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 O_3, NO and NO_2, surface O_3 diurnal variation, and O_3 vertical profiles); 3. Spatial distribution of O_3 and its comparison with MICS-Asia II, 4. Discussion (comparison with observed dry velocity and boundary layer height, relationships between O_3 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 NOx 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 R1site 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 μg/m^3

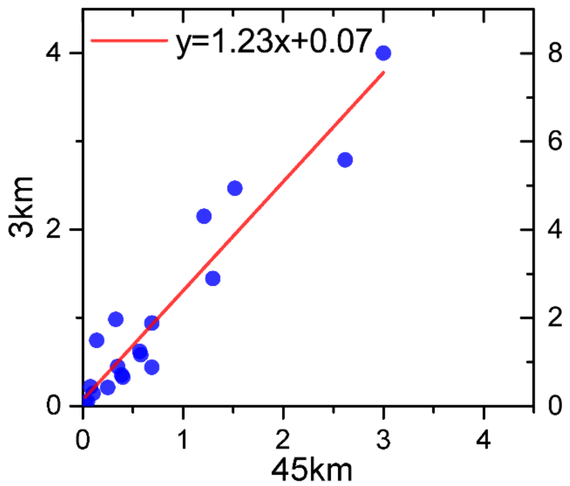


Fig.R1 Scatter plots of NO emission rates (μg/m^2/s) at observation sites in EA1 in 45km and 3km resolution emission inventory (MEIC)

• 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 NO_2 and O_3 in the PRD-RAQMN Network. Concentrations of pollutants were smooth. The effect of very local emissions was rarely seen.

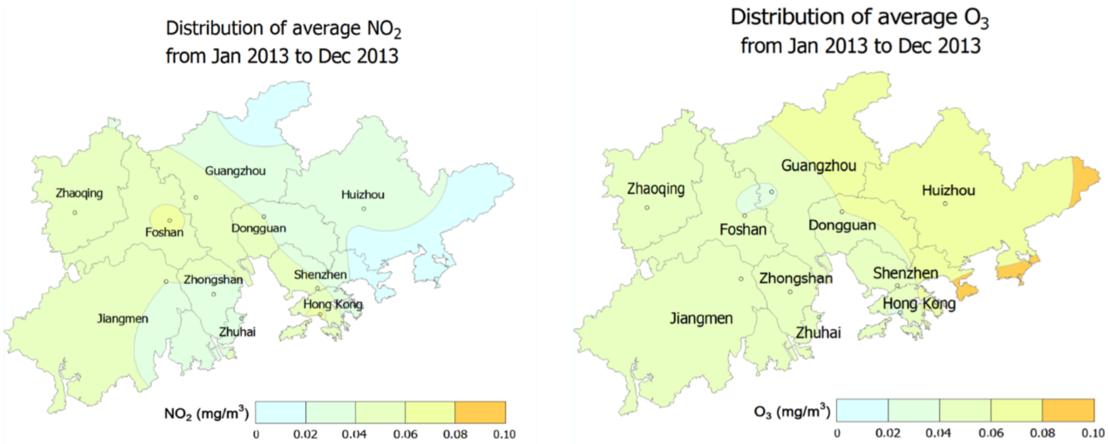


Fig.R2 Spatial distribution of average concentrations of NO_2 and O_3 in the PRD-RAQMN Network, figure is annual report of Pearl River Delta Regional Air Quality Monitoring Network in 2013 (https://www.epd.gov.hk/epd/sites/default/files//epd/english/resources_pub/publications/files/PRD_2013_report_en.pdf)

• 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).

Comment 3. About the investigation of intermodel variability on O3ã˘A˘A(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 O_3" to "Discussion".

As shown in Fig. R3, ensemble average dry deposition velocity of O_3 underestimated observations in August-September by 30-50% in EA1. This underestimation decreased the deposition amounts of surface O_3 and partly explained the overestimation of ensemble simulated O_3 in summer. This is consistent with intermodel comparison between M11 with M1-M6. M11 reproduced observed surface O_3 in EA1in May-July. The higher dry deposition velocities in M11 between May-July (0.3 cm/s) contributed to low surface O_3 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 O_3 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.

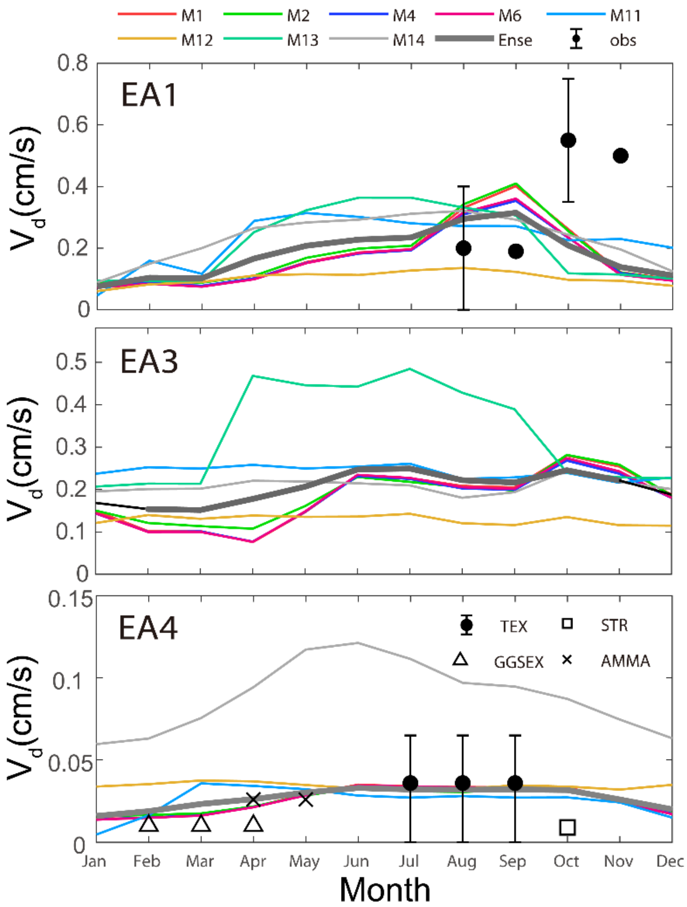
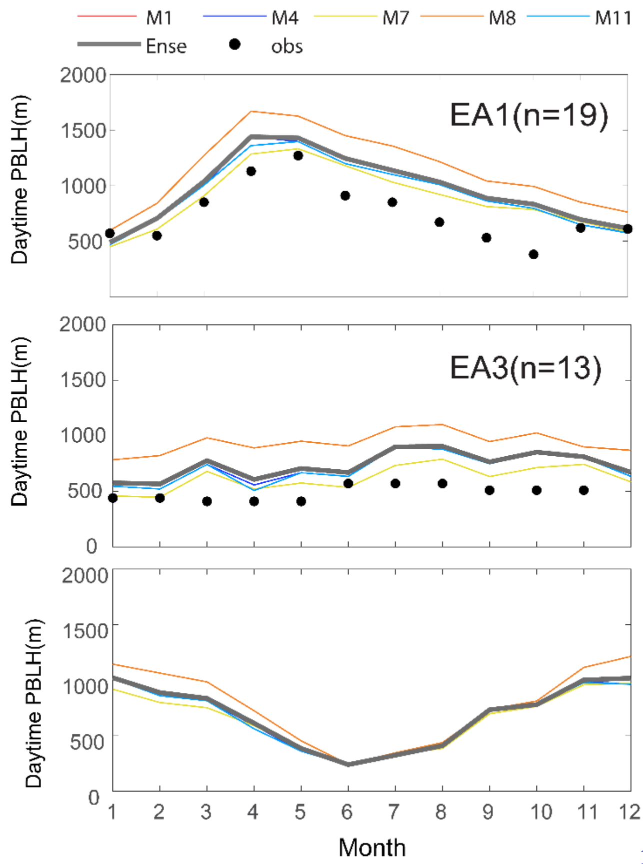


Fig. R3 Simulated and observed monthly dry deposition velocity

In the revised manuscript, we moved vertical profile of O_3 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 O_3 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.



now.

Fig. R4 Simulated and observed monthly daytime PBLH

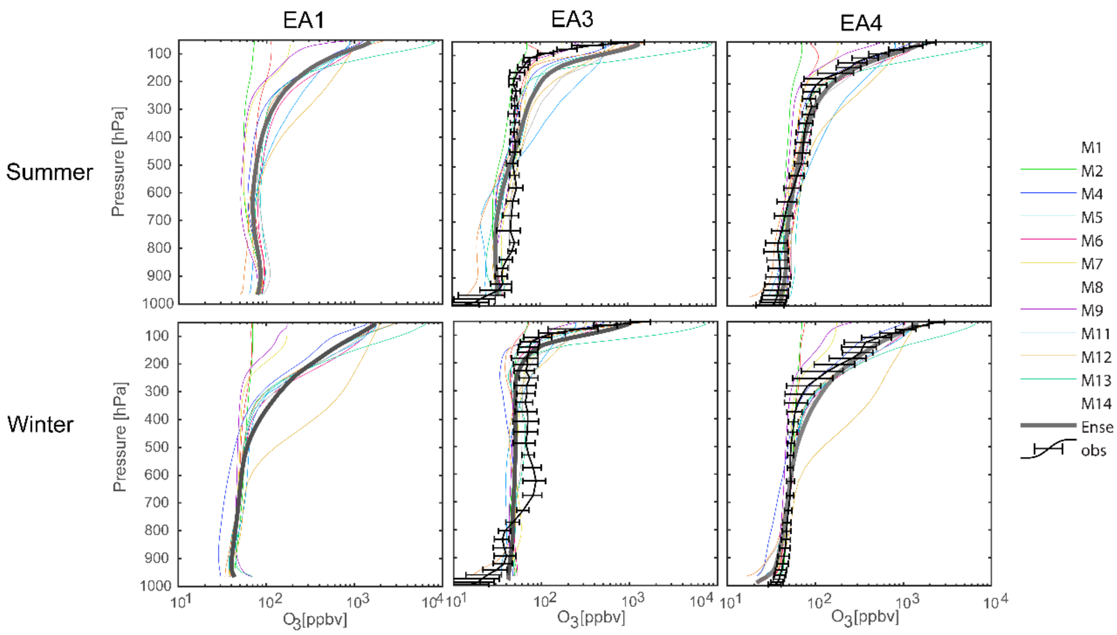


Fig. R5 Simulated and observed O_3 profiles in summer and winter of 2010, averaged over all observed stations in three subregions over East Asia (EA1: left column, EA3: middle column, EA4: bottom column).

The evaluation on chemistry in models is a difficult problem all along. As far as we know, there are no direct measurement on ozone production rates in East China till now. The relationships between O_3 with its precursors usually was regarded as an effective index on chemistry. We realized that the simple comparison between O_3 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 O_3 and NO_x in M11 (the best performance of O_3 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.

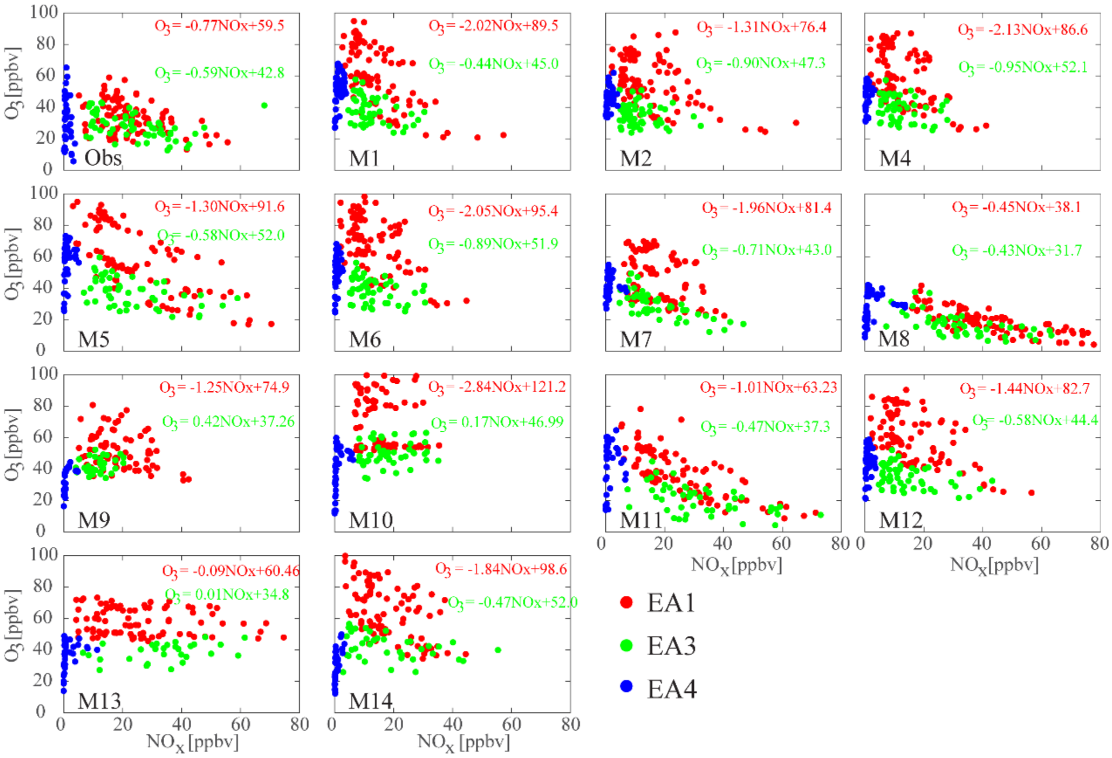


Fig. R9 Scatter plots between monthly daytime (08:00-20:00) surface NO_x and O_3 at each station over EA1(red), EA3(green)and EA4(blue) in May-October, for observations(obs) and models

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.

Comment 6: p.5 L10-11You 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.

In the revised manuscript, we added related descriptions.

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".

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

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..

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.

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

Reply: We deleted it.

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

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

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 O_3 concentrations in March-May and September-November. Observed O3 showed that the highest concentrations appeared in October-November."

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 O_3 was in May-October (L22 P9). The periods in P4 and P9 are different.

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 $m^2\ s^{-1}$. This value is a little higher than other models (e.g. CAMx: 0.1 $m^2\ s^{-1}$). 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.

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

Comment 17: p.11 L17Add 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.

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. 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 NOx 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.

Comment 20: p.16 L23-L24Jin et al (2015) perhaps showed the ozone formation regime at 1330 LST (overpasstime of OMI) while you show that between 1000-1800 LST. Also, your results include NOx 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. In the revised manuscript, we deleted this reference.

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

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

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 O_3 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.

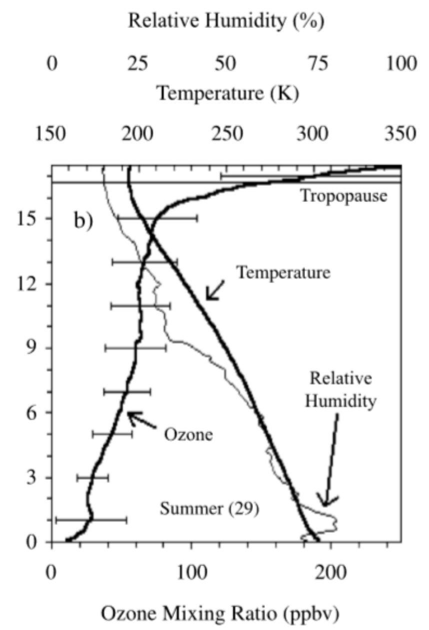


Fig. R10 Seasonally averaged ozone profiles in the troposphere above Hong Kong summer

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. R11 in the revised manuscript.

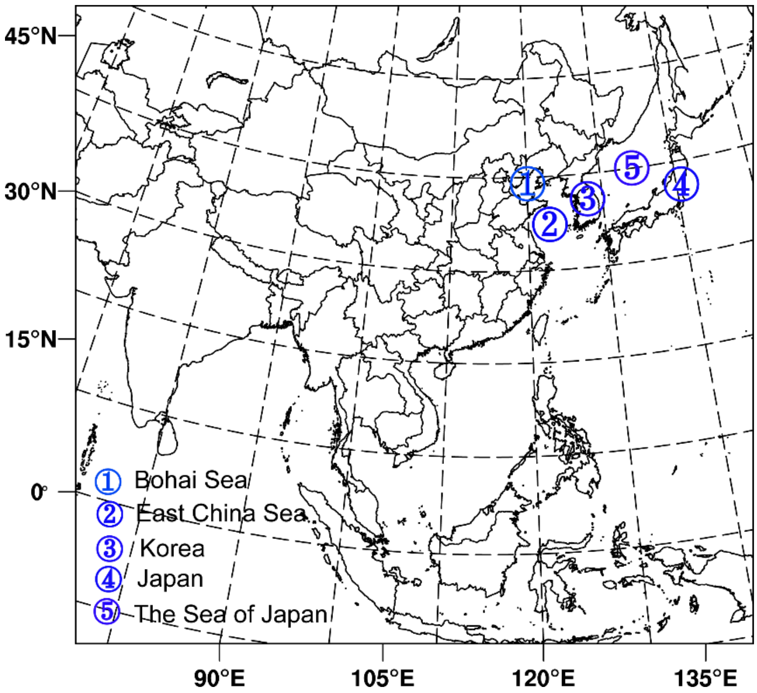


Fig. R11 Locations of related regions

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 O_3 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 O_3 between these three models the western Pacific Ocean showed a O_3 discrepancy. So, there could be other causes responsible for this discrepancy, besides emissions in source regions.

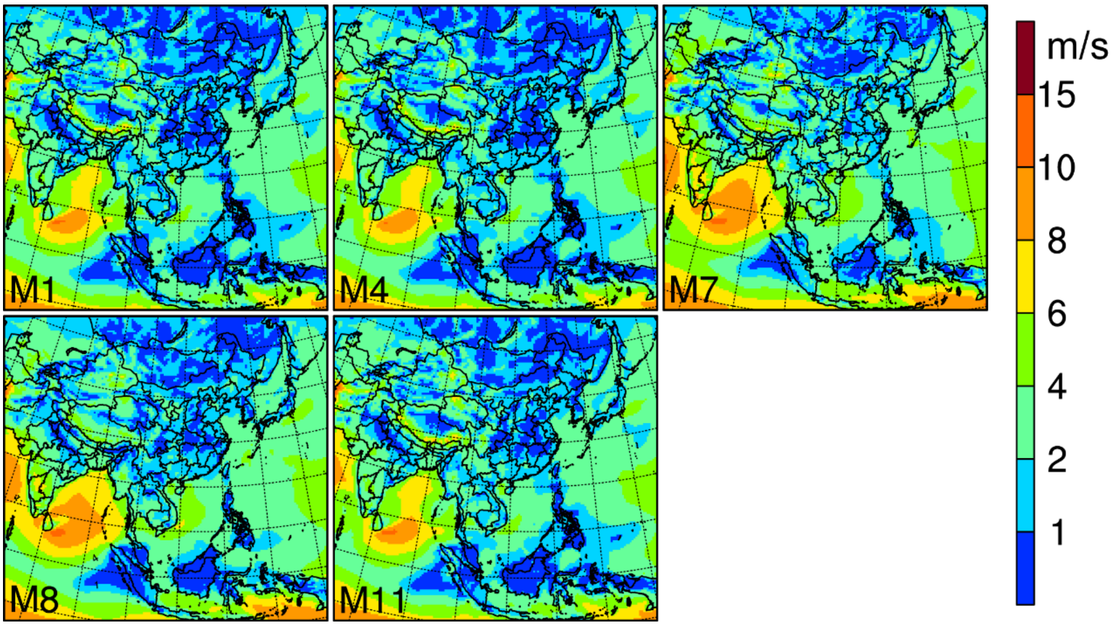


Fig. simulated surface wind velocities(m/s) in MICS-Asia III

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.

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.

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 O_3 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 O_3 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.

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

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

Reply: We revised it

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

Reply: We revised it

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

Reply: We revised 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., Engardtg, M., Hozumib, Y., Parkh, S.U., Kajinoi, M., Sarteletj, K., Fungk, 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