-model values had not represented all uncertainties of O₃ or most models in MICS-Asia III missed key processes. Unlike the performance level for O₃, ENSE demonstrated superior performance level than individual models for NO₂ in East Asia.

5 ~~The monthly O₃ series also revealed that some models performed better than others in some subregions (for example, M11 in EA1), but this behavior was not uniform in time and space. For NO₂ and NO, models appeared to be more consistent with observations than O₃.~~

~~Large intermodel variability of O₃ existed in all subregions over East Asia in this study, with model concentrations varying by a factor of 2 to 3 between different models. MICS-Asia II presented-guessed~~
10 ~~some potential reasons of variabilities among models, but did not explicitly examine the impact of these reasons. Quantifying the contribution of these processes to O₃ concentrations is one effective way to explain model biases by sensitivity simulations. But this required a tremendous amount of computational cost for 14 models. In this study, we directly-conducted a qualitative analysis on potential causes by comparison between models and observations on these processes to narrow sensitivity simulating~~
15 ~~scenarios for next phase of MICS-Asia. The comparison revealed that the ensemble model underestimated observed dry deposition velocities of O₃ in August-September in North China Plain, which could contribute to the overestimation of O₃ concentrations in summer. In western Pacific, simulated v_d agreed with observations reasonably. Photochemical treatment in models may contributed to the O₃ overestimation in North China Plain. Models captured the major pattern of climatology of~~
20 ~~PBLH observations in three subregions over East Asia. More evaluation on turbulent kinetic energy in PBL is urgent for assess the vertical mixing in future studies.~~

~~investigated the diversity of PBLH, emissions, dry deposition, O₃-NO_x relationships and vertical profiles among models. This investigation revealed that the internal chemical parameterizations of models (gaseous and heterogeneous chemistry) heavily contributed to the large variability among models,~~
25 ~~even though the native schemes in models were similar. Dry deposition and vertical mixing also played important roles.~~

~~This study revealed that ensemble average of 13 models on O₃ (ENSE) did not always exhibit a superior performance to certain individual models in East Asia, which contrasted with its performance~~

~~in Europe. This suggested that the spread of ensemble model values had not represented all uncertainties of O₃ or most models in MICS-Asia III missed key processes. Unlike the performance level for O₃, ENSE demonstrated superior performance level than individual models for NO₂ in East Asia.~~

~~Compared to MICS-Asia II, MICS-Asia III was less prone to underestimation of O₃ in March at Japanese sites. However, it predicted too enhanced surface O₃ concentrations at western Japan in July because of its overestimation in the North China Plain, which was not the case in MICS-Asia II. This indicated that the transboundary transport from Asian continent was likely overestimated in MICS-Asia III.~~

Author contribution:

JL, ZW and GC conducted the study design. JL, TN, BG, KY, JF, XW, QF, SI, HL, CK, CL, MZ, ZT, MK, HL contributed to modeling data. ML, JW, JK and QW provided the emission data. LK helped with data processing. HA, GC and ZW were involved in the scientific interpretation and discussion. JL prepared the manuscript with contributions from all co-authors.

Competing interests:

The authors declare that they have no conflict of interest.

Acknowledgements:

This work was supported by the Natural Science Foundation of China (41620104008, 41571130034; 91544227; 91744203), and National Key R&D Program of China (2017YFC0212402). This work was partly supported by the Environment Research and Technology Development Fund (S-12) of the Environmental Restoration and Conservation Agency of Japan and the Ministry of Environment, Japan. We thank the Pearl River Delta Regional Air Quality Monitoring Network for observations in Pearl River Delta. Dr. Kengo Sudo from Nagoya university and Prof. Rokjin J. Park provided us CHASER and GEOS-Chem outputs for boundary conditions. This manuscript was edited by Wallace Academic Editing.

Appendix A. Statistical Measures

Defining y_i and Obs_i modeled and observed concentrations of air pollutants at the i^{th} station, having mean value \bar{y} and \bar{obs}

Correlation coefficient (R)

$$R = \frac{\sum_{i=1}^n (y_i - \bar{y})(obs_i - \bar{obs})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (obs_i - \bar{obs})^2}} \quad (A1)$$

Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - obs_i)^2}{n}} \quad (A2)$$

Normalized Mean Bias (NMB)

$$5 \quad NMB = \frac{\sum_{i=1}^n (y_i - obs_i)}{n \times \bar{y} \times \bar{obs}} \quad (A3)$$

References:

Ackermann, I. J., Hass, H., Memmesheimer, M., Ebel, A., Binkowski, F.S., and Shankar, U.: Modal aerosol dynamics model for Europe: Development and first applications, *Atmos. Environ.*, 32, No.17, 2981-2999,1998.

10

Akimoto, H., Mori, Y., Sasaki, K., Nakanishi, H., Ohizumi, T., Itano, Y.: Analysis of monitoring data of ground-level ozone in japan for long-term trend during 1990–2010: causes of temporal and spatial variation. *Atmos. Environ.*, 102(9), 302-310,2015.

15

Akimoto, H., Nagashima, T., Li, J., Fu, J. S., Ji, D., Tan, J., and Wang, Z.: Comparison of surface ozone simulation among selected regional models in MICS-Asia III-effects of chemistry and vertical transport for the causes of difference, *Atmos. Chem. Phys.*, 19, 603-615, <https://doi.org/10.5194/acp-19-603-2019>, 2019.

20

Ban, S., Matsuda, K., Sato, K., Ohizumi, T.: Long-term assessment of nitrogen deposition at remote EANET sites in Japan. *Atmos. Environ.*, 146, 70-78,2016.

Banks, R. F., & Baldasano, J. M.: Impact of wrf model pbl schemes on air quality simulations over catalonia, spain. *Sci. Total Environ*, 572, 98-113,2016.

25

Binkowski, F.S. and Roselle, S. J.: Models 3-Community Multiscale Air Quality (CMAQ) model aerosol component:1. Model description, *J. Geophys. Res.-Atmos.*, 108(D6), 4183, doi:10.1029/2001JD001409, 2003.

- Byun, D.W., Dennis, R.: Design artifacts in Eulerian air-quality models e evaluation of the effects of layer thickness and vertical profile correction on surface ozone concentrations. *Atmos. Environ.*, 29, 105-126,1995.
- Carlton, A. G., Turpin, B. J., Altieri, K. E., Seitzinger, S., Reff, A., Lim, H.-J., and Ervens, B.:
5 Atmospheric oxalic acid and SOA production from glyoxal: Results of aqueous photooxidation experiment, *Atmos. Environ.*, 41, 7588–7602, 2007.
- Carmichael, G. R., Calori, G., Hayami, H., Uno, I., Cho, S. Y., Engardt, M., Kim, S. B., Ichikawa, Y., Ikeda, Y., Woo, J. H., Ueda, H., Amann, M.: The MICS-Asia study: model intercomparison of long-range transport and sulfur deposition in East Asia, *Atmos. Environ.*, 36, 175-199,2002.
- 10 Carmichael, G. R., Sakurai, T., Streets, D., Hozumi, Y., Ueda, H., Park, S.U., Fung, C., Han, Z., Kajino, M., Engardt, M., Bennet, C., Hayami, H., Sartelet, K., Holloway, T., Wang, Z., Kannari, A., Fu, J., Matsuda, K., Thongboonchoo, N., Amann, M., MICS-Asia II: The model intercomparison study for Asia Phase II methodology and overview of findings, *Atmos. Environ.*, 42(15), 3468-3490, 2008.
- Carter, W. P. L.: Implementation of the SAPRC-99 Chemical Mechanism into the Models-3 Framework,
15 Report to the United States Environmental Protection Agency, available at: <http://www.engr.ucr.edu/~carter/pubs/s99mod3.pdf> (last access: 6 February 2015), 2000.
- Colella, P., and Woodward, P. L.: The piecewise parabolic method (PPM) for gas-dynamical simulations, *J. Comput. Phys.*, 54, 174–201,1984
- Easter, R. C., Ghan, S. J., Zhang, Y., Saylor, R. D., Chapman, E. G., Laulainen, N. S., Abdul-Razzak,
20 H., Leung, L. R., Bian, X. D., and Zaveri, R. A.: MIRAGE: Model description and evaluation of aerosols and trace gases, *J. Geophys. Res.-Atmos.*, 109, 46, 10.1029/2004jd004571, 2004.
- ~~Engeln, A. V. and Teixeira, J. A.: Planetary boundary layer height climatology derived from ECMWF reanalysis Data, *J. Clim.*, 26(17), 6575–6590, 2013.~~
- Fiore, A. M., Dentener, F. J., Wild, O., Cuvelier, C., Schultz, M. G., Hess, P., Textor, C., Schulz, M.,
25 Doherty, R. M., Horowitz, L. W., MacKenzie, I. A., Sanderson, M. G., Shindell, D. T., Stevenson, D. S., Szopa, S., van Dingenen, R., Zeng, G., Atherton, C. S., Bergmann, D. J., Bey, I., Carmichael, G. R., Collins, W. J., Duncan, B. N., Faluvegi, G., Folberth, G. A., Gauss, M., Gong, S., Hauglustaine, D., Holloway, T., Isaksen, I. S. A., Jacob, D. J., Jonson, J. E., Kaminski, J. W.,

- Keating, T. J., Lupu, A., Marmer, E., Montanaro, V., Park, R. J., Pitari, G., Pringle, K. J., Pyle, J. A., Schroeder, S., Vivanco, M. G., Wind, P., Wojcik, G., Wu, S., and Zuber, A.: Multimodel estimates of intercontinental source-receptor relationships for ozone pollution, *J. Geophys. Res.-Atmos.*, 114(D4), 83-84, 2009.
- 5 Fountoukis, C. and Nenes, A.: ISORROPIA II: A Computationally Efficient Aerosol Thermodynamic Equilibrium Model for K^+ , Ca^{2+} , Mg^{2+} , NH_4^+ , Na^+ , SO_4^{2-} , NO_3^- , Cl^- , H_2O Aerosols, *Atmos. Chem. Phys.*, 7, 4639–4659, 2007.
- Ganzeveld, L., Helmig, D., Fairall, C. W., Hare, J., and Pozzer, A.: Atmosphere-ocean ozone exchange: A global modeling study of biogeochemical, atmospheric, and waterside turbulence dependencies,
10 *Global Biogeochem. Cy.*, 23, GB4021, doi:10.1029/2008GB003301, 2009.
- Gao, M., Han, Z., Liu, Z., Li, M., Xin, J., Tao, Z., Li, J., Kang, J.-E., Huang, K., Dong, X., Zhuang, B., Li, S., Ge, B., Wu, Q., Cheng, Y., Wang, Y., Lee, H.-J., Kim, C.-H., Fu, J. S., Wang, T., Chin, M., Woo, J.-H., Zhang, Q., Wang, Z., and Carmichael, G. R.: Air quality and climate change, Topic 3 of the Model Inter-Comparison Study for Asia Phase III (MICS-Asia III) – Part 1: Overview and
15 model evaluation, *Atmos. Chem. Phys.*, 18, 4859-4884, <https://doi.org/10.5194/acp-18-4859-2018>, 2018.
- Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B., Duncan, B.N., Martin, R.V., Logan, J., Higurashi, A., Nakajima T.: Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and sun photometer measurements, *J. Atmos. Phys.*, 59 , 461-483,2012.
- 20 [Gipson, G. L.: The Initial Concentration and Boundary Condition Processors. In Science algorithms of the EPA Models-3 Community Multiscale Air Quality \(CMAQ\) Modeling System, US Environmental Protection Agency Report, EPA-600/R-99/030, 12-1–12-91, 1999.](#)
- Goliff, W. S., Stockwell, W. R., Lawson, C. V.: The regional atmospheric chemistry mechanism, version 2. *Atmos. Environ.*, 68(1),174-185, 2013.
- 25 Guo, J., Miao, Y., Zhang, Y., Liu, H., Li, Z., Zhang, W., He, J., Lou, M., Yan, Y., Bian, L., and Zhai, P.: The climatology of planetary boundary layer height in china derived from radiosonde and reanalysis data, *Atmos. Chem. Phys.*, 16(20), 13309-13319, 2016.

Gutenther, A. K., T.; Harley, P.; Wiedinmyer, C.; Palmer, P.I.; Geron, C.: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols 921 from Nature, *Atmos. Chem. Phys.*, 6, 3181-3210, 2006.

Han, Z., Sakurai, T., Ueda, H., Carmichael, G. R., Streets, D., Hayami, H., Wang, Z., Holloway, T., Engardt, M., Hozumib, Y., Parkh, S.U., Kajinoi, M., Sarteletj, K., Funk, C., Bennetg, C., Thongboonchooc, N., Tangc, Y., Changk, A., Matsudal, K., Amannm, M. : MICS-ASIA II: model intercomparison and evaluation of ozone and relevant species, *Atmos. Environ.*, 42(15), 3491-3509,2008.

Hardacre, C., Wild, O., and Emberson, L.: An evaluation of ozone dry deposition in global scale chemistry climate models, *Atmos. Chem. Phys.*, 15, 6419-6436, <https://doi.org/10.5194/acp-15-6419-2015>, 2015.

He, J., Zhang, Y., Wang, K., Chen, Y., Leung, L. R., Fan, J., Li, M., Zheng, B., Zhang, Q., Duan, F., He, K. B.: Multi-year application of WRF-CAM5 over East Asia-Part II: Interannual variability, trend analysis, and aerosol indirect effects, *Atmos. Environ.*, 165,122-142, 2017.

[Helmig, D., Lang, E. K., Bariteau, L., Boylan, P., Fairall, C. W., Ganzeveld, L., Hare, J. E., Hueber, J., and Pallandt, M.: Atmosphere-ocean ozone fluxes during the TexAQS 2006, STRATUS 2006, GOMECC 2007, GasEx 2008, and AMMA 2008 cruises, *J. Geophys. Res.*, 117, D04305, doi:10.1029/2011JD015955, 2012.](#)

Hong, S. ~~Y~~->Y: A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, 134(9), 2318,2006.

Horowitz, L. W., Walters, S. M., Mauzerall, D. L., Emmons, L. K., Rasch, P. J., Granier, C., Tie, X., Lamarque, J.-F., Schultz, M. G., and Brasseur, G. P: A global simulation of tropospheric ozone and related tracers: description and evaluation of MOZART, version 2, *J. Geophys. Res.-Atmos.*, 108 (D24), 4784. <http://dx.doi.org/10.1029/2002JD002853>,2003.

Huang, X., Song, Y., Li, M., Li, J., Huo, Q., Cai, X., Zhu, T., Hu, M., and Zhang, H.: A high-resolution ammonia emission inventory in China, *Global Biogeochem. Cy.*, 26, GB1030, doi:10.1029/2011GB004161, 2012.

- Ji, D., Wang, Y., Wang, L., Chen, L., Hub, B., Tang, G., Xin, J., Song, T., Wen, T., Sun, Y., Pan, Y., and Liu, Z.: Analysis of heavy pollution episodes in selected cities of northern China. *Atmos. Environ.* 50, 338-348, 2012.
- Jin, X., Holloway, T.: Spatial and temporal variability of ozone sensitivity over China observed from the Ozone Monitoring Instrument, *J. Geophys. Res.-Atmos.*, 120(14),7229-7246,2015.
- Kajino, M., Inomata, Y., Sato, K., Ueda, H., Han, Z., An, J., Katata, G., Deushi, M., Maki, T., Oshima, N., Kurokawa, J., Ohara, T., Takami, A., and Hatakeyama, S.: Development of the RAQM2 aerosol chemical transport model and predictions of the Northeast Asian aerosol mass, size, chemistry, and mixing type, *Atmos. Chem. Phys.*, 12, 11833-11856, <https://doi.org/10.5194/acp-12-11833-2012>, 2012.
- Kajino, M., Deushi, M., Sekiyama, T. T., Oshima, N., Yumimoto, K., Tanaka, T. Y., Ching, J., Hashimoto, A., Yamamoto, T., Ikegami, M., Kamada, A., Miyashita, M., Inomata, Y., Shima, S., Adachi, K., Zaizen, Y., Igarashi, Y., Ueda, H., Maki, T., and Mikami, M.: NHM-Chem, the Japan Meteorological Agency's regional meteorology - chemistry model (v1.0): model description and aerosol representations, *Geosci. Model Dev. Discuss*, in review, doi:10.5194/gmd-2018-128, 2018.
- 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, doi:10.5194/acp-13-11019-2013, 2013.
- Lee, D. G., Lee, Y.M., Jang, K.W., Yoo, C., Kang, K.H., Lee, J.H., Jung, S.W., Park, J.M., Lee, S.B., Han, J.S., Hong, J.H., and Lee, S.J.: Korean national emissions inventory system and 2007 air pollutant emissions, *Asian J. Atmos. Environ.*, 5, 278-291, 2011.
- Lelieveld, J. and Dentener, F. J.: What controls tropospheric ozone?, *J. Geophys. Res.*, 105, 3531–3551, doi:10.1029/1999JD901011, 2000
- Li, J., Chen, X., Wang, Z., Du, H., Yang, W., Sun, Y., Hu, B., Li, J. J., Wang, W., Wang, T., Fu, P., Huang, H.: Radiative and heterogeneous chemical effects of aerosols on ozone and inorganic aerosols over East Asia, *Sci. Total Environ.*, 622-623,1327-1342, 2018.

- Li, J., Wang, Z., Akimoto, H., Gao, C., Pochanart, P., and Wang, X.: Modeling study of ozone seasonal cycle in lower troposphere over East Asia, *J. Geophys. Res.-Atmos.*, 112, D22S25, doi:10.1029/2006JD008209, 2007.
- Li, J., Wang, Z., Akimoto, H., Tang, J., Uno, I.: Near-ground ozone source attributions and outflow in central eastern China during MTX2006 *Atmos. Chem. Phys.*, 8, 7335-7351, 2008.
- Li, J., Yang, W., Wang, Z., Chen, H., Hu, B., Li, J., Sun, Y., Fu, P., Zhang, Y.: Modeling study of surface ozone source-receptor relationships in east Asia. *Atmos. Res.*, 167, 77-88, 2016.
- Li, M., Zhang, Q., Kurokawa, J. I., Woo, J. H., He, K., Lu, Z., Ohara, T., Song, Y., Streets, D. G., Carmichael, G. R., Cheng, Y., Hong, C., Huo, H., Jiang, X., Kang, S., Liu, F., Su, H., and Zheng, B.: MIX: a mosaic Asian anthropogenic emission inventory under the international collaboration framework of the MICS-Asia and HTAP, *Atmos. Chem. Phys.*, 17, 935-963, doi:10.5194/acp-17-935-2017, 2017.
- Li, M., Zhang, Q., Kurokawa, J., Woo, J.H., He, K.B., Lu, Z., Ohara, T., Song, Y., Streets, D.G., Carmichael, G.R., Cheng, Y.F., Hong, C.P., Huo, H., Jiang, X.J., Kang, S.C., Liu, F., Su, H., Zheng, B.: MIX: a mosaic Asian anthropogenic emission inventory for the MICS-Asia and the HTAP projects. *Atmos. Chem. Phys. Discuss.* 15 (23), 34813-34869. <http://dx.doi.org/10.5194/acpd-15-34813-2015>, 2015.
- [Loon, M., Vautard, R., Schaap, M., Bergstr, M. R., Bessagnet, B., Brandt, J., Builtjes, P.J.H., Christensen, J. H., Curvelier, C., Graff, A., Jonson, J. E., Krol, M., Langner, J., Roberts, P., Rouil, L.M., Stern, R., Tarrason, L., Thunis, P., Vignati, E., White, L., Wind, P.: Evaluation of long-term ozone simulations from seven regional air quality models and their ensemble. *Atmos. Environ.*, 41\(10\), 2083-2097, 2007.](#)
- Lin, J. T., & Mcelroy, M. B. : Impacts of boundary layer mixing on pollutant vertical profiles in the lower troposphere: implications to satellite remote sensing. *Atmos. Environ.*, 44(14), 1726-1739, 2010.
- [Liu, S. C., McKeen, S. A., Hsie, E-Y., Lin, X., Kelly, K. K., Bradshaw, J. D., Sandholm, S. T., Browell, E. V., Gregory, G. L., Sachse, G. W., Bandy, A. R., Thornton, D. C., Blake, D. R., Rowland, F. S., Newell, R., Heikes, B. G., Singh, H., and Talbot, R. W. : Model study of tropospheric trace species distributions during PEM-West A. *J. Geophys. Res.*, 101, 2073-2085, 1996.](#)

5 Liu, X. H., Zhang, Y., Xing, J., Zhang, Q., Wang, K., Streets, D. G., Jang, C., Wang, W. X., Hao, J. M.:
 Understanding of regional air pollution over China using CMAQ, part II. Process analysis and
 sensitivity of ozone and particulate matter to precursor emissions, *Atmos. Environ.* 44, 3719-3727,
 2010.

Lu, Z., and Streets, D. G.: Increase in NO_x Emissions from Indian Thermal Power Plants during 1996-
 2010: Unit-Based Inventories and Multisatellite Observations, *Environ. Sci. Technol.*, 46, 7463-
 7470, doi:10.1021/es300831w, 2012.

10 Lu, Z., Zhang, Q., and Streets, D. G.: Sulfur dioxide and primary carbonaceous aerosol emissions in
 China and India, 1996-2010, *Atmos. Chem. Phys.*, 11, 9839-9864, doi:10.5194/acp-11-9839-2011,
 2011.

~~Luecken, D. J., Phillips, S., Sarwar, G., and Jang, C.: Effects of using the CB05 vs. SAPRC99 vs. CB4
 chemical mechanism on model predictions: ozone and gas phase photochemical precursor
 concentrations. *Atmos. Environ.*, 42(23), 5805-5820, 2008.~~

15 Martin, R. V., Jacob, D. J., Logan, J. A., Bey, I., Yantosca, R. M., Staudt, A. C., Li, Q., Fiore, A. M.,
 Duncan, B. N., and Liu, H.: Interpretation of TOMs observations of tropical tropospheric ozone
 with a global model and in situ observations, *J. Geophys. Res.-Atmos.*, 107(D18), ACH 4-1-ACH
 4-27, 2002.

20 Nagashima, T., Ohara, T., Sudo, K., and Akimoto, H.: The relative importance of various source regions
 on East Asian surface ozone, *Atmos. Chem. Phys.*, 10, 11305-11322, <https://doi.org/10.5194/acp-10-11305-2010>, 2010.

Nenes, A., Pandis, S.N., Pilinis, C. : ISORROPIA: A new thermodynamic equilibrium model for
 multiphase multicomponent inorganic aerosols, *Aquat. Geoch.*, 4, 123-152, 1998.

25 Olson, J., Prather, M., Berntsen, T., Carmichael, G., Chatfield, R., Connell, P., Derwent, R., Horowitz,
 L., Jin, S., Kanakidou, M., Kasibhatla, P., Kotamarthi, R., Kuhn, M., Law, K., Penner, J., Perliski,
 L., Sillman, S., Stordal, F., Thompson, A., and Wild, O.: Results from the intergovernmental panel
 on climatic change photochemical model intercomparison(PhotoComp), *J. Geophys. Res.-Atmos.*,
 102 (D5), 5979-5991, 1997.

Pan, X., Wang Z., Wang X., Dong H., Xie, F., Guo, Y.: An observation study of ozone dry deposition over grassland in the suburban area of Beijing. Chinese Journal of Atmospheric Sciences (in Chinese), 34(1), 120-130, 2010.

5 Pleim, J. E., Xiu, A., Finkelstein, P. L., and Otte, T. L.: A Coupled Land-Surface and Dry Deposition Model and Comparison to Field Measurements of Surface Heat, Moisture, and Ozone Fluxes, *Water Air Soil Poll.*, 1, 243–252, 2001.

Pleim, J. E.: A combined local and nonlocal closure model for the atmospheric boundary layer, Part I: Model description and testing, *J. Appl. Meteor. Climatol.*, 46, 1383–1395, 2007.

10 Quan, J., Tie, X., Zhang, Q., Liu, Q., Li, X., Gao, Y., and Zhao, D.: Evolution of planetary boundary layer under different weather conditions, and its impact on aerosol concentrations, *Particuology*, 11(1), 34-40, 2013.

Rao, S. T., Galmarini, S., Puckett, A. K.: Air quality model evaluation international initiative (AQMEII): advancing state-of-science in regional photochemical modeling and its applications, *BAMS*, 23-30, 2011.

15 Santanello, A., Lidard, C. D., Kennedy, A., Kumar, S.V.: Diagnosing the nature of land-atmosphere coupling: a case study of dry/wet extremes in the U.S. Southern Great Plains, *J. Hydrometeor.*, 14, 3-24, 10.1175/JHM-D-12-023.1,2013.

Sillman, S., and He, D.: Some theoretical results concerning O₃-NO_x-VOC chemistry and NO_x-VOC indicators, *J. Geophys. Res.-Atmos.*, 107(D22), 4659, doi:10.1029/2001JD001123, 2002.

20 Solazzo, E., Bianconi, R., Pirovano, G., Matthias, V., Vautard, R., Appel, K. W., Bessagnet, B., Brandt, J., Christensen, J. H., Chemel, C., Coll, I., Ferreira, J., Forkel, R., Francis, X. V., Grell, G., Grossi, P., Hansen, A., Miranda, A. I., Moran, M. D., Nopmongco, U., Parnk, M., Sartelet, K. N., Schaap, M., D. Silver, J., Sokhi, R. S., Vira, J., Werhahn, J., Wolke, R., Yarwood, G., Zhang, J., Rao, S. T.,
25 Galmarin, S.: Model evaluation and ensemble modelling of surface-level ozone in Europe and north America in the context of AQMEII, *Atmos. Environ.*, 53(6), 60-74,2012.

Sorimachi, A, Sakamoto, K, Ishihara H, Fukuyama, T., Utiyama, M., Liu, H., Wang, W., Tang, D., Dong, X., Qian, H.: Measurements of sulfur dioxide and ozone dry deposition over short vegetation in northern China-A preliminary study. *Atmos. Environ.*, 37(22), 3157-3166, 2003.

5 Stockwell, W. R., Middleton, P., Chang, J. S. and Tang, X.: The second generation regional Acid Deposition Model chemical mechanism for regional air quality modeling, *J. Geophys. Res.*, 95, 16,343-16,367, 1990.

Streets, D. G., Fu, J. S., Jang, C. J., Hao, J. M., He, K. B., Tang, X. Y., Zhang, Y. H., Wang, Z. F., Li, Z. P., Zhang, Q., Wang, L. T., Wang, B. Y., and Yu, C: Air quality during the 2008 Beijing Olympic
10 games, *Atmos. Environ*, 41(3), 480-492, 2007.

Sudo, K., Takahashi, M., Kurokawa, J. I., Akimoto, H.: Chaser: a global chemical model of the troposphere 1. model description. *J. Geophys. Res.-Atmos.*, 107(D17), ACH 7-1–ACH 7-20., 2002a.

Sudo, K., Takahashi, M., Akimoto, H., CHASER: A global chemical model of the troposphere 2. Model results and evaluation, *J. Geophys. Res.*, 107, 10.1029/2001/JD001114, 2002b.

15 Tang, H., Takigawa, M., Liu, G. , Zhu, J., Kobayashi, K.: A projection of ozone induced wheat production loss in China and India for the years 2000 and 2020 with exposure-based and flux-based approaches, *Glob. Change Biol.*, 19, 2739-2752, 2013.

Tang, X, Wang, Z., Zhu, J., Abaguiddi, A., Wu, Q., Li, J., Zhu, T.: Sensitivity of ozone to precursor emissions in urban Beijing with a Monte Carlo scheme. *Atmos. Environ.*, 44(31),3833-3842, 2010.

20 Tao, Z., Santanello, J. A., Chin, M., Zhou, S., Tan, Q., Kemp, E. M., and Peters-Lidard, C. D.: Effect of land cover on atmospheric processes and air quality over the continental United States – a NASA Unified WRF (NU-WRF) model study, *Atmos. Chem. Phys.*, 13, 6207-6226, <https://doi.org/10.5194/acp-13-6207-2013>, 2013.

The Royal Society: Ground-level ozone in the 21st century: future trends, impacts and policy implications,
25 Policy Document, 15/08,2008.

von Engel, A. V. and Teixeira, J. A.: Planetary boundary layer height climatology derived from ECMWF reanalysis Data, *J. Clim.*, 26(17), 6575-6590, 2013.

- Walcek, C. J. and Aleksic, N. M.: A simple but accurate mass conservative peak-preserving, mixing ratio bounded advection algorithm with fortran code, *Atmos. Environ.*, 32, 3863–3880, 1998
- Walcek, C. J. and Taylor, G. R.: A theoretical method for computing vertical distributions of acidity and sulfate production within cumulus clouds, *J. Atmos. Sci.*, 43, 339–355, 1986.
- 5 Wang, S., Ackermann, R., and Stutz, J.: Vertical profiles of O₃ and NO_x chemistry in the polluted nocturnal boundary layer in Phoenix, AZ: I. Field observations by long-path DOAS, *Atmos. Chem. Phys.*, 6, 2671-2693, <https://doi.org/10.5194/acp-6-2671-2006>, 2006.
- Wang, W. N., Cheng, T. H., Gu, X. F., Chen, H., Guo, H., Wang, Y., Bao, F. W., Shi, S. Y., Xu, B. R., Zuo, X., Meng, C., Zhang, X. C.. Assessing spatial and temporal patterns of observed ground-level
10 ozone in China. *Scientific Reports*, 7(1), 3651. doi:10.1038/s41598-017-03929-w, 2007.
- Wang, Y. X., Shen, L. L., Wu, S., Mickley, L., He, J. W., Hao, J.: Sensitivity of surface ozone over China to 2000–2050 global changes of climate and emissions, *Atmos. Environ.*, 75, 374-382, 2013.
- Wesely, M. L.: Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models, *Atmos. Environ.*, 23(6), 1293-1304,1989.
- 15 World Health Organization (WMO), WHO air quality guidelines global update, report on a working group meeting. Born, Germany,18-20 October, 2005, *Rep.* ,25 pp., Geneva, 2005.
- Yamaji, K., Ohara, T., Uno, I., Tanimoto, H., Kurokawa, J. I., Akimoto, H.: Analysis of the seasonal variation of ozone in the boundary layer in east Asia using the community multi-scale air quality model: what controls surface ozone levels over Japan?, *Atmos. Environ.*, 40(10), 1856-1868, 2006.
- 20 Yamartino, R. J.: Nonnegative, conserved scalar transport using grid-cell-centered, spectrally constrained Blackman cubics for applications on a variable-thickness mesh, *Mon. Weather Rev.*, 121, 753–763, 1993
- Yarwood, G., Rao, S., Yocke, M., Whitten, G.: Updates to the Carbon Bond Chemical Mechanism: CB05 Final Report to the US EPA, RT-0400675,2005.
- 25 Zaveri, R. A., Peters, L. K.: A new lumped structure photochemical mechanism for large-scale applications. *J. Geophys. Res.*104, 30387-30415,1999.
- Zhang, J., Trivikrama, Rao, S.: The role of vertical mixing in the temporal evolution of ground-level ozone concentrations, *J. Appl. Meteo.*, 38(38), 1674-1691,1998.

Zhang, L., Brook, J. L., and Vet, R.: A revised parameterization for gaseous dry deposition in air-quality models, *Atmos. Chem. Phys.*, 3, 2067-2082, 2003.

Zhang, Y.H., Hu, M., Zhong, L. J. , Wiedensohler, A., Liu, S.C., Andreae, M.O., Wang, W. , Fan, S. J.:
Regional Integrated Experiments on Air Quality over Pearl River Delta 2004 (PRIDE-PRD2004):
5 Overview, *Atmos. Environ.*, 42(25), 6157-6173, 2008.

Zhao, C., Wang, Y., Zeng. T.: East China Plains: a “Basin” of ozone pollution, *Environ. Sci. Technol.*,
43, 1911-1915, 2009.

Zhong, L., Louie, P. K., Zheng, J., Wai, K. M., Ho, J. W. K., Yuan, Z., Lau A. K. H., Yue D. L., Zhou
Y.: The pearl river delta regional air quality monitoring network - regional collaborative efforts on
10 joint air quality management. *Aero. Air Qual. Res.*, 13(5), 1582-1597, 2013.

15

20

25

Table and Figure captions:

Table.1 Basic structures, schemes and relevant parameters of the fourteen participating models

Table. 2 Statistical analysis for surface O₃ in three subregions over East Asia (R: correlation coefficient; NMB: Normalized Mean Bias; RMSE: Root Mean Square Error)

5 Table. 3 Statistical analysis for surface NO in three subregions over East Asia (R: correlation coefficient; NMB: Normalized Mean Bias; RMSE: Root Mean Square Error)

Table. 4 Statistical analysis for surface NO₂ in three subregions over East Asia (R: correlation coefficient; NMB: Normalized Mean Bias; RMSE: Root Mean Square Error)

10 Fig.1 Model domain of models except M13 and M14 with locations of ~~four~~three sub-regions marked in this study. Also show are locations of surface monitoring stations in this study. The meteorological model used for providing meteorological fields with most models also use this domain. Note that the domains of M13 and M14 are shown in Fig.10.

15 Fig.2 Box-plots of observed and simulated annual NO₂(left column), NO (middle column) and O₃ (right column) frequency distribution by 13 models, averaged in stations over EA1, ~~EA3~~EA2 and ~~EA4~~EA3, and in time for the whole 2010 year. n represents the numbers of stations. The rectangle represents the inter-quantile range (25th to 75th percentile). The small star identifies the mean, the continuous horizontal line inside the rectangle identifies the median, the whiskers extend between the minimum and maximum values.

20 Fig.3 Time series of monthly NO₂, NO and O₃ simulated by all models and their ensembles_(Ense), in ppbv, averaged over all observed stations in three subregions over East Asia (EA1: top row, ~~EA2~~EA3: middle row, ~~EA3~~EA4: bottom row). Observations are also shown by the black line. n represents the numbers of stations

25 Fig. 4 Seasonal mean diurnal cycle of surface O₃, in ppbv, as a function of hour, for all models and their ensembles, averaged over all observed stations in three subregions over East Asia (EA1: top row, ~~EA2~~EA3: middle row, ~~EA3~~EA4: bottom row). Observations are also shown by the black line. n represents the numbers of stations

30 Fig.5 Simulated O₃ profiles in summer and winter of 2010, averaged over all observed stations in three subregions over East Asia (EA1: left column, EA2: middle column, EA3: bottom column). The ozonesonde data observe in 2010 was taken from the data base stored by World Ozone and Ultraviolet Radiation Data Centre (WOUDC)

~~Simulated monthly daytime (08:00-20:00 LST) PBL height (m) by M1, M4, M7, M8 and M11, averaged over all observed stations in three subregions over East Asia (EA1: top row, EA3: middle row, EA4: bottom row). n represents the numbers of stations~~

Fig. 6 The ensemble mean seasonal surface O₃ concentrations and CV for the different seasons. CV is defined as the standard deviation of the modeled fields divided by the average, for the different seasons

The same as Fig.5, but for NO emission fluxes on the first day in each month. M2 was also shown

Fig.7 Surface O₃ spatial distribution from 13 models for summer 2010 (unit: ppbv).

5 The same as Fig.5, but for O₃ dry deposition velocities (V_d) of M1, M2, M4, M6, M11, M12, M13 and M14

Fig. 8 The modeled and observed monthly mean concentrations of O₃ at EANET sites in the phase II (left panel) and III (right panel) of MICS-ASIA project. Solid line represents ensemble mean. Note that data in MICS-ASIA II and III are in the period of March, July and December of 2001 and 2010, respectively. ID of Monitoring sites represents: 1: Rishiri (45.12°N, 141.23°E), 2: Ogasawara (27.83°N, 142.22°E), 3: Sado-seki (38.23°N, 138.4°E), 4: Oki (36.28°N, 133.18°E), 5: Hedo (26.85°N, 128.25°E), 6: Banryu (34.67°N, 131.80°E)

10 Scatter plots between monthly daytime (08:00-20:00) surface NO_x and O₃ at each station over EA1 (red), EA3 (green) and EA4 (blue) in May-October, for observations (obs) and models.

15 Fig. 9 Simulated and observed monthly O₃ dry deposition velocities (V_d) of M1, M2, M4, M6, M11, M12, M13 and M14 in three subregions over East Asia (EA1: top row, EA3: middle row, EA3: bottom row). TEX, STR, GGSEX and AMMA represents observations in TexAQSO6 (7 July–12 September 2006; north-western Gulf of Mexico), STRATUS06 (9–27 October 2006; the persistent stratus cloud region off Chile in the eastern Pacific Ocean), GasEx08 (29 February– 11 April 2008; the Southern Ocean), and AMMA08 (27 April–18 May 2008; the southern and northern Atlantic Ocean). Observation data is from Sorimachi et al. (2003), Pan et al. (2010), and Helmig et al. (2012).

20 Fig. 10 Scatter plots between monthly daytime (08:00-20:00) surface NO_x and O₃ at each station over EA1 (red), EA2 (green) and EA3 (blue) in May-October, for observations (obs) and models. Also shown are the linear regression equations between NO_x and O₃ in EA1 (red) and EA2 (green).

25 Simulated O_x (O₃+NO₂) profiles in summer and winter of 2010, averaged over all observed stations in three subregions over East Asia (EA1: left column, EA3: middle column, EA4: bottom column).

Simulated monthly daytime (08:00-20:00 LST) PBL height (m) by M1, M4, M7, M8 and M11, averaged over all observed stations in three subregions over East Asia (EA1: top row, EA3: middle row, EA4: bottom row). n represents the numbers of stations

30

Fig.10 Surface O₃ spatial distribution from 13 models for summer 2010 (unit:ppbv).

Fig.11 The same as Fig.10, but in winter 2010.

Fig.12 The ensemble mean seasonal surface O₃ concentrations and CV for the different seasons. CV is defined as the standard deviation of the modeled fields divided by the average, for the different seasons

5

Fig.13 The modeled and observed monthly mean concentrations of O_3 at EANET sites in the phase II (left panel) and III (right panel) of MICS ASIA project. Solid line represents ensemble mean. Note that data in MICS ASIA II and III are in the period of March, July and December of 2001 and 2010, respectively. ID of Monitoring sites represents: 1: Rishiri(45.12°N, 141.23°E), 2: Ogasawara(27.83°N, 142.22°E), 3: Sado-seki (38.23°N, 138.4°E), 4: Oki (36.28°N, 133.18°E), 5: Hedo(26.85°N,128.25°E), 6: Banryu (34.67°N,131.80°E)

Table1 Basic structures, schemes and relevant parameters of the fourteen participating models

Models	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14
Domain	Ref ^a	Ref ^a	Ref ^a	Ref ^a	Ref ^a	Ref ^a	Ref ^a	Ref ^a	Ref ^a	Ref ^a	Ref ^a	Ref ^a	Global	10 °N -50°N; 80 °E -135 °E
Horizontal resolution	45km	45km	45km	45km	45km	45km	45km	45km	45km	45km	45km	45km	0.5 ° ×0.667°	45km
Vertical resolution	40σ _p levels	40σ _p levels	40σ _p levels	40σ _p levels	40σ _p levels	40σ _p levels	40σ _p levels	40σ _p levels	40σ _p levels	60σ _p levels	20σ _z levels	40σ _p levels	47σ _p levels	15σ _z levels
Depth of first layer	58m	58m	58m	58m	58m	58m	29m	58m	16m	44m	48m	27m		100m
Meteorology	Standard ^b	Standard ^b	Standard ^b	Standard ^b	Standard ^b	Standard ^b	WRF/NCEP ^b	WRF/NCEP ^b	WRF/NCEP ^b	WRF/ MERRA2 ^b	Standard ^b	Standard ^b	GEOS-5	RAMS/NCEP ^b
Advection	Yamo (Yamartino, 1993)	Yamo	Yamo	PPM(Colle lla and Woodward 1984)	PPM	Yamo	5 th order monotonic	5 th order monotonic	5 th order monotonic	5 th order monotonic	Walcek and Aleksic (1998)	Walcek and Aleksic (1998)	PPM	PPM
Vertical diffusion	ACM2 (Pleim,2007)	ACM2	ACM2	ACM2	ACM2	ACM2	3 th order Monotonic	3 th order Monotonic	YSU	YSU	K-theory	FTCS (Forward in Time, Center in Space)	Lin and McElroy, (2010)	ACM2
Dry deposition	Wesely (1989)	Wesely (1989)	Wesely (1989)	M3DRY (Pleim et al., 2001)	M3DRY	M3DRY	Wesely (1989)	Wesely (1989)	Wesely (1989)	Wesely (1989)	Wesely (1989)	Wesely(1989)and Zhang et al. (2003)	Wesely (1989)	Wesely (1989)

Wet deposition	Henry's Law	Henry's Law	Henry's Law	Henry's Law	Henry's Law	ACM	Henry's Law	AQCHEM	Easter et al., (2004)	Grell	Henry's Law	Henry's Law	Henry's Law	Henry's Law
Gas chemistry	SAPRC99(Carter,2000)	SAPRC99	CBM05(Yarwood et al.,2005)	SAPRC99	SAPRC99	SAPRC99	RACM-ESRL with KPP	RACM (Goliff et al., 2013)	RADM2 (Stockwell et al., 1990)	RADM2	CBMZ (Zaveri et al.,1999)	SAPRC99(Carter,2000)	NOx-Ox-HC chemistry mechanism	SAPRC99
Aqueous chemistry	ACM-ae6	ACM-ae6	ACM-ae5	ACM-ae5	ACM-ae5	ACM-ae5	CMAQ simplified Aqueous chemistry	AQCHEM	Walcek and Taylor (1986)	None	RADM2 (Stockwell et al., 1990)	Walcek and Teylor (1986) Carlton et al. (2007)	-	ACM
Inorganic mechanism	AER06(Binkowski and Roselle, 2003)	AER06	AER05	AER05	AER05	AER05	MADE (Ackermann et al., 1998)	MADE	MADE	GOCART	ISORROP IAv1.7(Nenes et al.,1998)	Kajino et al. (2012)	ISORROPIAv1.7	ISORROPIAv1.7
Boundary conditions	GEOS-Chem global model (Martin et al.,2002)	DefaultGip son (1999)	GEOS-Chem global model	CHASER global model (Sudo et al., 2002a, 2002b)	CHASER global model	CHASER global model	Liu et al. (1996) Default	CHASER global model	GEOS-Chem global model	MOZART + GOCART global models ^c	CHASER global model	CHASER global model	/	GEOS-Chem global model
Two-way feedback	Off-line	Off-line	Off-line	Off-line	Off-line	Off-line	On-line	On-line	On-line	Off-line	Off-line	On-line	Off-line	Off-line

^a Ref represent the referenced domain by MICS-ASIA III project.

^b [StandardUnified](#) represents the reference meteorological field provided by MICS-ASIAIII project; WRF/NCEP and WRF/MERRA represents the meteorological field of the participating model itself, which was run by WRF driven by the NCEP and Modern Era Retrospective-analysis for Research and Applications (MERRA) reanalysis dataset.

*Boundary conditions of M10 are from MOZART and GOCART (Chin et al., 2002; Horowitz et al.,2003), which provided results for gaseous pollutants and aerosols, respectively.

Table 2 Statistical analysis for surface O₃ in three subregions over East Asia (R: correlation coefficient; NMB: Normalized Mean Bias; RMSE: Root Mean Square Error, unit is ppbv)

Models	Region	R	NMB	RMSE	Region	R	NMB	RMSE	Region	R	NMB	RMSE
M1		0.89	0.52	19.79		0.48	0.31	14.41		0.57	0.28	15.49
M2		0.90	0.64	18.13		0.10	0.35	15.06		0.66	0.24	13.83
M4		0.87	0.44	18.78		0.41	0.36	14.15		0.01	0.05	17.57
M5		0.87	0.42	19.00		0.30	0.14	13.38		0.34	0.31	19.28
M6		0.90	0.88	25.41		0.15	0.44	17.41		0.52	0.31	16.52
M7	EA1 (n=19) ^a	0.84	0.25	10.03	EA3 EA2 (n=13)	0.29	-0.08	11.11	EA4 EA3 (n=8)	0.60	0.02	10.97
M8		0.78	-0.47	13.52		0.20	-0.59	19.54		0.55	-0.27	15.32
M9		0.85	0.59	14.84		0.63	0.48	15.69		0.26	-0.09	13.27
M10		0.82	1.24	32.70		0.51	0.72	21.71		0.52	0.11	12.68
M11		0.81	0.09	9.46		0.34	-0.25	13.40		0.65	0.15	12.09
M12		0.89	0.55	18.53		0.36	0.30	13.31		0.57	0.11	11.81

M13	0.86	0.95	22.69	0.25	0.50	17.04	0.63	0.09	11.04
M14	0.86	0.75	23.33	0.12	0.40	17.01	-0.13	-0.30	20.03
Ensemble Mean	0.89	0.53	15.92	0.38	0.23	11.76	0.52	0.08	11.93
Ensemble Media	0.89	0.56	17.86	0.37	0.31	13.29	0.54	0.11	12.06

a: n represents the numbers of observation stations

Table 3 Statistical analysis for surface NO in three subregions over East Asia (R: correlation coefficient; NMB: Normalized Mean Bias; RMSE: Root Mean Square Error, unit is ppbv)

Models	Region	R	NMB	RMSE	Region	R	NMB	RMSE	Region	R	NMB	RMSE
M1		0.58	-0.35	20.68		0.22	-0.81	15.16		0.03	-0.35	0.23
M2		0.57	-0.14	23.73		0.14	-0.73	15.21		0.06	-0.27	0.19
M4		0.60	-0.61	22.29		0.18	-0.87	15.72		0.00	-0.39	0.20
M5		0.57	-0.07	20.34		0.24	-0.29	13.80		0.02	0.08	0.35
M6	EA1_	0.60	-0.71	23.36	EA3EA2	0.11	-0.89	15.94	EA4EA3	0.15	-0.70	0.16
M7	(n=19)	0.63	-0.75	24.91	(n=13)	0.04	-0.78	15.32	(n=8)	0.27	-0.40	0.15
M8		0.65	0.91	26.89		0.29	1.14	25.06		0.24	3.53	0.94
M9		0.58	-0.82	27.73		0.32	-0.93	16.72		0.22	-0.54	0.14
M10		0.63	-0.90	27.97		0.27	-0.94	16.30		0.39	-0.51	0.14
M11		0.61	-0.34	19.92		0.04	-0.05	14.86		0.41	0.09	0.14

M12	0.62	-0.55	21.19	0.13	-0.85	15.64	0.17	-0.48	0.16
M13	-	-	-	-	-	-	-	-	-
M14	0.68	-0.66	22.74	0.01	-0.66	14.77	0.24	-0.50	0.15
Ensemble Mean	0.63	-0.42	20.12	0.21	-0.55	13.58	0.20	-0.03	0.19
Ensemble Media	0.62	-0.58	21.66	0.17	-0.83	15.40	0.17	-0.45	0.16

a: n represents the numbers of observation stations

Table 4 Statistical analysis for surface NO₂ in three subregions over East Asia (R: correlation coefficient; NMB: Normalized Mean Bias; RMSE: Root Mean Square Error, unit is ppbv)

Models	Region	R	NMB	RMSE	Region	R	NMB	RMSE	Region	R	NMB	RMSE
M1		0.59	-0.18	11.08		0.33	-0.30	12.92		0.54	0.27	1.51
M2		0.64	-0.25	11.30		0.25	-0.43	14.85		0.43	-0.07	1.13
M4		0.65	-0.28	11.62		0.26	-0.32	13.79		0.56	-0.07	1.04
M5		0.57	0.08	10.86		0.30	0.09	12.91		0.60	0.46	1.79
M6	EA1_ (n=19)	0.65	-0.22	11.04	EA32_ (n=13)	0.23	-0.30	13.86	EA4EA3 (n=8)	0.56	-0.23	0.90
M7		0.59	-0.22	11.42		0.20	-0.25	13.24		0.65	0.19	1.42
M8		0.43	14.32	11.90		0.43	0.15	10.97		0.72	2.38	4.46
M9		0.60	32.30	18.80		0.51	-0.37	12.66		0.49	0.05	1.66
M10		0.61	-10.61	10.65		0.15	-0.08	12.81		0.63	0.06	1.33

M11	0.54	0.00	10.82	0.24	0.13	13.56	0.69	0.36	1.58
M12	0.63	-0.16	10.76	0.25	-0.24	13.78	0.61	-0.05	0.91
M13	-	-	-	-	-	-	-	-	-
M14	0.66	-0.12	10.00	0.08	-0.22	14.50	0.60	0.42	0.91
Ensemble Mean	0.65	-0.09	9.89	0.29	-0.18	12.16	0.64	0.25	1.33
Ensemble Media	0.65	-0.13	10.07	0.27	-0.23	12.85	0.59	0.06	1.23

a: n represents the numbers of observation stations

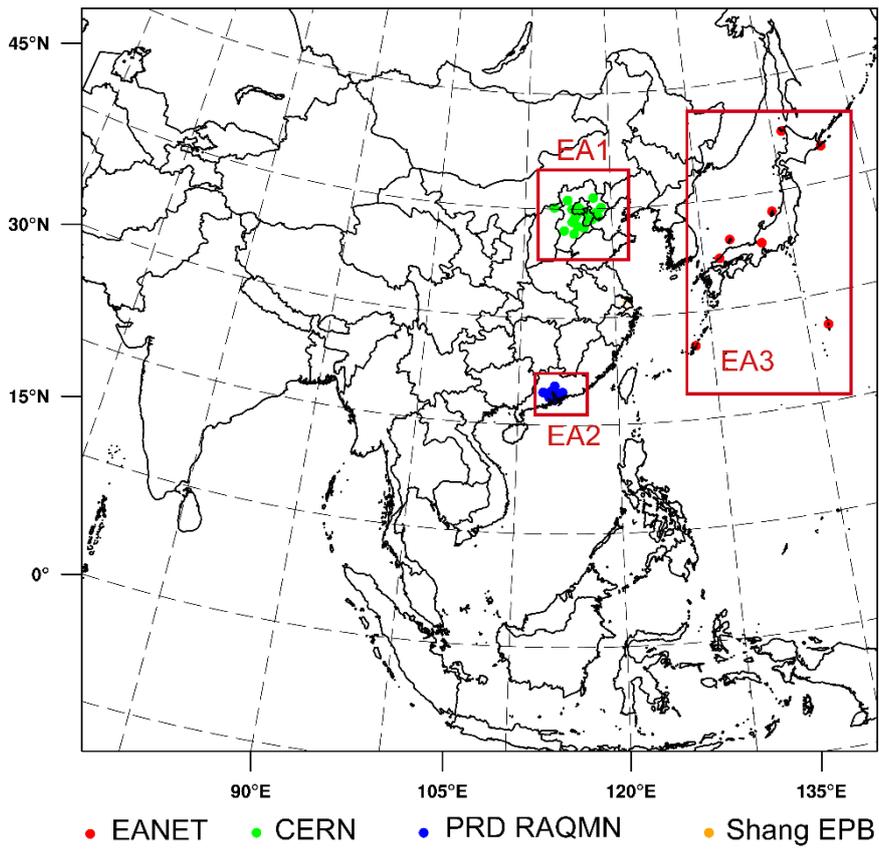
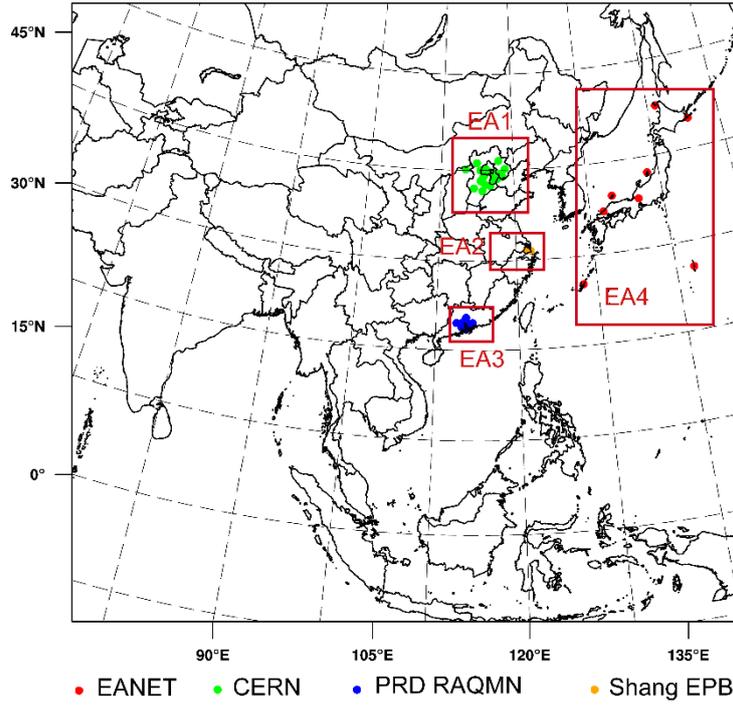


Fig.1 Li et al., 2018

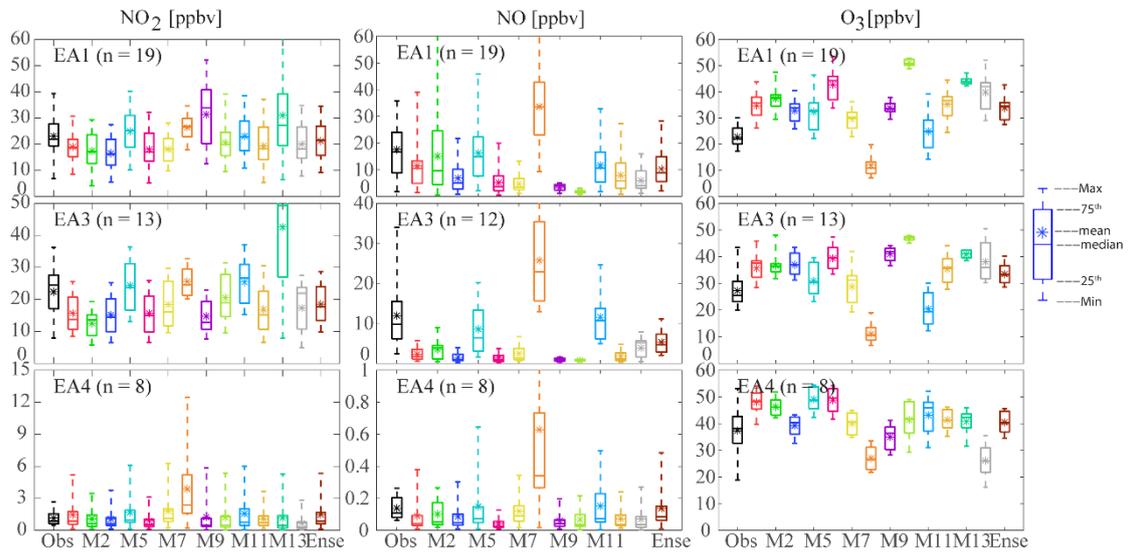
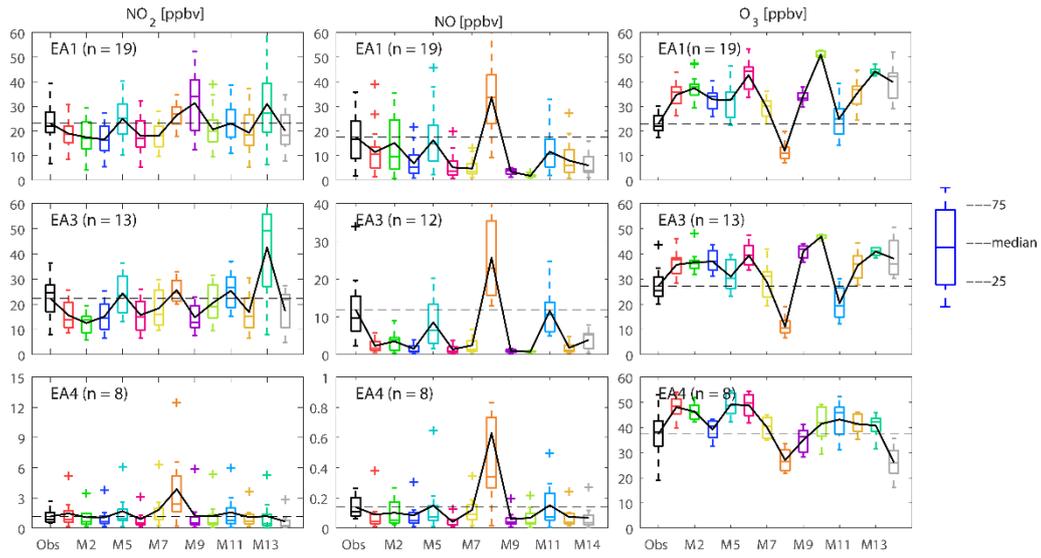


Fig.2 Li et al., 2018

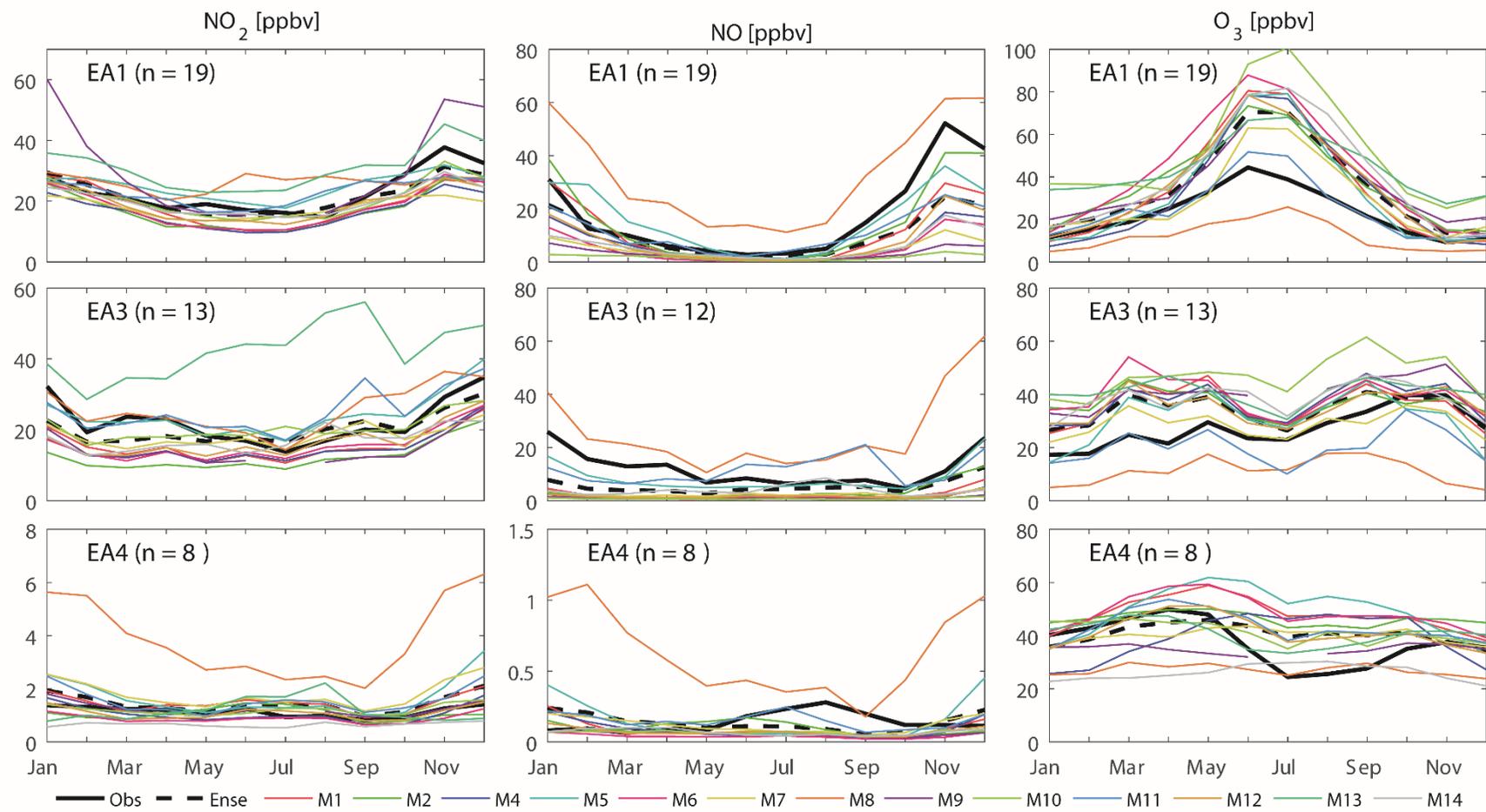


Fig.3 Li et al., 2018

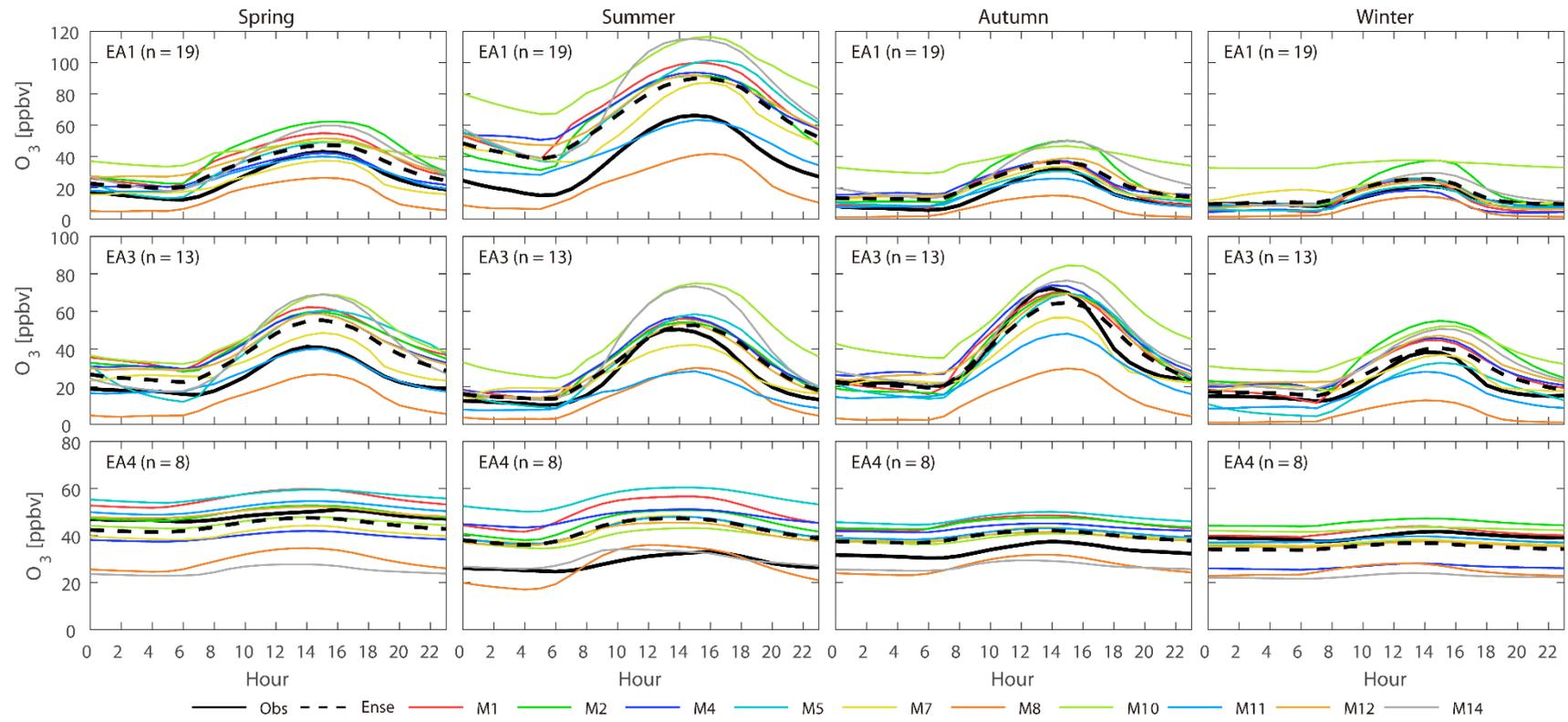


Fig.4 Li et al., 2018

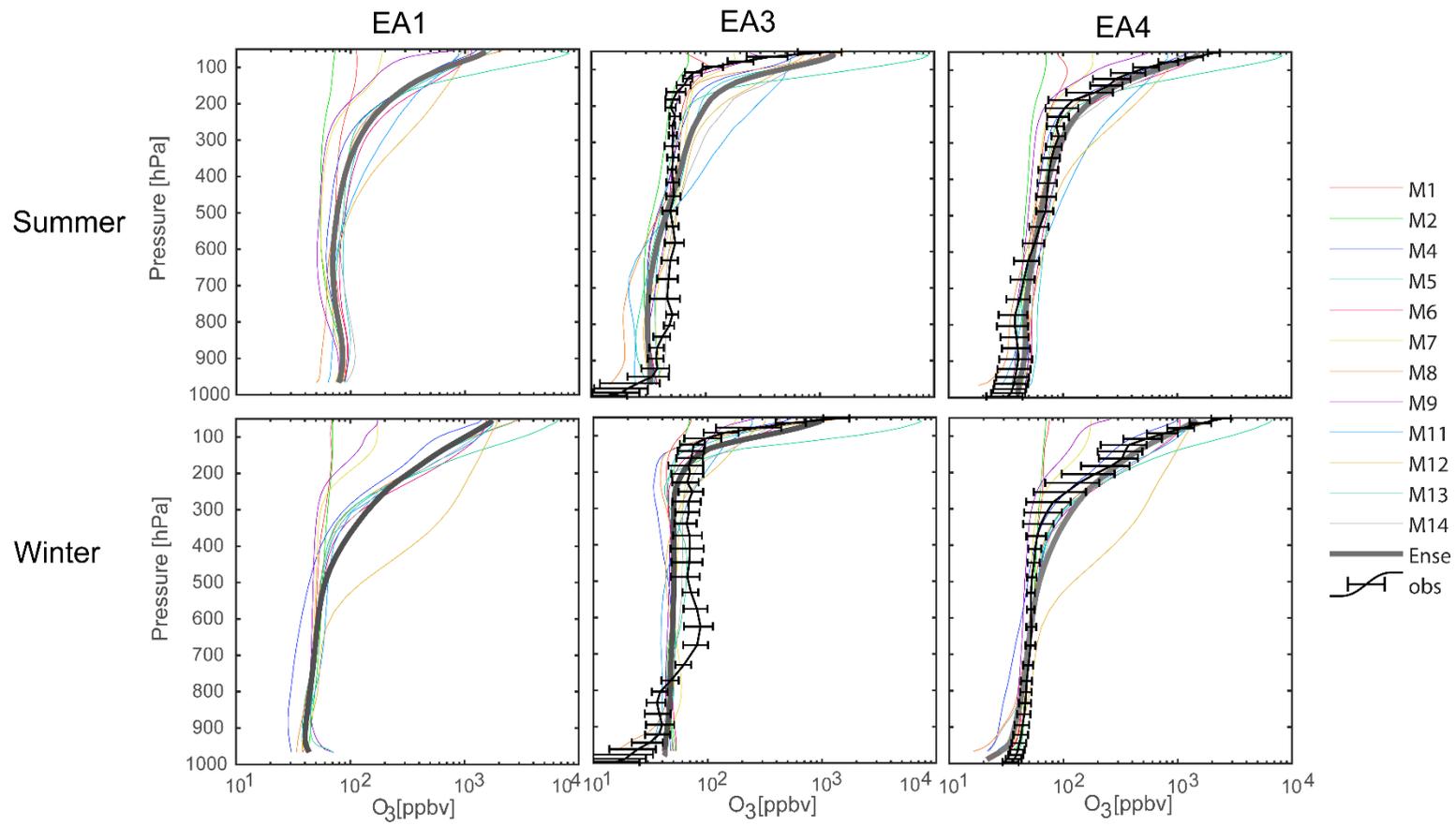


Fig.5 Li et al., 2018

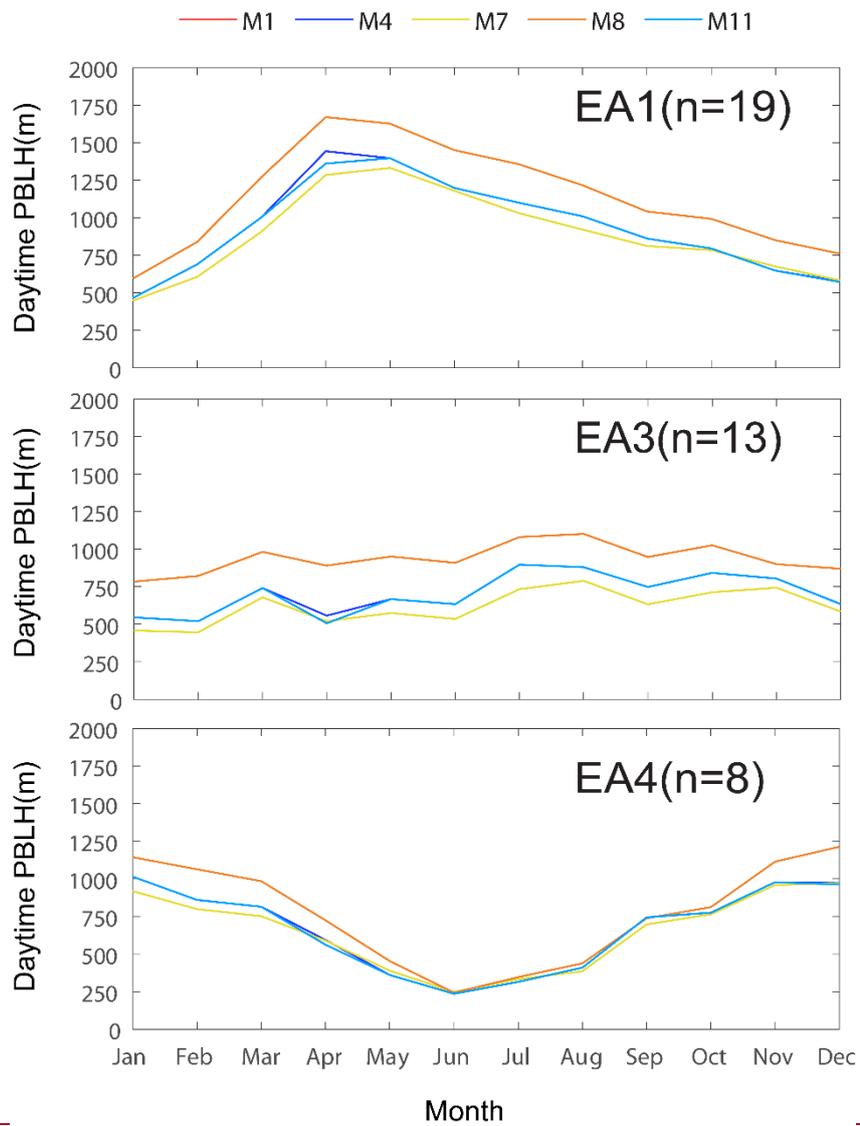
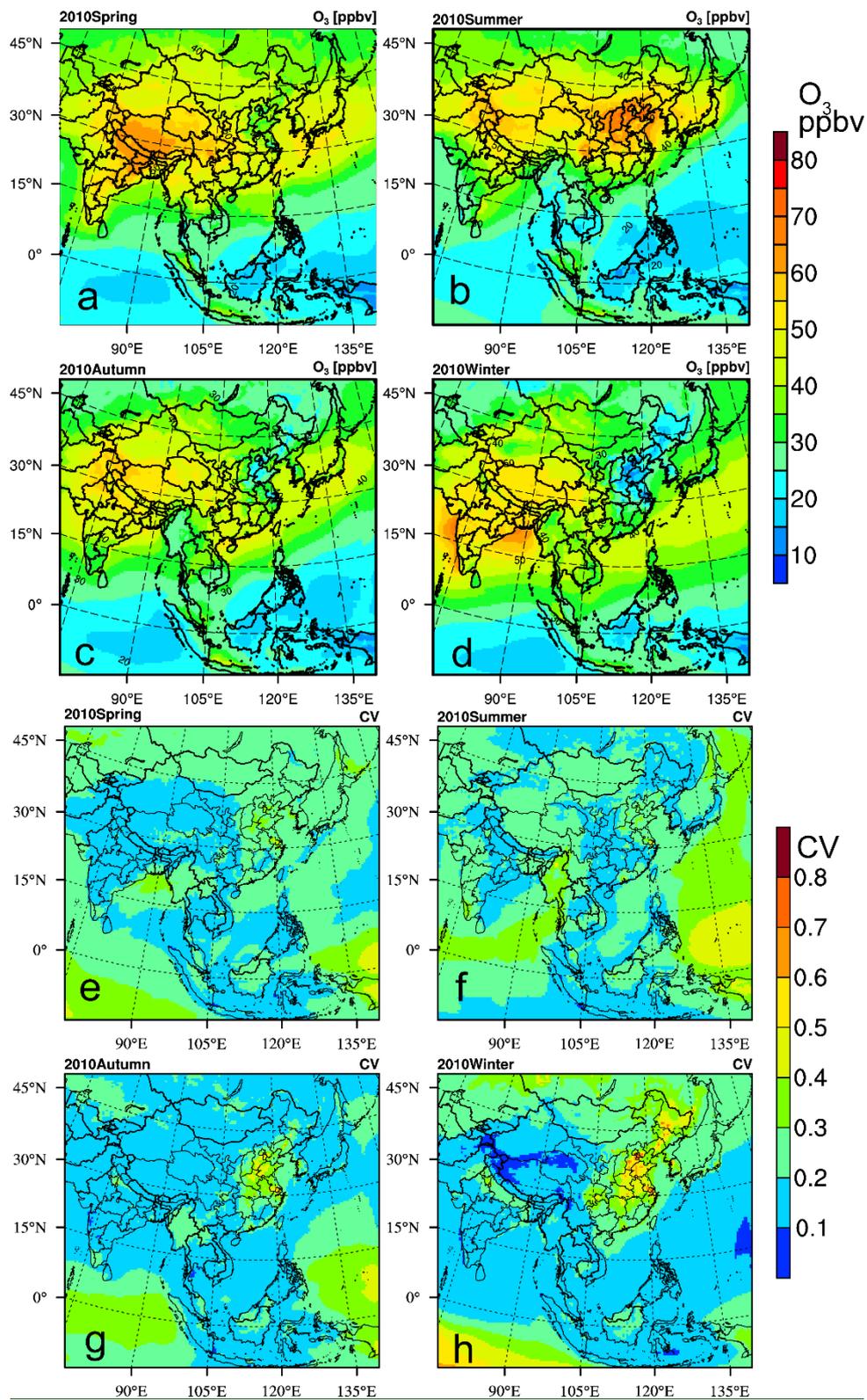


Fig. 5 Li et al., 2018



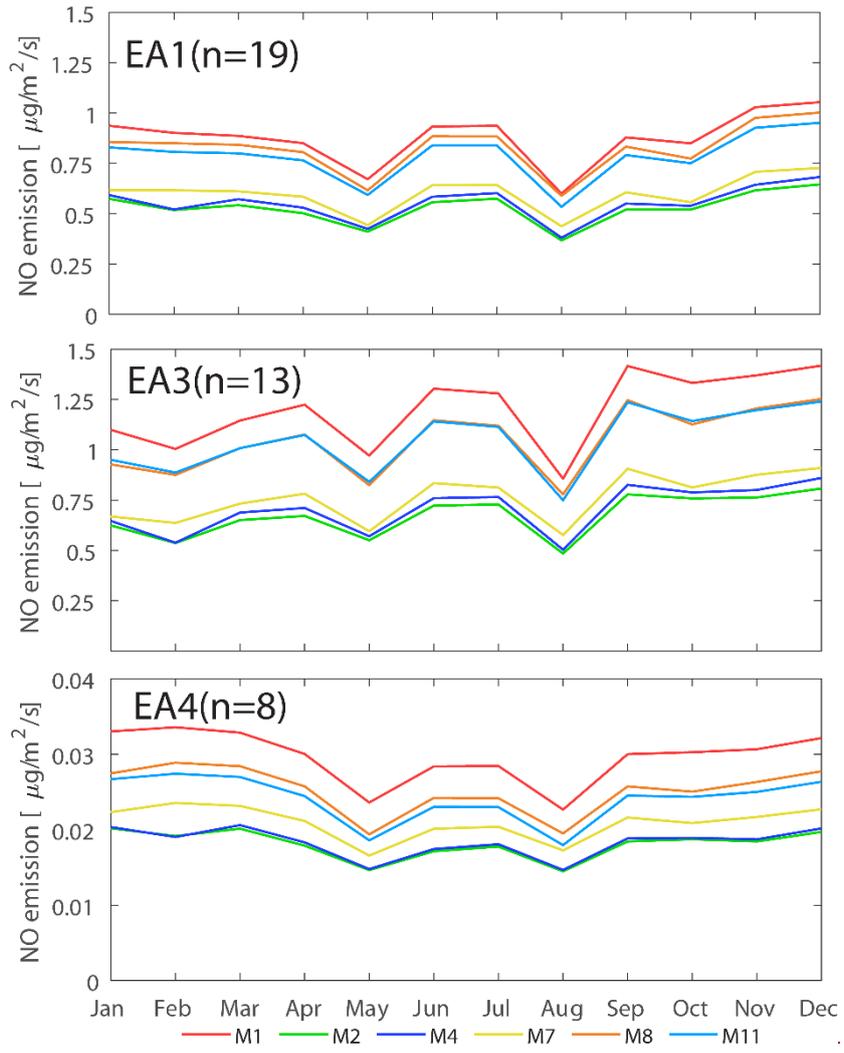


Fig.6 Li et al., 2018

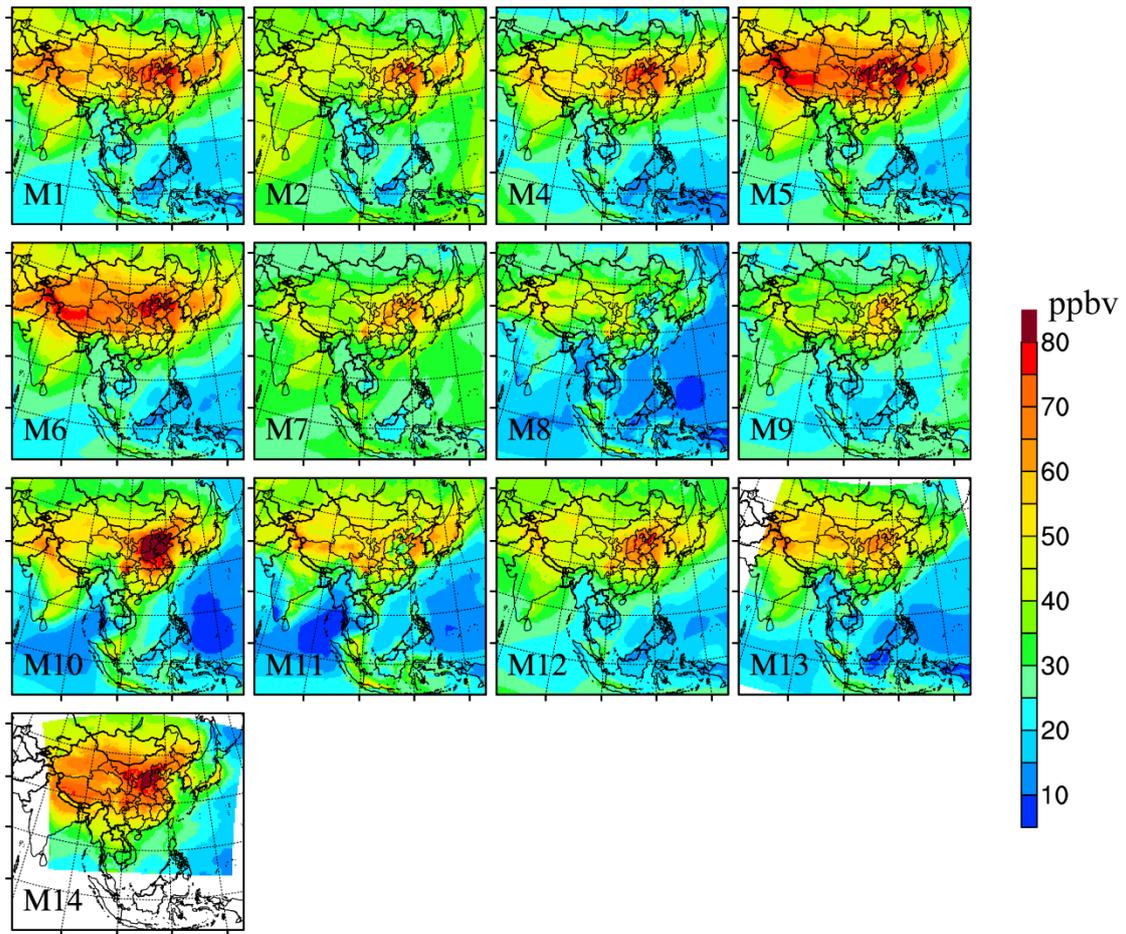


Fig.7 Li et al., 2018

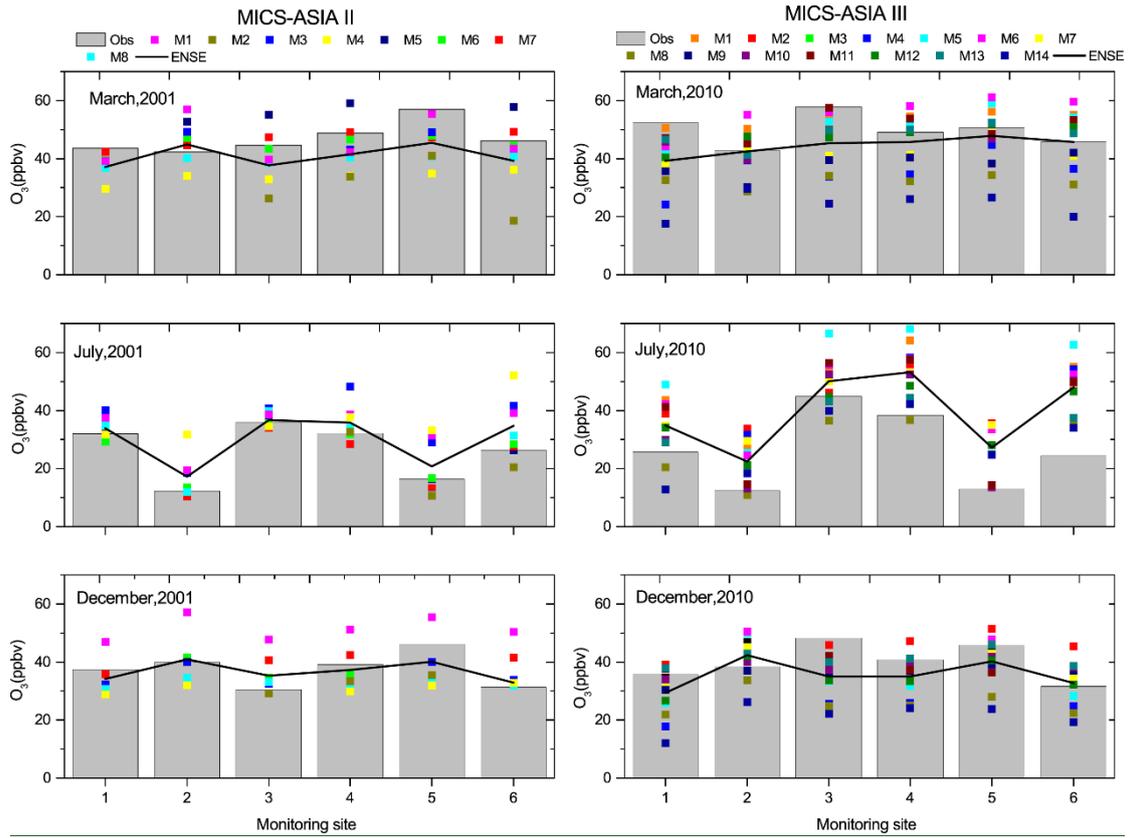
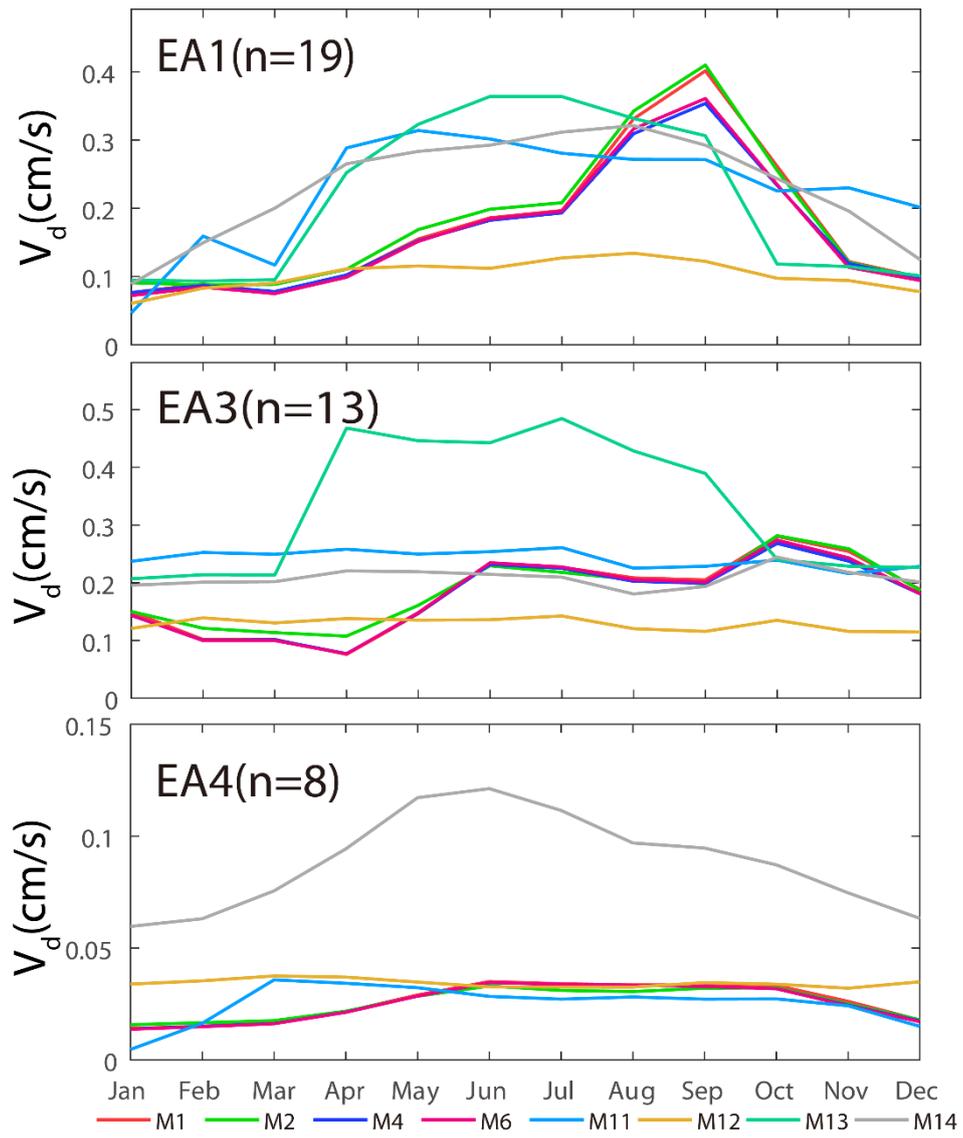


Fig.8 Li et al., 2018



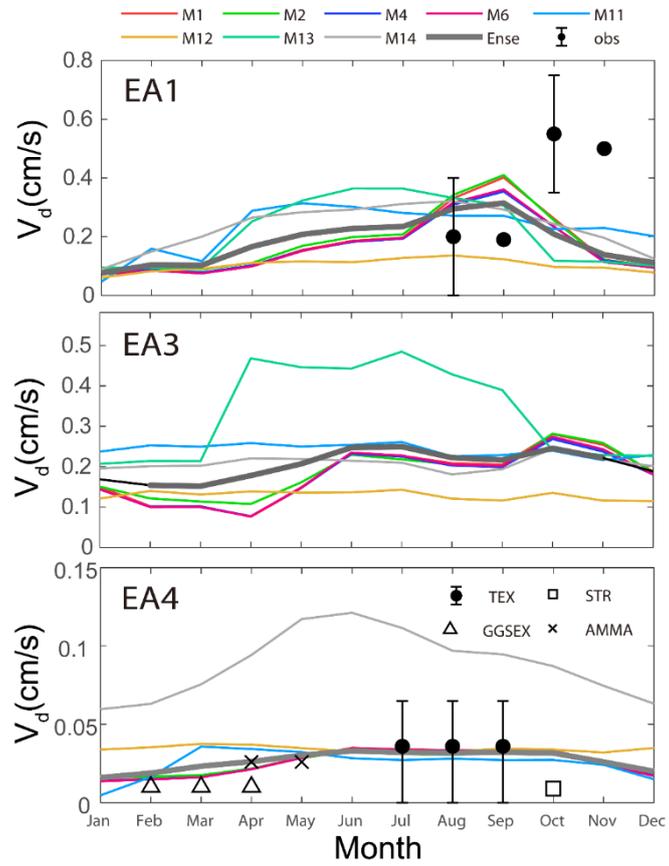
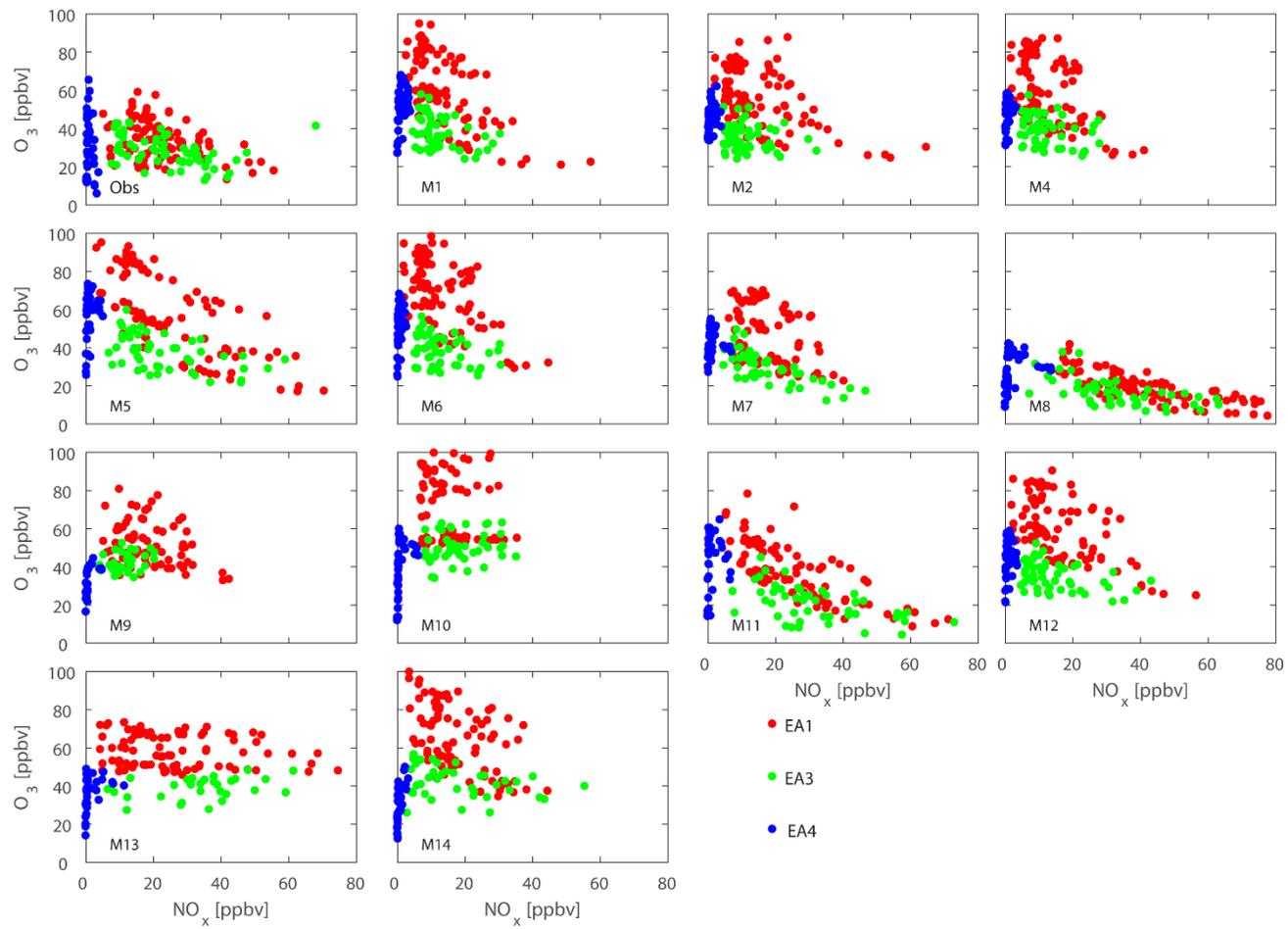


Fig.97 Li et al., 2018



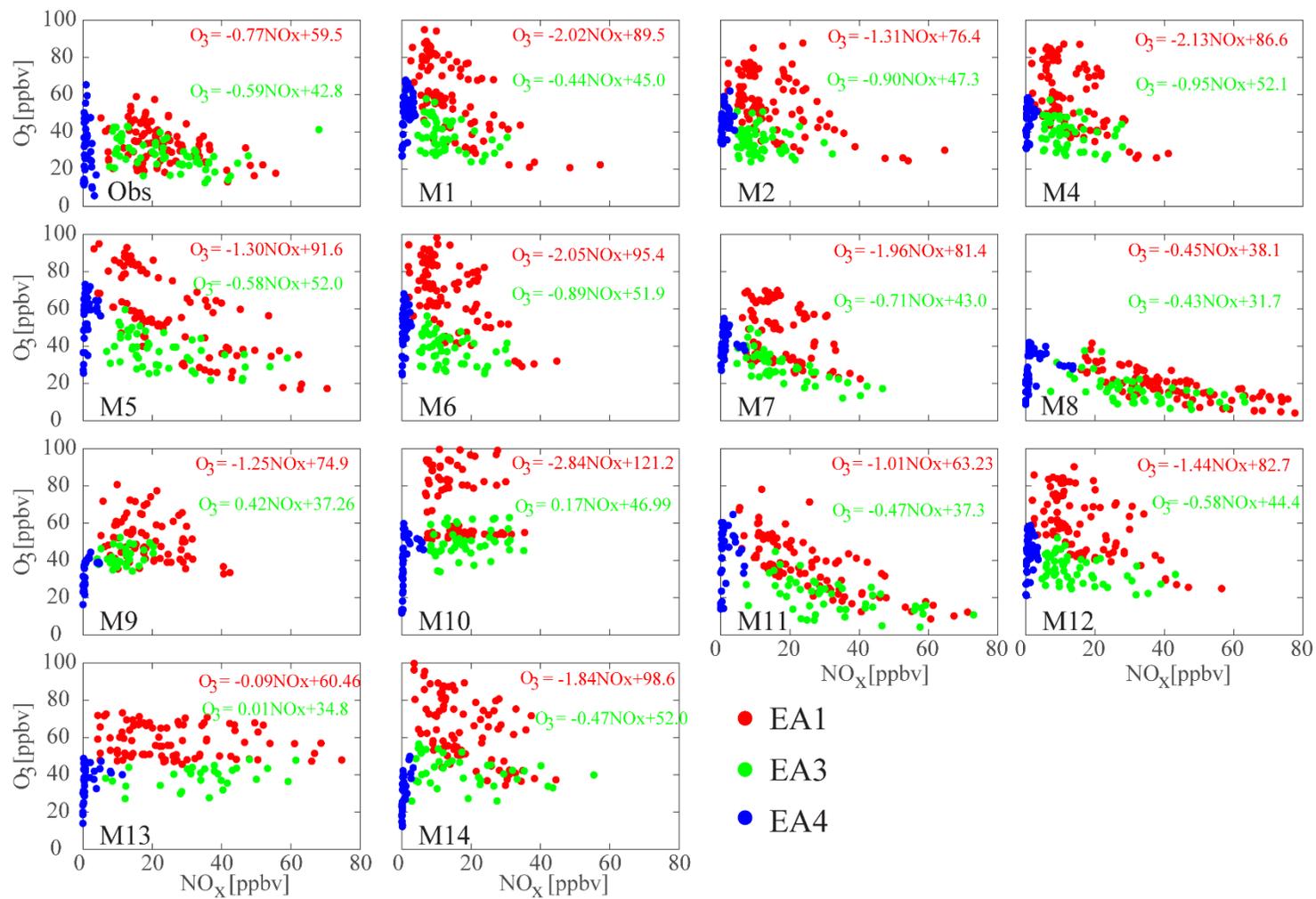


Fig. 8-10 Li et al., 2018

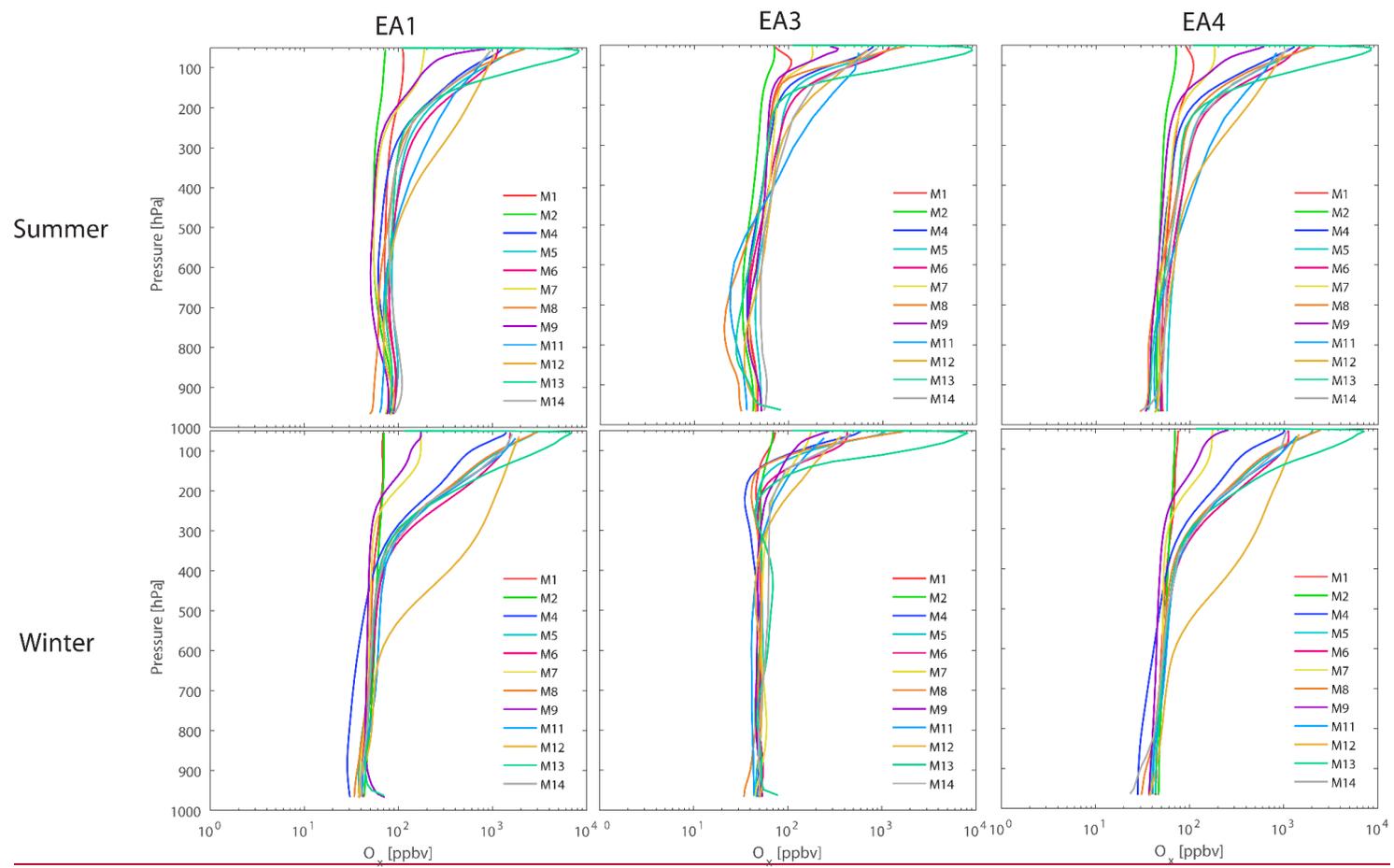
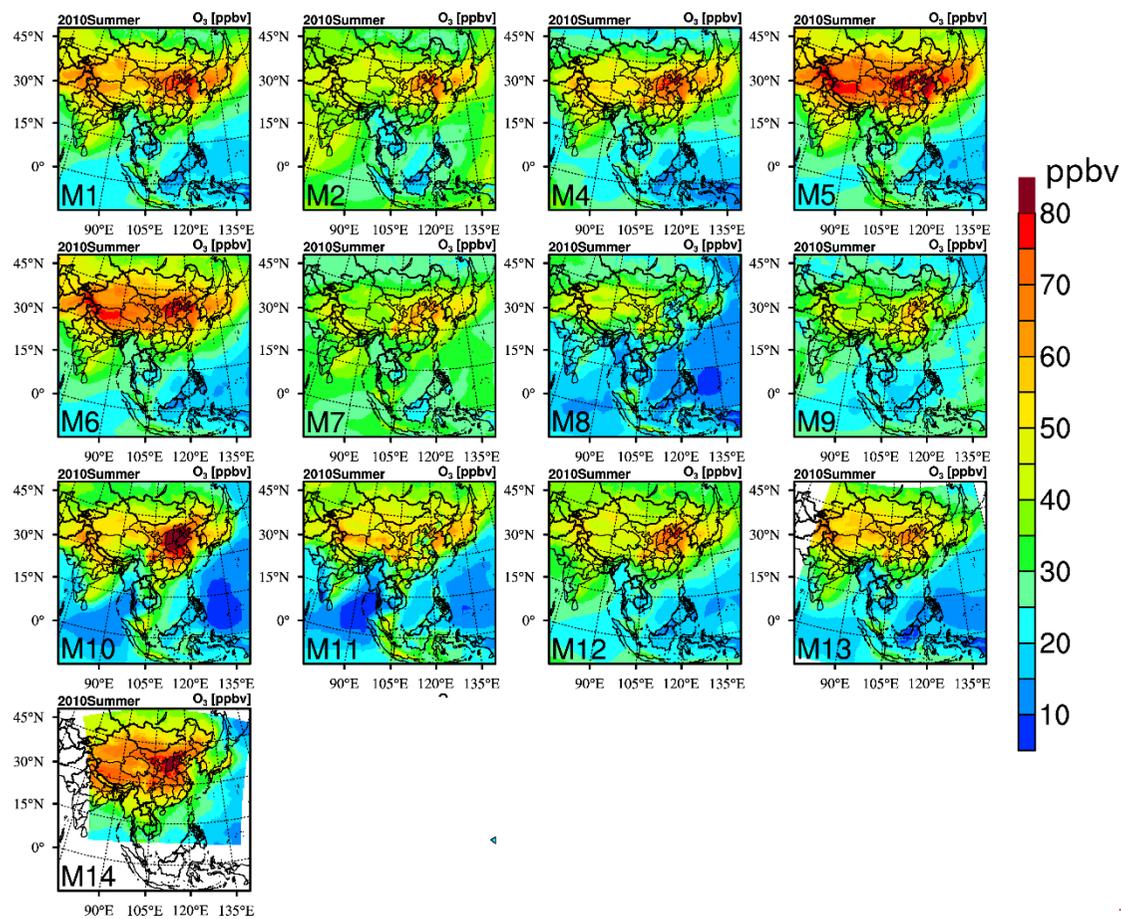
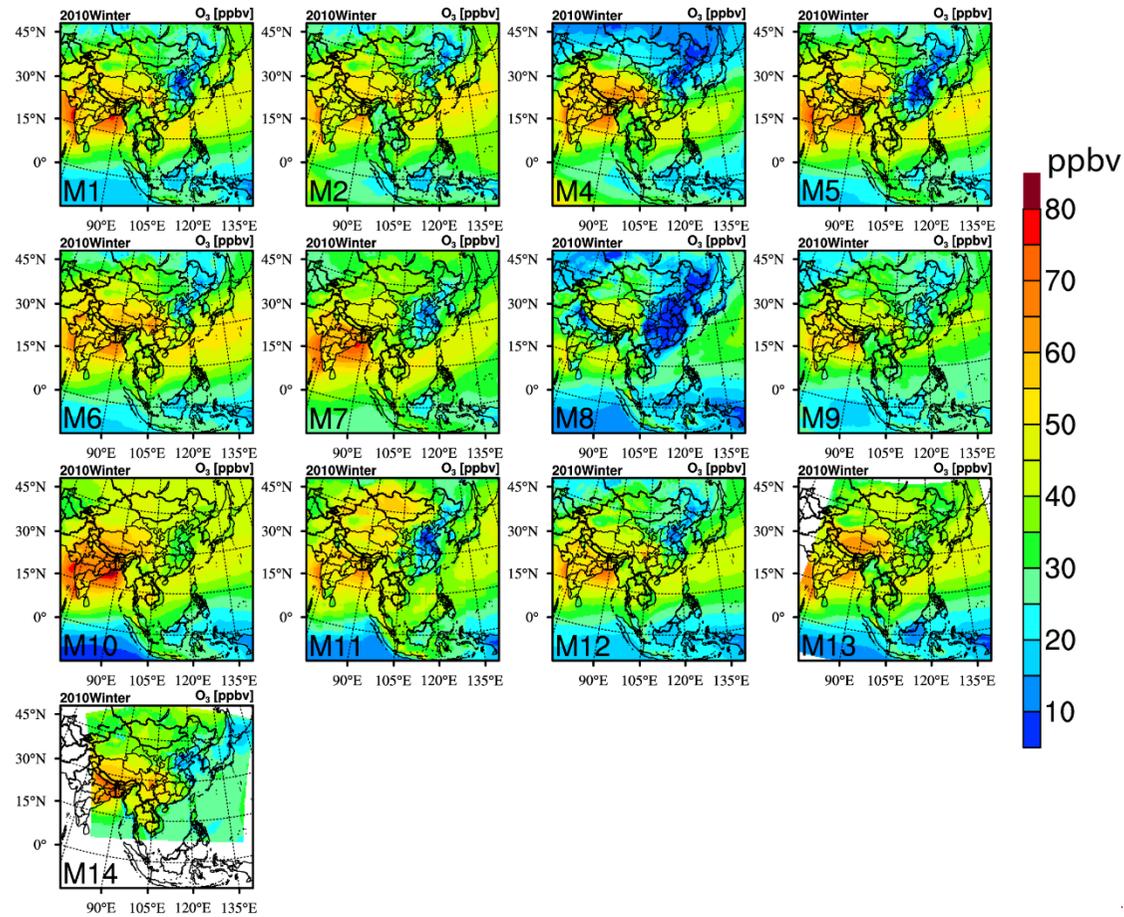


Fig.9 Li et al., 2018



— Fig.10 Li et al.,2018



—Fig.11 Li et al.,2018

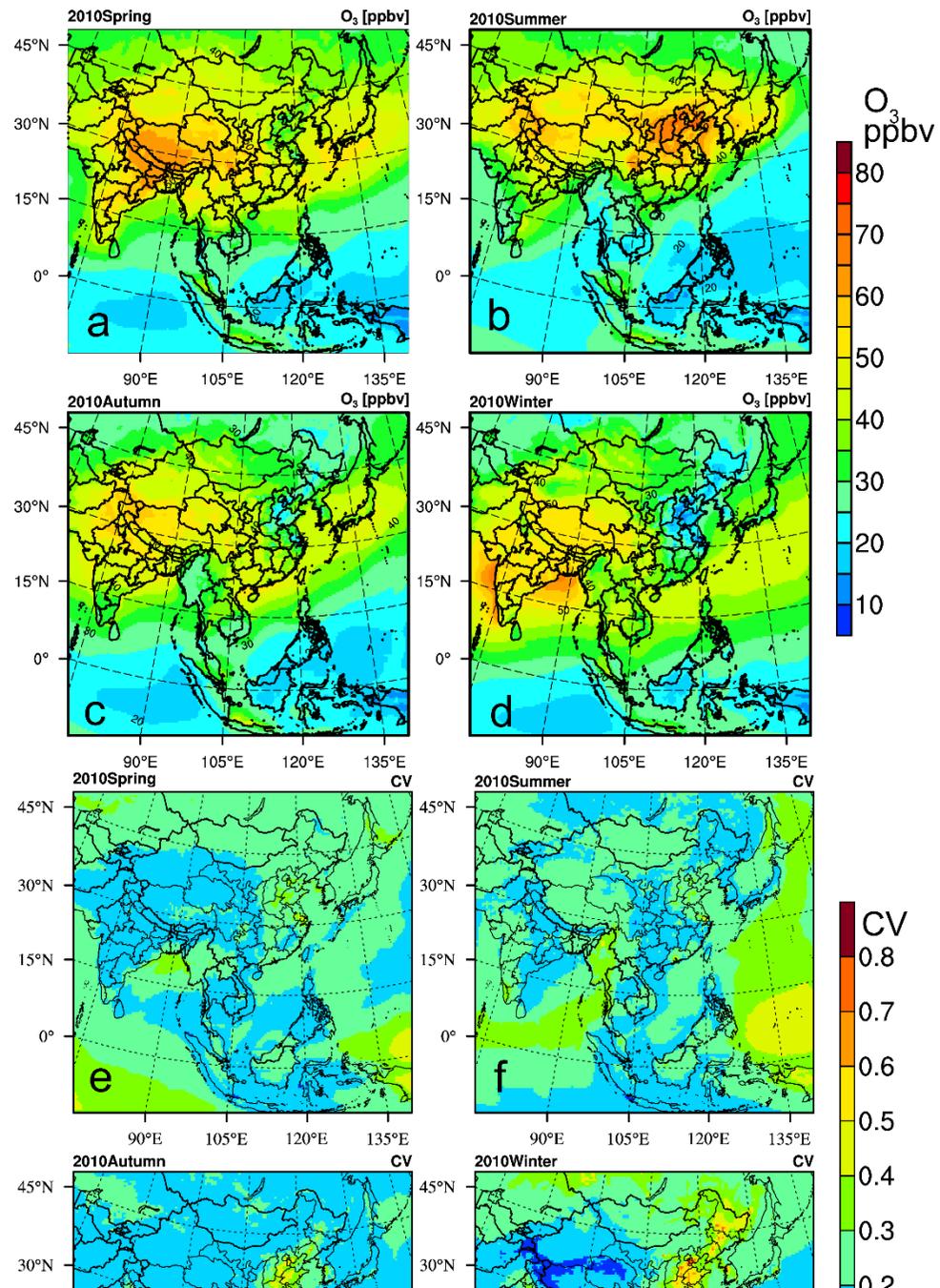


Fig.12 Li et al., 2018

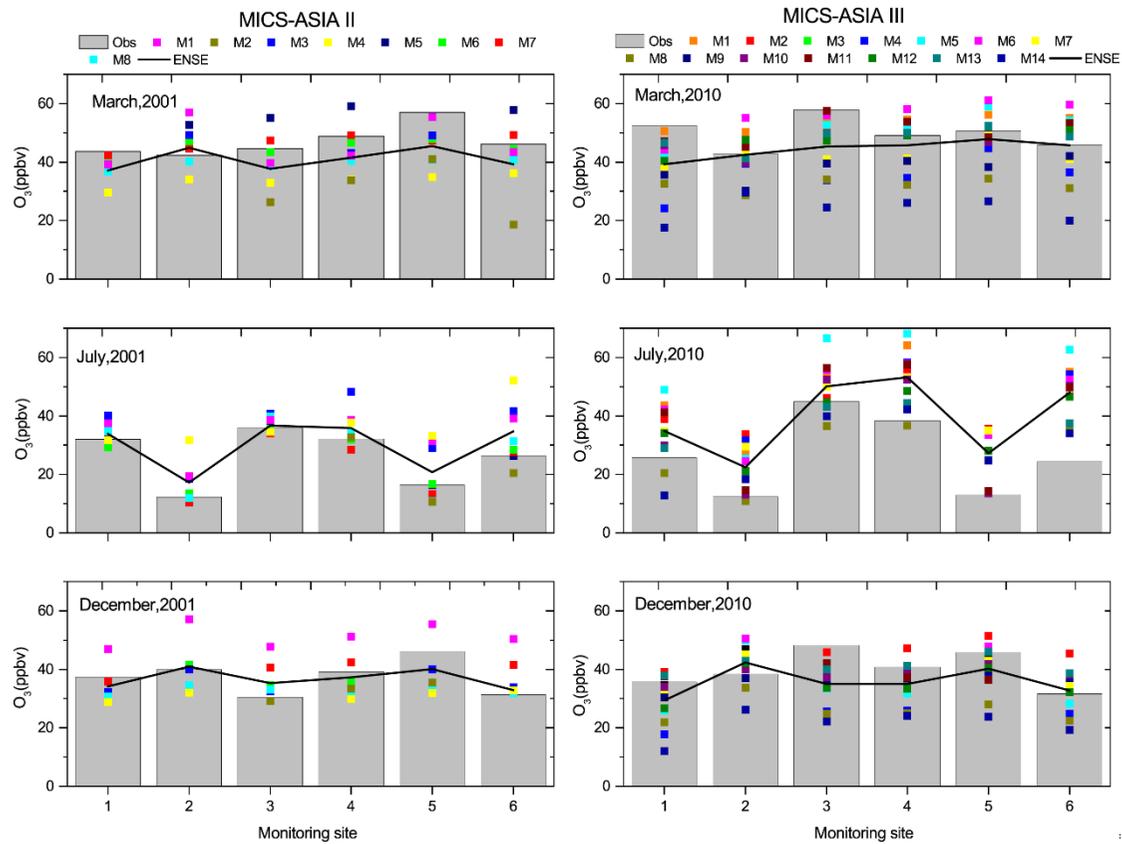


Fig.13 Li et al., 2018