We thank the reviewer for his/her constructive comments.
Response to the Specific comments.

Comment 1: What does the vertical axis on the right of Fig. R1 mean?
If it represents a 3 km mesh emission, contrary to what is stated in the response document or in the revised manuscript, it is likely that the observation points receive local emissions.

Reply: Before the intercomparison between emission rates in 45 km and 3 km emission inventory, we provide the equation how to calculate the emission rates (μg/m^2/s) in 45 km resolution (E_{45km}) from those in 3 km resolution (E_{3km}).

$$E_{45km} = \sum_{i=1}^{m=15} \sum_{j=1}^{n=15} (E_{3km\,ij} / (m \times n)) \quad (1)$$

As shown in Eq (1), some $E_{3km\,ij}$ were higher than E_{45km}. Others could be less than E_{45km}. In general, there could be local emissions in ith and jth grid when E_{3km} in this grid is much higher than E_{45km}. When E_{3km} is much less than E_{45km}, the ith and jth grid represented a regional background condition. If E_{3km} is close to E_{45km}, the emission rates in the ith and jth grid represented the averaged conditions of 45 km × 45 km areas. So, the comparison of E_{3km} with E_{45km} is a good way to examine if this 3km grid receives local emissions.

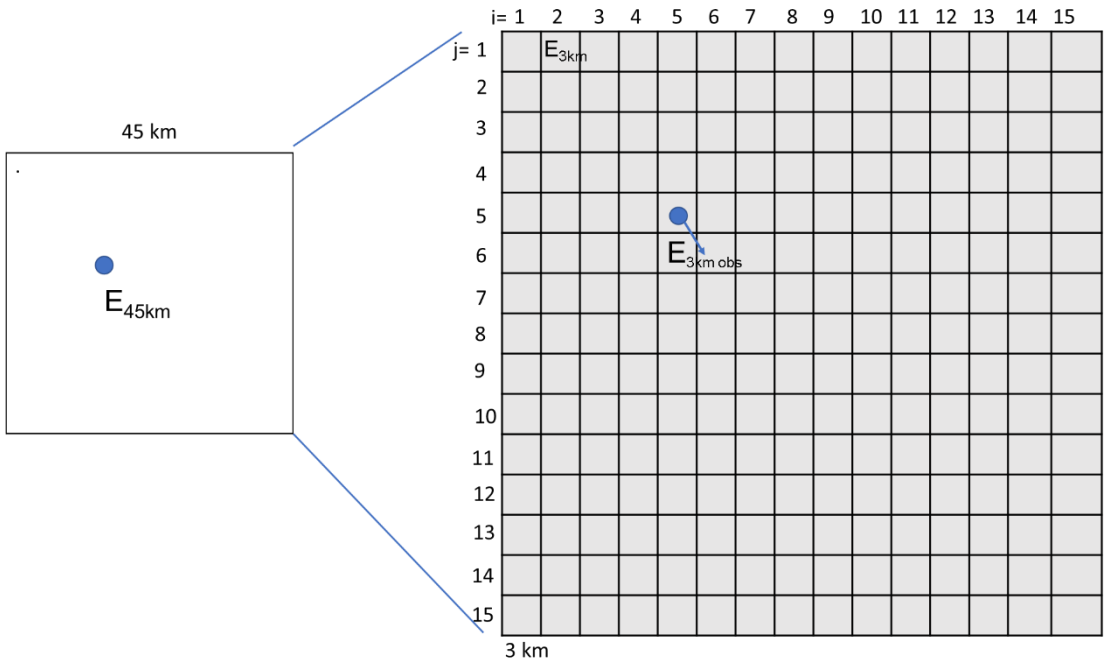


Fig. R1 The grids of observations in 45 (left) and 3 km (right) emissions inventory. The solid cycle represents the location of observation site.

In this study, we calculated the NO emission rates of observation sites in 45 km grid (E_{45km}). Emission rates (E_{3km}) at each 3 km grid (i=1,15; j=1,15) within this 45km grid was compared with E_{45km}. Meanwhile, Emission rates ($E_{3km\ obs}$) at this 3 km grid of

observation site was shown. Fig. R2 and R3 showed the comparison of E_{45km}, E_{3km} and $E_{3km\ obs}$ for NO and C_2H_4 emission rates. Clearly, $E_{3km\ obs}$ at the observation sites were close to E_{45km}, which indicated that these observation sites rarely receive local emissions.

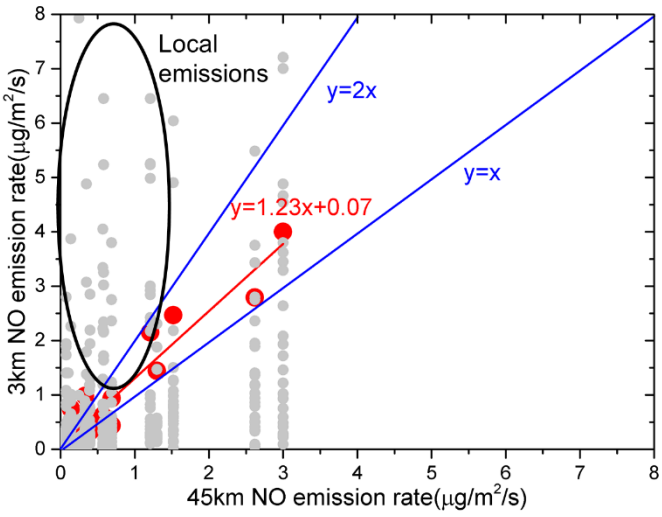


Fig. R2 Scatter plots of E_{45km}, E_{3km} and $E_{3km\ obs}$ for NO emission rates (μg/m²/s). The gray and red solid cycles represented the relationships of E_{45km}-E_{3km} and E_{45km}- $E_{3km\ obs.}$ The gray solid cycles above the line y=2x represented the grids receive the local emissions.

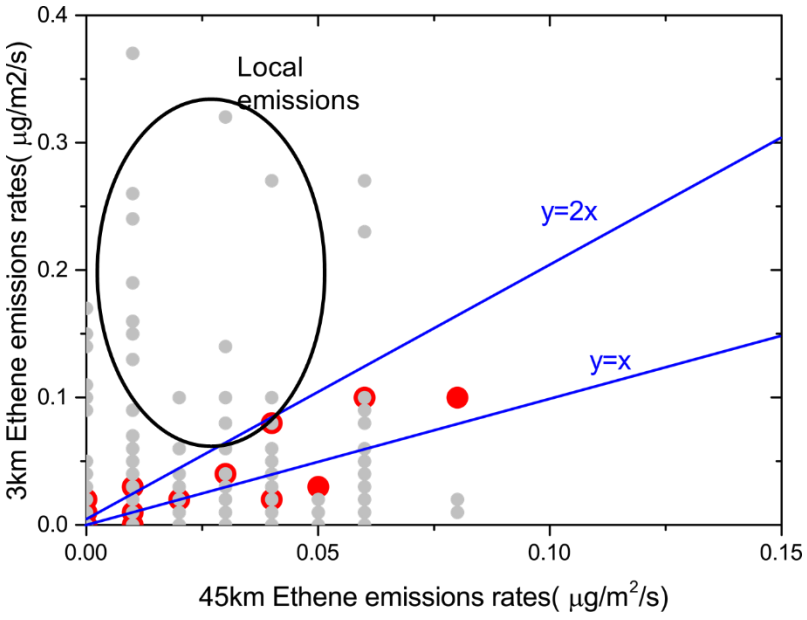


Fig. R3 The same as Fig.2, but for ethene

The following is a comment on the revised manuscript other than the above.

Comment 2： Page 11 Line 11: It describes the measuring instruments, but which measuring stations adopt these measuring instruments?

Reply: In this study, NO_x was measured by Thermo Scientific 42C NO-NO_2-NO_x

Analyzer with chemiluminescence technology at 40 sites in all three network (CERN, PRD-RAQMN). O_3 were measured by Thermo Scientific 49i with UV photometric technology from USA in CERN network and by Thermo Scientific 49C in PRD-RAQMN and EANET network.

Changes in the revised manuscript: Page 8 Line 14-17.

Comment 3：<About the response of my Comment 3>
It is said that the ensemble of models underestimates the measured values of dry deposition velocity from August to September in EA1, but according to Figure R3 it is actually overestimated.
Therefore, the claims in the response document or in the revised manuscript that underestimation of dry deposition velocity contributed to overestimation of summer ozone concentrations are false and need to be reviewed.
Reply: We agree. In the revised manuscript, we reworded related discussions and conclusions.
"In EA1, ensemble mean values overestimated observed dry deposition velocities of O_3 (vd) in August-September, but still fell into the range of observed standard deviation. This indicated that there must be other factors rather than dry deposition playing important roles in the overestimation of August-September O_3 in EA1. In October-November, simulated vd apparently underestimated observations by 30-50%. Among models, the lower dry deposition velocities in May-July from M1, M2, M4 and M6 than that of M11 partly explained higher May-July 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)."

Changes in the revised manuscript: Page 17 Line 14-23.

The following are comments on the revised manuscript other than the above.

Comment 4：Section 3.5: The sentences are difficult to understand. I think that English native check is necessary first.
Reply: We invited an English native speaker to check our manuscript through AJE company.

Comment 5：Page 16 Line 4-5: I do not understand the meaning of this sentence.
Reply: We revised to ". Quantifying the contributions of these processes can help explain model biases through sensitivity simulations"

Changes in the revised manuscript: Page 16 Line 24-25.

Comment 6：Page 16 Line 27 "summer": A statement from May to July, not summer,

is more logical.
Reply: We agree and revised it.

Comment 7： Page 17 Line 3-4: According to Fig.9, contrary to the description in the revised manuscript, the model underestimated the observed values of Vd.
Reply: We agree and corrected it.
Changes in the revised manuscript: Page 16 Line 24-25.

Comment 8： Page 18 Line 7-8: The description in this sentence should indicate that it is a description of the observed value.
Reply: We agree. "The slope and intercept of regression line between observed O_3 and NOx were -0.77 ppbv/ppbv and 59.5 ppbv, respectively"

Changes in the revised manuscript: Page 18 Line 27.

Comment 9： Page 18 Line 16: What are "previous theoretical results"?
Reply: According to comments by co-editor, we deleted these sentences.

Comment 10： Page 19 Line 15-25: At the beginning of chapter 5, you have mentioned that you evaluated dry deposition, PBL, and chemistry.
I do not understand why the discussion about the meteorological field comes out suddenly here.
Reply: In the revised manuscript, we remove this discussion.

Comment 11： <About the response of my Comment 6>

Page 5 Line 9-10: This sentence is about how to validate the model,
so this sentence should be moved to chapter 3 which discusses it.
Reply: We moved it to section 3.2.

Changes in the revised manuscript: Page 9 Line 25.

Comment 12： <About the response of my Comment 9>

Page 7 Line 2-6: Make a description that shows the correspondence between the model and the university.
Reply: We agree. "GEOS-Chem was run with a 2.5º×2º horizontal resolution and 47 vertical layers by University of Tennessee and Chemical AGCM for Study of Atmospheric Environment and Radiative Forcing (CHASER) was run with a 2.8º× 2.8º

horizontal resolution and 32 vertical layers by and Nagoya University."

Changes in the revised manuscript: Page 7 Line 6-8.

Comment 13：<About the response of my Comment 17>

It is understood from Appendix A that various statistics are derived taking into account only the variation due to the location (i) of the measuring station, but in fact it seems that temporal variations are also taken into consideration.
This is because, for example, in Page 11 Line 22-24, as the reason for the high correlation, the high reproducibility of the monthly variation is mentioned.
Reply: We totally agree. We added the variation due to temporal variations in Appendix A.

Comment 14：<About the response of my Comment 27>

I understood that all models must be identified in Fig.2 and Fig.3.
However, since it is still difficult to identify models by color, I hope that at least models that appear in the discussion in chapter 5 can be identified by such as drawn lines.
Reply: We agree. In Fig. 3 and 4 in the revised manuscript, we only plotted models in appear in the discussion in chapter 5 by colors. The others are plotted in gray lines.

<Other technical corrections to the revised manuscript>

As there are many necessary technical corrections, I would like you to carefully review the entire manuscript. .
Some of them are shown below.

Comment 15：Page 2 Line 2: "Evaluated and intercompared to O3 observations" perhaps should be "intercompared and evaluated to O3 observations".
Reply: We revised it.

Comment 16：page 2 Line 7-8: "western pacific rim" is repeated twice.
Reply: We deleted the second one.

Comment 17：Page 8 Line 25: "Table1" should be "Table2".
Reply: We revised it.

Comment 18：Page 11 Line 27: "Table1" should be "Table2".

Reply: We revised it.

Comment 19： page 14 line 27: The word "other" should be added before "combined influence".
Reply: We revised it.

Comment 20： Page 17 Line 24 "Sillman and He et al.": "Et al" should be removed.
Reply: We revised it.

Comment 21： Page 20 Line 10-11: There is an incomplete sentence.
Some of figures: "EA3" should be replaced by "EA2", and "EA4" should be replaced by "EA3".
Reply: We revised it. “For the North China Plain and western Pacific Rim, the model ensemble severely overestimated surface O3 levels for May-September by 10-30 ppbv.”

Changes in the revised manuscript: Page 23-25

We thank co-editor for his constructive comments.
Response to the Specific comments.

General comment: Although the manuscript is improved, still many fundamental and technical points are present needing revision. Please examine comments from the reviewer #1 carefully, and the following comments by the Co-Editor. Recheck English throughout the manuscript with a native speaker.
Reply: Thanks a lot for insightful comments. In the revised manuscript, we invited a native speaker by AJE company to improve our language in this manuscript.

(pages and lines are for the change-track version)

Comment 1： Page 2, line 18: In terms of
Reply: We revised it.

Comment 2： Page 2, line 20: hPa
Reply: We revised it.

Comment 3： Page 3, line 21: do not
Reply: We revised it.

Comment 4： Page 5, lines 19-20: Better to place the following inserted sentence elsewhere (page 8-9?): If two or more observation sites were in the same grid of model, their mean values will be used to evaluate model performance.
Reply: We placed it into Page 9 Line 25.

Comment 5： Page 6, line 18. possibly
Reply: We revised it.

Comment 6： Page 8, lines 18-20: More explanation and clarification about the local sources and sinks are necessary, as requested by the reviewer. Not only for NOx but also VOC, representativeness is necessary.
Reply: We agree. In the revised manuscript, we plotted NO and VOCs emission rates.

Before the intercomparison between emission rates in 45 km and 3 km emission inventory, we provide the equation how to calculate the emission rates (μg/m^2/s) in 45 km resolution (E_{45km}) from those in 3 km resolution (E_{3km}).

$$E_{45km} = \sum_{i=1}^{m=15} \sum_{j=1}^{n=15} (E_{3km\,ij} / (m \times n)) \quad (1)$$

As shown in Eq (1), some $E_{3km\,ij}$ were higher than E_{45km}. Others could be less than E_{45km}. In general, there could be local emissions in ith and jth grid when E_{3km} in this grid is much higher than E_{45km}. When E_{3km} is much less than E_{45km}, the ith and jth grid represented a regional background condition. If E_{3km} is close to E_{45km}, the emission

rates in the i^{th} and j^{th} grid represented the averaged conditions of 45 km × 45 km areas. So, the comparison of E_{3km} with E_{45km} is a good way to examine if this 3km grid receives local emissions.

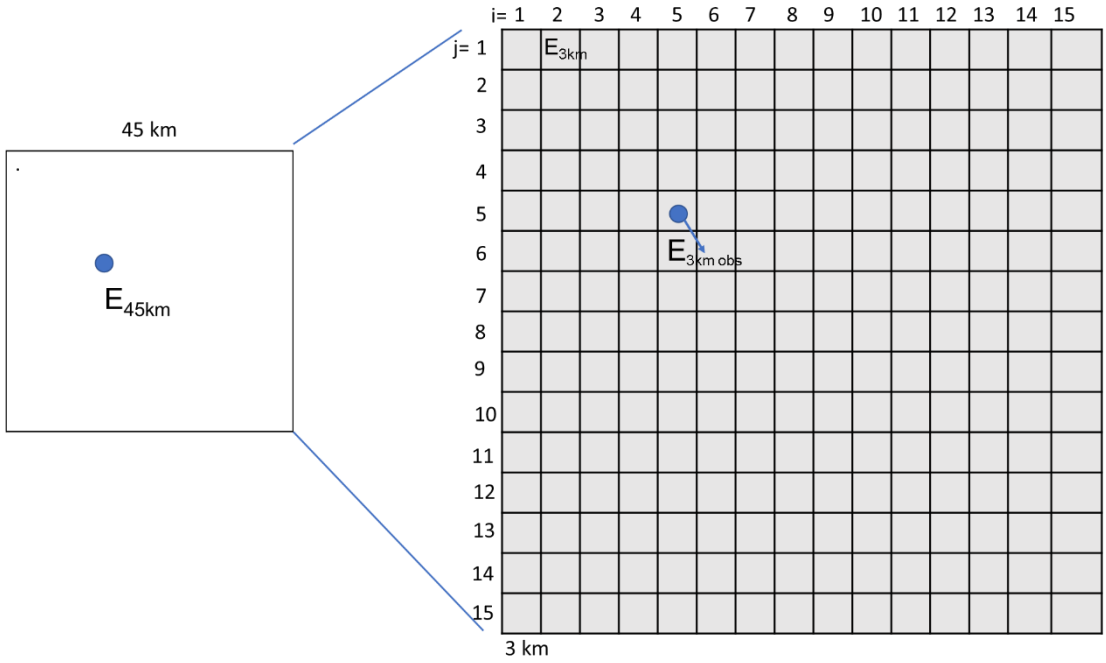


Fig. R1 The grids of observations in 45 (left) and 3 km (right) emissions inventory. The solid cycle represents the location of observation site.

In this study, we calculated the NO emission rates of observation sites in 45 km grid (E_{45km}). Emission rates (E_{3km}) at each 3 km grid (i=1,15; j=1,15) within this 45km grid was compared with E_{45km}. Meanwhile, Emission rates ($E_{3km\ obs}$) at this 3 km grid of observation site was shown. Fig. R2 and R3 showed the comparison of E_{45km}, E_{3km} and $E_{3km\ obs}$ for NO and C_2H_4 emission rates. Clearly, $E_{3km\ obs}$ at the observation sites were close to E_{45km}, which indicated that these observation sites rarely receive local emissions.

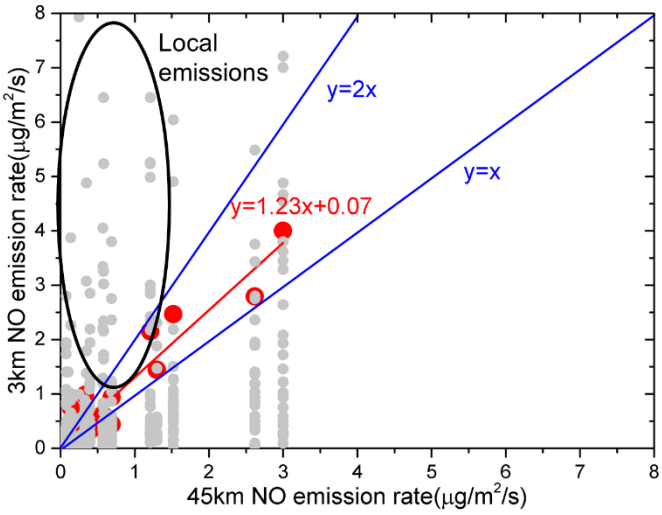


Fig. R2 Scatter plots of E_{45km}, E_{3km} and $E_{3km\ obs}$ for NO emission rates (μg/m²/s). The gray and red solid cycles represented the relationships of E_{45km}-E_{3km} and E_{45km}- $E_{3km\ obs.}$ The gray solid cycles above the line y=2x represented the grids receive the local emissions.

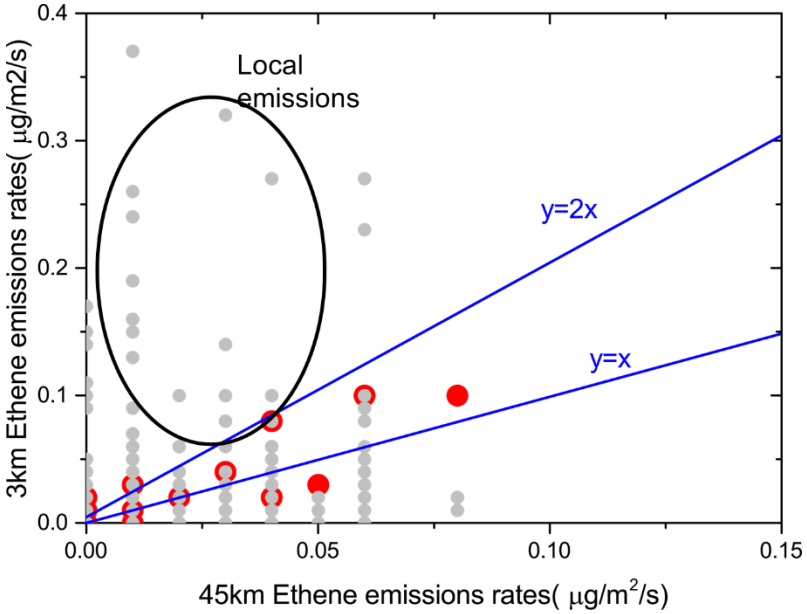


Fig. R3 The same as Fig.2, but for Ethene

Comment 8: Page 8, line 28: exhibited
Reply: We revised it.

Comment 9 Page 9, line 1: Molybdenum converters
Reply: We revised it.

Comment 10： Page 9, lines 2-5. The results must be dependent on the conditions (with different NOz/NOx). Where did the comparison take place? Uncertainty should be estimated to cover all possible conditions (urban to rural) used for the study.

Reply: We agree. In this study, the comparison was conducted in an urban site in Beijing from 1 to 29 August 2012. As said by the editor, the difference between NO_2 and NO_2* is dependent on the conditions (Fig. R1). In general, their differences appeared within 10-15% when NO_2*>15 ppbv and NOz<10 ppbv in Beijing. In low NO_2 (<15 ppbv) and high NOz (>10 ppbv) conditions, the NO_2^* usually was higher than NO_2 (10-30%). Unfortunately, there are very few CAPS NO_2 measurements in China. Jung reported their NO_2 and NO_2* measurements in a rural site in South Korea (Fig. R2). Similar as Beijing, NO_2 and NO_2* usually appeared within 20% when NO_2^* >20 ppbv in Spring and Summer. NO_2* overestimated NO_2 by 20-40% when NO_2^* <20 ppbv. In Fall and Winter, NO_2* usually overestimated NO_2 by 10-20 ppbv in all conditions.

In the revised manuscript, we added a short discussion on NO_2* uncertainty.

"This bias was dependent on the chemical conditions. A one-month continuous measurement in August by a chemiluminescence analyzer and Aerodyne Cavity Attenuated Phase Shift Spectroscopy (CAPS) at an urban site in Beijing 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). Measurements at a rural site in South Korea also revealed a similar pattern (Jung et al., 2017). These comparisons suggested that observations by molybdenum converters may overestimated NO2 by 10-20% in EA1 and EA2, and 30% in EA3. This brings uncertainties for the model evaluation on NO_2 in this study. "

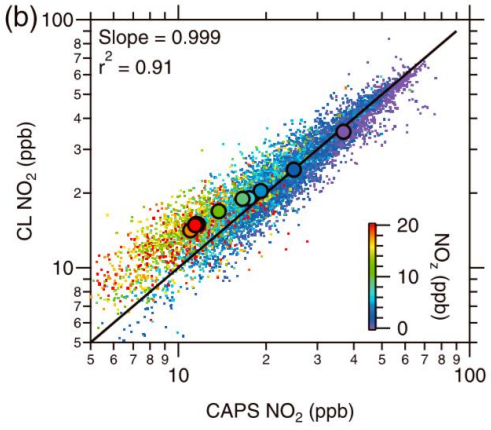


Fig. R1 Comparison of hourly NO_2 measured by the CL NO_x analyzer and the CAPS NO_2 monitor in Beijing

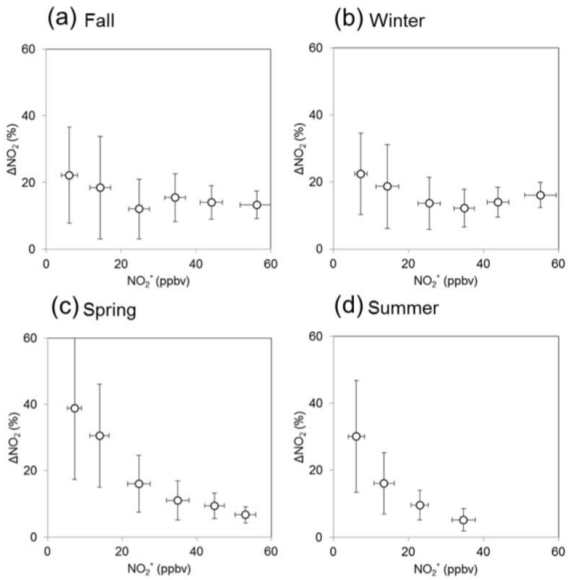


Fig. R2 The NO_2/NO_2* ratios as a function of NO_2 * concentrations during (a) fall, (b) winter, (c) spring, and (d) summer in a rural site in South Korea (Jung et al., 2017). NO_2 and NO_2 * is the measurements by the photolytic converterver and molybdenum converter。

Changes in the revised manuscript: Page 8 Line 20-27.

Comment 11：Page 9, line 16: parts per billion by volume
Reply: We revised it.

Comment 12：Page 9, line 19: and M7
Reply: We revised it.

Comment 13：Page 9, line 20: closer to
Reply: We revised it.

Comment 14：Page 9, line 20: M11 simulated O_3 with RMSEs of
Reply: We revised it.

Comment 15：Page 12, line 26: existed
Reply: We revised it.

Comment 16：Page 13, line 21. approximately 20 and 10 ppbv
Reply: We revised it.

Comment 17：Page 14, line 7: showed a better performance

Reply: We revised it.

Comment 18: Page 15, line 5: What is meant by short-lived species?
Reply: 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 (e.g. NO).

Comment 19: Page 15, line 25: Figure 5 shows
Reply: We revised it.

Comment 20: Page 15, lines 27-28 and elsewhere: hPa
Reply: We revised it.

Comment 21: Page 16, line 7: The mean standard deviations (SD) of models (1σ) Use SD instead of σ later.
Reply: We revised it.

Comment 22: Page 23, line 16. MICS
Reply: We revised it.

Comment 23: Page 27, lines 23. Ensemble mean values rather OVERESTIMATES vd (Fig. 9). Discussion needs to be revised (conclusion also). Representativeness of deposition velocity observations over the domains needs to be discussed. I doubt that grassland represents the whole domain region.
Reply: We agree. We reworded discussions (and conclusions) on dry depositions in the revised manuscript.
"In EA1, ensemble mean values overestimated observed dry deposition velocities of O3 (vd) in August-September, but still fell into the range of observed standard deviation. This indicated that there must be other factors rather than dry deposition playing important roles in the overestimation of August-September O_3 in EA1. In October-November, simulated vd apparently underestimated observations by 30-50%. Among models, the lower dry deposition velocities in May-July from M1, M2, M4 and M6 than that of M11 partly explained higher May-July 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."

In this study, we selected observations at two sites in EA1. One site was located in a valley of Mangshan Forest Park, the other (CREAS) was in a suburb about 10km north of the Beijing city. The CREAS site had an area of 200m×200m, and was thickly covered with short grass about 10cm high (Dense grassland). Fig.R3 presents the land cover classification map in EA1 from MODIS satellite data (Zhang et al., 2008). It can be found that dense grassland is one of the dominated landcover classes and covers ~20% land area. Another dominated landuse in EA1 is crop class, which covers more

than 50% land area. Unfortunately, there are few observations on O_3 dry deposition on the crop class in China. This may bring uncertainties for model evaluation. Hardacre et al. (2015) reported O_3 dry deposition measurements on crops in Europe and simulated O_3 dry deposition in 15 global models. Both observations and simulations showed that O_3 dry deposition velocities on agriculture crop class were quite similar as grasslands (Fig. R4). This indicated that the uncertainties on representativeness of measurement sites in this study did not affect our conclusions.

We added related discussions in the revised manuscript.

Changes in the revised manuscript: Page17 Line 25-Page 18 Line 2.

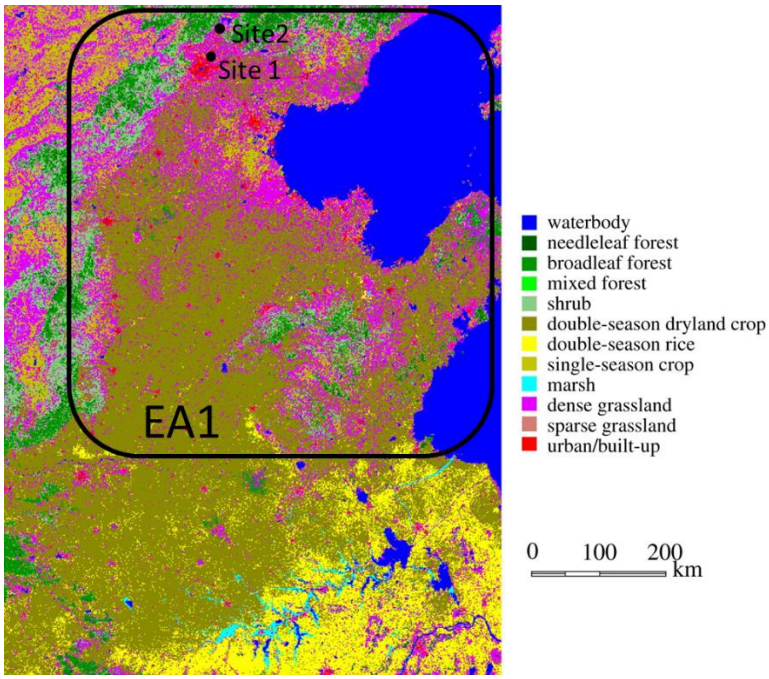


Fig. R3 The land cover classification map from MODIS satellite data

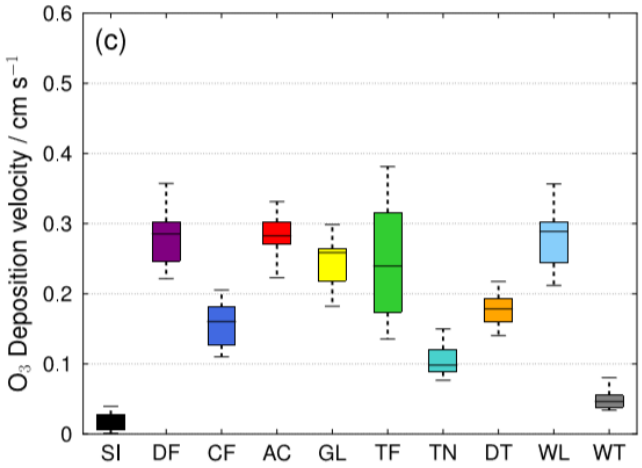


Fig. R4 Total annual O_3 dry deposition and annual average O_3 deposition velocity partitioned to land cover classes (GL: grassland; AC: agriculture crop class)

Comment 24： . Page 28, lines 3-4. October-November, simulated vd apparently overestimated observations by 30-50%. This is also opposite.

Reply: We agree. In the revised manuscript, we corrected it. "In October-November, simulated vd apparently underestimated observations by 30-50%."

Changes in the revised manuscript: Page 17 Line 18.

Comment 25： Page 29, line 6. Is this only because O3 is titrated with NO? The analysis never tells the regimes. Consider cases of O3 buildup and transport by consuming NOx. Even in the NOx limited regime, the NOx-O3 relationship constructed from geographical distribution will show a negative correlation. At least potential ozone (O_3 + NO_2) should be used for the analysis, and remove all statements on the regimes (particularly those in lines 22-23, page 29).

Reply: We agree. In the revised manuscript, we remove all statements on the regimes.

Fig. R5 shows the relationship between NO_x and $O_x(O_3+NO_2)$ in this study. For EA1, observed O_x increases with the increase of NO_x levels, with coefficient of determination (R^2) of 0.61. Most of the models (except for M8, M11 and M13) failed to reproduced observed positive correlations between O_x and NO_x, and their R^2 only ranged from 0.01-0.08. The slope, intercept and R^2 of M8 and M11 are relative agreement with observations. For EA2, all models reproduced observed key patterns in which O_x positively correlated with NO_x.

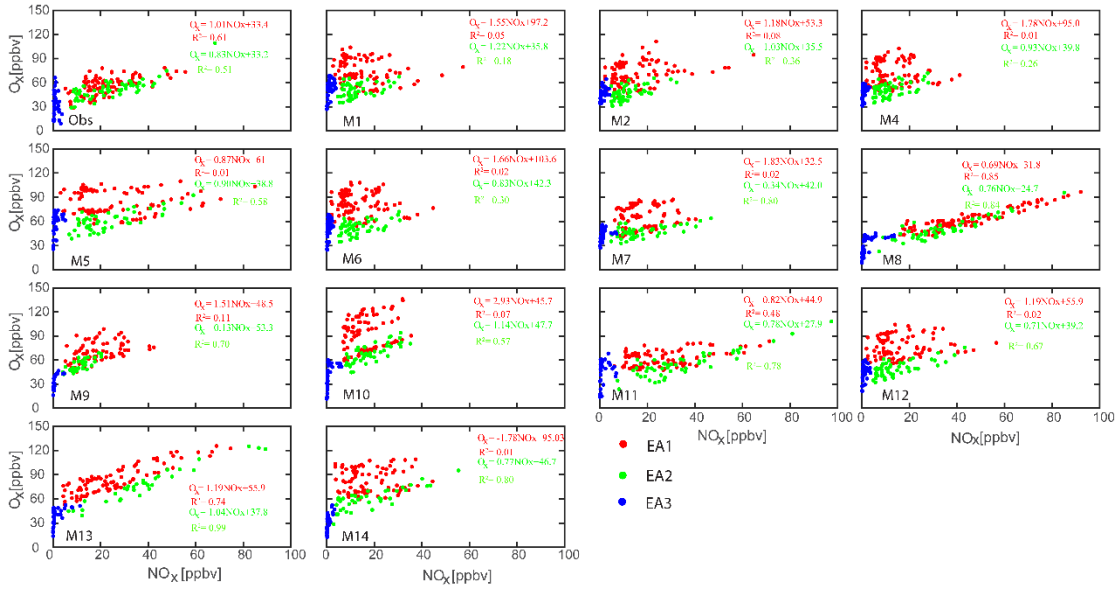


Fig. R5 Scatter plots for monthly daytime (08:00-20:00) surface NO_x and O_x for each station in EA1 (red), EA2 (green)and EA3 (blue) in May-October, for observations (obs) and models. Also shown are the linear regression equations and coefficient of determination (R^2) for NO_x and O_x (O_3+NO_2) in EA1 (red) and EA2 (green).

Changes in the revised manuscript: Page 19 Line 7-15.

Comment 26: Page 30, line 6. Briefly mention how PBLH was measured.
Reply: In this study, the observed PBLH was derived from the radiosonde network of the L-band sounding system of the China Meteorological Administration (CMA). The system provides fine-resolution profiles of temperature, pressure relative humidity, wind speed and direction. the bulk Richardson number (Ri) method (Vogelezang and Holtslag,1996) was taken to simultaneously estimate the PBLH from CMA soundings. R_i is defined as the ratio of turbulence associated with buoyancy to that induced by mechanical shear, which is expressed as

$$Ri(z) = \frac{(g/\theta_{vs})(\theta_{vz} - \theta_{vs})(z - z_s)}{(u_z - u_s)^2 + (v_z - u_s)^2 + (bu_*^2)},$$

where z denotes height above ground, s the surface, g the acceleration due to gravity, θv virtual potential temperature, u and v the component of wind speed and u∗ the surface friction velocity. u∗ can be ignored here due to the much smaller magnitude compared with bulk wind shear term in the denominator (Vogelezang and Holtslag, 1996). The critical value of 0.25 (Ri) is referred to as PBLH in this study, similar to the criteria used by Seidel et al. (2012).

Changes in the revised manuscript: Page 19 Line 27-Page 20 Line 5.

Comment 27: Page 30, line 8. caused by the sampling bias between
Reply: We revised it.

Comment 28: Page 31, line 9. A model ensemble was produced
Reply: We revised it.

Comment 29: Page 31, line 14. A period character must be a comma?
Reply: We revised it.
"In North China Plain and western Pacific rim, the model ensemble severely overestimated surface O_3 in May-September by 10-30 ppbv."

Comment 30: At many times figures contain EA1, 3, and 4, instead of EA1, 2, and 3.
Reply: We corrected it.

Jinsang Jung, JaeYong Lee, ByungMoon Kim, SangHyub Oh, Seasonal variations in the NO2 artifact from chemiluminescence measurements with a molybdenum converter at a suburban site in Korea (downwind of the Asian continental outflow) during 2015–2016, Atmospheric Environment, Volume 165,2017,Pages 290-300.

Seidel, D. J., Zhang, Y., Beljaars, A., Golaz, J.-C., Jacobson, A. R., and Medeiros, B.: Climatology of the planetary boundary layer over the continental United States and Europe, J.Geophys.Res.Atmos., 117, D17106, doi:10.1029/2012JD018143, 2012

Vogelezang, D. H. P. and Holtslag, A. A. M.: Evaluation and model impacts of alternative boundary-layer height for mulations, Bound.-Lay. Meteorol., 81, 245–

269, doi:10.1007/BF02430331, 1996.
Xia Zhang, Rui Sun, Bing Zhang, Qingxi Tong, Land cover classification of the North China Plain using MODIS_EVI time series, ISPRS Journal of Photogrammetry and Remote Sensing, Volume 63, Issue 4,2008,Pages 476-484.
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.

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 Nagashima[4], Lei Kong[1,2], Baozhu Ge[1,2,3], Kazuyo Yamaji[5], Joshua S. Fu[6], Xuemei Wang[7], Qi Fan[8], Syuichi Itahashi[9], Hyo-Jung Lee[10], Cheol-Hee Kim[10], Chuan-Yao Lin[11], Meigen Zhang[1,2,3], Zhining Tao[12], Mizuo Kajino[13,14], Hong Liao[15], Meng Li[16], Jung-Hun Woo[10], Jun-ichi Kurokawa[17], Qizhong Wu[18], Hajime Akimoto[4], Gregory R. Carmichael[19] and Zifa Wang[1,2,3]

[1]LAPC, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China
[2]College of Earth Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China
[3]Center for Excellence in Urban Atmospheric Environment, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen, 361021, China
[4]National Institute for Environmental Studies, Onogawa, Tsukuba, 305-8506, Japan
[5]Graduate School of Maritime Sciences, Kobe University, Kobe, 657-8501, Japan
[6]Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN, 37996, USA
[7]Institute for Environment and Climate Research, Jinan University, Guangzhou, 510632, China
[8]School of Atmospheric Sciences, Sun Yat-sen University, Guangzhou, 510275, China
[9]Central Research Institute of Electric Power Industry, Tokyo, 100-8126, Japan
[10]Department of Atmospheric Sciences, Pusan National University, Pusan, 46241, South Korea
[11]Research Center for Environmental Changes/Academia Sinica, 11529, Taipei
[12]Universities Space Research Association, Columbia, MD, 21046, USA
[13]Meteorological Research Institute, Tsukuba,305-8506, Japan
[14]Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, 305-8506, Japan
[15]Jiangsu 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
[16]Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, 100084, China
[17]Japan Environmental Sanitation Center, Asia Center for Air Pollution Research, Niigata, 950-2144, Japan
[18]Beijing Normal University, Beijing, 100875, China
[19]Center 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 intercompared and evaluated ~~evaluated and intercompared to~~ with 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~~ for Asia and to 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~~ to MICS-Asia II, the evaluation ~~against~~ of observations was extended to ~~be~~ one-full year ~~in~~ across China and the western Pacific Rim from four months ~~and the western Pacific Rim~~. In general, ~~the~~ model performance levels for O_3 varied widely ~~, depending on~~by region and season~~s~~. Most models captured ~~the~~ key pattern~~s~~ of monthly and diurnal variation of surface O_3 and its precursors in North China Plain and western Pacific Rim~~,~~ but failed to do so for the ~~in~~ Pearl River Delta. A significant overestimation of surface O_3 was evident ~~in~~ from May-September/October and from January-May over the North China Plain, western Pacific Rim and Pearl River Delta. Comparison~~s~~ ~~with~~ drawn from observations ~~revealed~~ show ~~that underestimation on dry deposition velocities and large~~considerable diversity ~~of~~ in photochemical production partly contributed to this overestimation and ~~large~~ to high levels of intermodel variability i~~o~~n O_3 for~~in~~ North China. In term~~s~~ of O_3 soundings, the ensemble average of the models reproduced the vertical structure ~~in~~ for the western Pacific, but overestimated O_3 levels to below 800 h~~p~~Pa in the summer. In the industrialized Pearl River Delta, the ensemble average presented an overestimation ~~in~~ for the lower troposphere and an underestimation ~~in~~ for the middle troposphere. The ensemble average of 13 models for O_3 did not always exhibit ~~a~~ superior performance compared to certain individual model~~s~~~~,~~ in contrast to its superior value for~~ity in~~ Europe. This finding suggest~~s~~~~ed~~ that the spread of ensemble-model values ~~had~~ does not represent~~ed~~ all uncertainties of O_3 or that most ~~models in~~ MICS-Asia III models miss~~ed~~ key processes. This study improve~~s~~ ~~d~~ the performance of modeling O_3 in March at Japanese sites than the previous phase of MICS-Asia ~~(~~(MICS-Asia II)~~)~~. However, it ~~overpredicte~~overpredict~~s~~~~d~~ surface O_3 concentrations ~~in~~ for western Japan in July, which ~~has~~ was not ~~been~~ found ~~in~~ by MICS-Asia II. Major challenges still remain in regard to identifying the sources of bias in surface O_3 over East Asia in CTMs.

1. Introduction:

Tropospheric ozone (O_3) is a significant secondary air pollutant produced through thousands of photochemical reactions and that is detrimental to human health, ecosystems, and climate change as a strong oxidant (WHO, 2005; The Royal Society, 2008). With rapid ~~the fast~~ industrialization and urbanization in the last two decades, O_3 concentration is rising at a higher rate in East Asia than in other regions, and on 30% of ~~the~~ days in megacities (e.g. Beijing, Shanghai Guangzhou in China) values exceed~~s~~ air quality standard of World Health Organization (100 μg/m^3) for 8-hour average surface O_3 concentration (Wang et al.,2017). ~~The~~H~~-h~~igh O_3 concentrations have 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 the 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~~ under which a limit on the 8-hour O_3 maximum was set ~~limits~~ for the first time. However, these measures ~~don't~~ do not prevent the persistent increase of the ground-level O_3 in East Asia. The average mixing ratio of O_3 has increased 20-30% in Japan over the last 20 years (Akimoto et al.,2015). In Chinese megacities, 8-hr O_3 concentrations have increased 10-30~~-~~% since 2013 (Wang et al.,2017).

The ~~primary~~ main method used for the detailed evaluation of ~~the~~ effects of air quality policies at the scale of East Asia is that of numerical air quality modeling. Several global and regional scale CTMs (e.g. GEOS-Chem, CHASER, CMAQ, CAMx, WRF-Chem and NAQPMS) have been developed over the past few decades ~~have been developed~~ and have been widely used to simulate the O_3 formation process and to evaluate strategies for 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~~ Such simulations have identified the key precursors of O_3 formation in East Asia (Zhang et al.,2008; Liu et al., 2010; Tang et al., 2011; He et al., 2017), have assessed the contributions of international and regional transport (Streets et al., 2008; Li et al., 2008), and have predicted ~~the~~ O_3 mixing ratios under~~in~~ different future emission scenarios (Wang et al., 2013). However, discrepancies remain between models and observations, indicating that model simulations of O_3 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~~ addressed using ~~one~~ a 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, O_3, nitrogen compounds and aerosols (Carmichael et al., 2002, 2008). They ~~found~~ study shows that the predicted temporal variations ~~of~~ found for surface O_3 in eight regional CTMs generally tend~~ed~~ to be lower than ~~that~~ those observed in 2001 with poor correlations in the western Pacific in March and December (Han et al., 2008). Model performance levels for O_3 were found to var~~y~~ied ~~largely~~ greatly in southern China. ~~The i~~Inconsistenci~~es~~ ~~i~~of horizontal grids, emissions and meteorological inputs used among models increased ~~the difficulty of explaining~~have rendering explaining intermodel variability ~~in the~~using MICS-Asia II results more difficult. More importantly, model evaluations for~~in~~ industrialized China ha~~ve~~s not been conducted ~~because of~~due to a lack of ~~few~~ observations, which has been detrimental to efforts made to improve O_3model performance levels ~~on O_3~~.

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~~ of 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~~ to simulate~~ing~~ current air quality levels over East Asia. In 2010, MICS-Asia was expanded to Phase III wherein,~~, in which~~ 13 regional CTMs and 1 global CTM ~~are~~ were run over one-full year by 14 independent groups from East Asia and North America, ~~under~~ using 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~~ for MICS-Asia II, new observational data from China were made available for MICS-Asia III and~~, 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 the western Pacific ~~over~~ for one full year ha~~s~~d never

before been performed, ~~which provided~~creating a broader ~~wider~~ database to use for~~in~~ ~~the~~ comparisons. The completeness of MICS-Asia III is therefore unique.

In this paper, we mainly evaluate the ~~abilities~~ capacities of participating models in MICS-Asia III ~~for~~ to simulate~~ing~~ ~~the~~ concentrations ~~–~~of O_3 and its related species ~~in the framework~~within ~~of~~ the MICS-Asia III framework. ~~Several~~ The follwing questions are addressed: (1) ~~What is the performance level of~~How well do various air quality models ~~for~~ perform in simulating O_3 levels in East Asia? (2) How consistent or discrepant are the models? (3) How do muti-model ensembles improve O_3~~the~~ simulation accuracy ~~for O_3~~? This paper is expected to provide valuable insights into the ~~abilities~~ capacities and limitations of CTMs when applied to~~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 O_3 and its precursors mixing ratios in the lowest model layer. For O_3, 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.**~~

~~1.1~~

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), the 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 have been widely applied in modeling studies ~~over~~ of East Asia. Table 1 ~~did~~ does not list model names to maintain each model's anonymity ~~for each participating model~~. Similar behavior was ~~also found~~ observed from~~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 (ds083.2) from the National Centers for Environmental Prediction (NCEP), with 1° × 1° 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 the WRF. The horizontal model domain of~~, 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 top of model ~~top~~ (10 hPa.). ~~The s~~Standard meteorological fields were applied by the majority of groups. Several other models were employed to perform~~ed~~ 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~~ may be slightly different from the standard. GEOS-Chem is driven by the GEOS-5 assimilated meteorological fields taken from the Goddard Earth Observing System of the NASA Global Modeling Assimilation Office. The couples of meteorological data and CTMs vary for each group, likely resulting in a diversified set of model outputs.

MICS-Asia III ~~provided~~ provides a set of monthly anthropogenic emission ~~inventory~~ inventories 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 that includes Regional Emission inventory in ASia (REAS) version 2.1 for the whole ~~of~~ Asian region (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., 2 011; Lu and Streets, 2012), and the official Korean emission inventory from the Clean Air Policy Support System (CAPSS; Lee et al., 2011). B~~The b~~iogenic 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~~ Global Fire Emissions Database version 3 (GFEDv3) (0.5 by 0.5 degree). Volcano SO_2 emissions were provided, with a daily temporal resolution

by the Asia Center for Air Pollution Research (ACAP). The ~~emission group in~~ MICS-ASIA III emission group 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 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. M13 and M14 ma~~k~~de the projections 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 for~~at~~ the top and lateral boundaries of the model domain, which were derived from ~~the~~ 3-hourly global model outputs for the year 2010. The global models were run by University of Tennessee (http://acmg.seas.harvard.edu/geos/) and Nagoya University (Sudo et al., 2002). GEOS-Chem was run with a 2.5º×2º horizontal resolution and 47 vertical layers by University of Tennessee, and Chemical AGCM for Study of Atmospheric Environment and Radiative Forcing (CHASER) was run with a 2.8º× 2.8º horizontal resolution ~~and~~ with 32 vertical layers by Nagoya ~~,~~University. Some models ~~made~~ applied boundary conditions depending on their own ~~previous~~ past experiences.

2.3 Observational data for O_3

In this study, East Asia ~~has been~~was divided into three sub-regions as shown in Fig. 1. The selection of ~~the~~ sub-regions was~~is~~ based on emission~~s~~, climate and observation data coverage. The North China Plain (EA1) and Pearl River Delta (EA2) represent ~~the~~ highly industrialized regions of~~in~~ the mid-latitudes. EA1 is characterized by~~have~~ a temperate and tropical continental monsoon climate with marked seasonality~~, respectively~~. EA2 is located in ~~the~~ south~~-of~~ern China~~,~~ and is less affected by ~~the~~ continental air masses. EA3 ~~consists~~ covers ~~of~~ the northwest Pacific and the Sea of Japan~~,~~ and represents the downwind regions of the Asian continent with a marine climate.

Hourly O_3 and NO_x observations ~~in~~ for the year 2010 in East Asia were obtained from the CERN, PRD-RAQMN~~)~~, and EANET. The CERN was built by the Institute of Atmospheric Physics, Chinese Academy of Sciences and ~~consists of~~includes 19 surface stations ~~within~~ covering an area of 500 × 500 km^2 ~~in~~ across North China Plain (EA1 sub-region; Ji et al., 2012). The ~~se~~ 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. Nine~~9~~ sites were located ~~in the~~within meteorological stations or on campuses of universities in urban regions, with little influence from local sources and sinks. ~~The~~ A comparison of NO and ethene emission rates at these sites in 45 km and 3 km resolution emission inventories show~~ed~~s that observations generally represent~~ed~~ 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 located across the EA2 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 designed to probe ~~the~~ regional air quality, to assess the effectiveness of emission reduction measures and to enhance the roles of monitoring networks in characterizing regional air quality and in supporting air quality management. ~~So~~Thus, the 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, eight remote stations in the northwestern Pacific and Japan (EA3 sub-region) were selected ~~for use in~~ to evaluate~~ing~~ ~~the~~ model performance level~~s~~ ~~in~~ for 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 supplementary section~~s~~ ~~lists~~ provides detailed site description. Our comparisons of NO_x and VOCs emission rates conducted on grid for these stations at 45 km and 3 km resolution emission inventories suggest that our selected stations have rarely received local emissions.

O_3 were measured by Thermo Scientific 49i with UV photometric technology in CERN network and by Thermo Scientific 49C in PRD-RAQMN and EANET network. NO_x was measured by Thermo Scientific 42C NO-NO2-NOx Analyzer with chemiluminescence technology at 40 sites in all three networks (CERN, PRD-RAQMN and EANET). ~~The O_3 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 measurements ~~existed~~ exhibited sometimes biases (especially for stations located far from sources) when using molybdenum converter ~~molybden~~

~~convertors~~ devices since all nitrogen oxydes ~~are~~ were measured. This bias was found to be dependent on chemical conditions. A one-month continuous measurement collected in August ~~by~~ with a chemiluminescence analyzer and Aerodyne Cavity Attenuated Phase Shift Spectroscopy (CAPS) from an urban site in Beijing ~~showed~~ shows that this bias from ~~a~~the chemiluminescence analyzer ~~was~~ is minor ~~small~~ when NO_2 concentrations ~~were more than~~exceed 10-15 ppbv, ~~and~~ rang~~inged~~ from 10% to 30% under low NO_2 conditions (<10 ppbv) (Ge et al., 2013). Measurements collected from a rural site in South Korea reveal a similar pattern across all seasons (Jung et al., 2017). These comparisons suggest that observations made using molybdenum converters may overestimate NO_2 by 10-20% for EA1 and EA2 and 30% for EA3, introducing uncertainties into the NO_2 model evaluation in this study.

3. Model validation and general statistics

3.1 Annual concentrations of surface O_3, nitric oxide (NO) and nitrogen dioxide (NO_2)

Fig. 2 provides a concise comparison of model performance~~s~~ ~~on~~ for annual O_3, NO and NO_2 ~~in~~ for three sub regions ~~in~~ of East Asia. A box-and-whisker representation was used to show the frequency distribution of monthly concentrations measured from~~at~~ stations in each sub-region. The O_3 normalized mean bias (NMB) and root mean square error (RMSE) of the ensemble mean were found to be significantly less than the ensemble median in most c~~situations~~ ases (Table ~~1~~2). Therefore, we only present~~ed the results of~~ multi-model mean ensemble results (Ense). In general, the majority of the models significantly overestimated annual surface O_3 relative to ~~compared with the~~ observations of~~in~~ EA1, EA2 and EA3. Ense overestimated surface O_3 by 10-30 parts per billion by volume (ppbv) for~~in~~ these subregions. Ense NO_2 levels ~~was~~ closely ~~–~~reflect ~~to the~~ observations to within ±20% across~~in~~ all subregions. In EA1 and EA2, Ense NO levels were found to be ~~was~~ 5-10 ppbv lower than those observed ~~observation,~~ while ~~it showed~~exhibiting ~~a~~ reasonable ~~performance in~~levels for EA3.

~~Among all~~Of the models, M11 ~~in~~ for sub-regions EA1 and EA2 ~~,~~and M7 in for EA2 and EA3 more closely reflect ~~were closer~~ O_3 observations. M11 simulated O_3 ~~achieved~~ with~~a~~ RMSEs of 9.5 ppbv and 13.3 ppbv ~~in~~ for EA1 and EA2, respectively (Table 2). The ~~performance levels of~~ models' performance in ~~for~~ simulating O_3 ~~were~~ was found to be closely related to their performance~~s~~ for NO_2 and NO. In

highly polluted regions (EA1 and EA2), a persistent underestimation of NO was evident across most models. As an interesting phenomenon, we found the ~~was that~~ models' performance regarding O_3 measurements to varied~~y~~ greatly for~~in~~ EA3, ~~al~~though ~~they but~~ M8 ~~showed~~ exhibited ~~a~~ a consistent performance with respect to NO and NO_2. This finding suggests that O_3 was significantly affected by other factors in addition to local chemistry in EA3. M8 underestimated O_3 and overestimated NO in all sub-regions by 40-50%. The ~~strongest~~ highest O_3 titration level observed in M8 may have ~~result~~ generate~~in~~ lower O_3 levels than those indicated by other models and observations.

3.2 Monthly variation of surface O_3, NO and NO_2

Fig. 3 presents ~~the~~ monthly mean concentrations of O_3, NO and NO_2 ~~in~~ for the three sub-regions ~~over~~ across East Asia. When~~If~~ two or more observation sites are located~~were~~ in the same model grid ~~of model~~, their mean values ~~will~~are ~~be~~ used to evaluate model performance.

All models captured the observed seasonal cycles of O_3, NO and NO_2 ~~in~~ for EA1. ~~In~~ From May-September, Ense O_3 was 10-30 ppbv higher than observed values ~~observations,~~(30-70% of observed values~~)~~, while Ense NO and NO_2 levels appeared to be consistent with observations~~, attaining~~ with mean biases of < 3 ppbv. This finding suggests that an~~the~~ intercomparison o~~f~~n O_3 production efficiency levels per NO_x with observations is needed. ~~In~~ For EA2, Ense O_3 values agree~~d~~ well with observed high autumn O_3 levels~~,~~ but are overestimated from January to September by 5-15 ppbv (15-60% of observations). This overestimation reached the highest point from in March-April (15ppbv) and led to a spring peak in simulated O_3 values ~~which was~~ not found in the observations. This overestimation i~~wa~~s partly related to the underestimation of NO in the same months, which decreased the titration effect. For NO_2, Ense value agree~~d~~ well with observed values for~~in~~ June-December, and slightly underestimated observations for~~in~~ January-May. For~~In~~ EA3, the ensemble NO_2 was generally close to ~~the~~ observed values~~ations~~ ~~to~~ within ±0.5 ppbv. ~~A~~ S~~s~~ignificant overestimations of O_3 and underestimations of NO were observed from~~existed in~~ June-October. Similar results have been found from~~in~~ MICS-Asia II and through another model inter-comparison project ~~under~~ of the Task Force on Hemispheric Transport of Air Pollution (TF HTAP), ~~which~~ suggesting~~ed~~ that such results may stem from ~~the~~ differences in ~~the~~ representations of ~~dispersion by~~ southwesterly clean marine air masses dispersion observed across ~~in~~ different metrological fields used in CTMs (Han et al.,2008; Fiore et al., 2009).

For individual models, M11 achieved the ~~best~~ highest degree of model reproductivity for~~of~~ monthly mean O_3 levels in EA1 ~~among~~ of the examined models~~.~~. Most of the ~~O~~other ~~most~~ models overestimated O_3 by 100-200% for~~in~~ May-October. The largest levels of model bias and intermodel variability for NO and NO_2 appeared in the winter, ~~which~~ and likely came from the NO_x vertical diffusion and heterogeneous chemistry (Akimoto et al., 2019). In EA2, M7 seems to have achieved the ~~best~~ highest levels of O_3reproducibility ~~for O_3~~. Most of the models (except for M11 and M12) ~~exhibited~~ show high O_3 concentrations for~~in~~ March-May and September-November. Observed O_3 values show~~ed~~ that the highest concentrations appeared from~~in~~ October-November. M11 captured the observed January-May O_3 value ~~because of~~ due to relatively high NO concentrations. However, NO was overestimated by M11 from~~in~~ May-September, ~~which~~ leading to ~~the~~ an underestimation of O_3 levels. In EA3, spatially averaged O_3 concentrations often differ by more than 20 ppbv in ~~the~~ individual models. The highest levels of intermodel variability i~~o~~n O_3 values appeared from~~in~~ May-October, ~~which~~ overestimat~~inged~~ O_3 levels ~~in comparison to~~relative to observations by 10-40 ppbv. Interestingly, although M8, M9 and M14 exhibited ~~a~~ similar magnitudes with observations for~~in~~ June-September, they significantly underestimated observations in other months by 200-300%. A detailed investigation is required in future studies.

3.3 Diurnal concentrations of surface O_3

Sub-regional O_3 diurnal variations are shown in Fig. 4. In general, model results for three sub-regions exhibited a larger spread with a magnitude of 10-50 ppbv ~~throughout~~ across the diurnal cycle than ~~that~~ those observed in Europe and North America (Solazzo et al., 2012). ~~The~~ Summer Ense O_3 ~~in summer~~ levels exhibited a systematic pattern of overestimation (20 ppbv) throughout the diurnal cycle in EA1.This indicate~~d~~s that the models had difficulty ~~dealing with~~measuring summer O_3 levels ~~in~~for the North China Plain. Compared ~~with~~ to summer conditions, ~~there was~~ only a slightly systematic overestimation of Ense O_3 levels ~~in~~ was observed for the other seasons (3-5 ppbv). In EA2, Ense O_3 levels generally agree~~d~~ with ~~the observations in~~ summer, autumn and winter observations. In particular, the O_3 maximum occurring at around noon was ~~reproduced,~~ reasonably reproduced. ~~There was o~~Only a 3-5 ppbv overestimation was observed from~~during~~ 16:00-23:00 and in early morning (6:00-10:00). In the spring, a systematic overestimation of Ense O_3 values ~~exited~~ was observed across ~~in~~ the whole diurnal cycle (5-10 ppbv). In EA3, Ense captured the minor ~~small~~ diurnal variations ~~of~~in O_3 across all ~~in~~ four

seasons, but significantly overestimated observations for the ~~in~~ summer and autumn (5-20 ppbv). In the spring and winter, differences between Ense and observations fell~~were~~ within 5 ppbv.

~~Among~~ Of all of the models, M11 exhibited the best model performance level in measuring~~on~~ peak daily O_3 concentrations of 60 ppbv from~~in~~ 14:00-16:00 in EA1, but it still overestimated nighttime O_3 levels by 10 ppbv. Compared to their ~~with~~ performance in measuring summer patterns, the models~~' performances had~~ performed ~~a~~ significantly better~~improvement~~ in measuring winter conditions ~~because of~~due to the weak intensity of photochemical reactions~~,~~ except in case of M2, M10 and M8. Differences between observations and most simulations ~~in~~ for both the nighttime and daytime ~~were~~ fell within 5 ppbv. These differences ~~e contrast of~~in the models' performances between the summer and winter imply ~~ied~~ that the variety of ~~parametrizations on~~ chemistry parametrizations applied to ~~in~~ different models partly explain~~ed~~ the intermodel variability of simulated O_3 levels in EA1 (North China Plain). For~~In~~ EA2, the majority of models agree~~d~~ well with ~~the~~ diurnal variations occurring in the summer and autumn. However, most models exhibited~~had~~ a tendency to overestimate the O_3 concentrations for~~in both~~ both the daytime and nighttime in the spring. The overestimated magnitudes exceeded 10 ppbv and 25 ppbv (~~out~~ of observed values of 20-35 ppbv) for the ~~in~~ nighttime and daytime, respectively. M11 reproduced ~~d the~~ observed O_3 levels for ~~in~~ the spring~~,~~ but underestimated O_3 levels for the ~~in~~ summer and autumn. For~~In~~ EA3, ~~the~~ significant levels of intermodel variability persisted ~~still existed~~ throughout the year. A~~The~~ ~~a~~mplitudes of intermodel variability except for those of M8 and M14 reached approximately 20 ppbv and ~~approximately~~ 10 ppbv in the spring-summer and autumn-winter, respectively. M8 and M14 ~~exhibited~~ generated the lowest O_3 values of the ~~among~~ models for~~in~~ the whole year.

3.4 Error statistics on surface concentrations

In this section, we present statistics ~~concerning~~ on the models' performance ~~levels of the models~~ based on monthly values. ~~They~~ Values are calculated ~~by~~ with equations shown in Appendix A. On a yearly basis, all models ~~showed~~ observe the highest (0.8-0.9) and lowest (0.1-0.6) correlation coefficients for O_3 for~~in~~ EA1 and EA2, respectively (Table 2). ~~The h~~High correlations were observed in EA1~~ were~~ mainly because the summer-maximum and winter-minimum seasonal cycles are~~is~~ ~~the~~ typical ~~pattern in~~of polluted regions ~~that were well~~ represented in all of the participating models. In general, Ense ~~performed~~ performed~~a~~ better ~~performance level~~ than the individual models ~~for~~ in representing NO_2

for~~in~~ East Asia, reproducing the observed seasonal cycles and magnitudes. However, Ense did not always exhibit ~~ed~~ a superior performance in measuring ~~for~~ O_3 levels over ~~certain~~ individual models for~~in~~ East Asia, which stands~~was~~ in contrast to its performance ~~in~~ for Europe (Table ~~1~~2). M7 and M11 agree~~d~~ well with observations for~~in~~ EA1 and EA2~~,~~ while ENSE tend~~sed~~ to overestimate O_3 concentrations for~~in~~ May-September in EA1 and for January-September in EA2. Loon et al. (2007) indicated that Ense~~NSE~~ exhibit~~s~~ ~~ed~~ a superior performance level only when the spread of ensemble-model values i~~was~~ representative of ~~the~~ O_3uncertainty ~~of O_3~~. This indicate~~sd~~ that most models d~~oid~~ not reflect this uncertainty or missed key processes ~~in~~ of MICS-Asia III.

~~The large~~Considerable overestimation s made by~~of~~ most of the models for~~in~~ May-September led to high NMB (0.25-1.25) and RMSE (10-33 ppbv) values for ~~in~~ EA1. M11 generated~~had~~ the lowest NMB (0.09) and RMSE (9.46 ppbv) ~~among~~ values of the examined models. For~~In~~ EA2, M9 and M10 generated~~had~~ stronger~~larger~~ correlations than the other models. However, their corresponding NMB and RMSE values were also the highest. These findings ~~is~~ impli~~ed~~y that systematic model biases are present ~~existed~~ in these two models. M7 exhibited ~~a~~ lower NMB and RMSE values than the other models, but its correlation was measurements as only 0.29. For~~In~~ EA3, ~~the~~ correlations exhibited the largest degree of intermodel variability ~~among~~ across all sub-regions, ranging from -0.13-0.65. M7 generated~~showed~~ the lowest NMB and RMSE and ~~. This is~~ likely ~~caused by~~due to ~~the~~ cancelling effect of its overestimation for the~~in~~ summer and underestimation ~~in~~ for other seasons (Fig. 3).

For NO, ~~correlations of~~ model correlations for~~in~~ EA1 ranged from 0.57-0.68, ~~which indicated~~showing that all of models ~~did a good job in~~effectively reproduc~~ed~~ing ~~the~~ spatial variability in~~of~~ NO for ~~in~~ this sub-region (Table 3). ~~The~~ NMBs indicated underestimation by the models except in case of M8, ~~which~~ mostly occurred for the~~in~~ winter. This underestimation can be partly ~~was~~ attributed to the coarse model horizontal resolution (45 km) used in ~~the~~ MICS-Asia III, which hardly reproduced concentrations of short-lived species (e.g., NO). In contrast to most of the other models, M8 overestimated NO concentrations ~~in~~ for all three sub-regions. It is noted that NO observations ~~of NO~~ for EA3 were too low (<0.3 ppbv) ~~in EA3~~ to be discussed in this study.

Table 4 shows ~~the~~ statistics ~~of~~ on the models' performance ~~levels~~ in measuring~~for~~ NO_2 levels. In general, most of the models ~~exhibited a better~~ performan~~ce~~ed better in~~levels for~~ representing NO_2 than

O_3 and NO ~~in~~ for EA1. ~~The~~ NMBs ranged from -0.28-0.32, ~~which were much l~~falling far below ~~ower than~~those measured for O_3 ~~((~~0.48-1.25). ~~The c~~Correlations of~~were~~ 0.54-0.66 were recorded, implying the models' reliable ~~model~~ performance ~~levels for~~in reproducing ~~the s~~spatial and ~~month to~~ monthly variability of NO_2 for~~in~~ EA1. Similar those for~~to~~ O_3 and NO, ~~the~~ correlation coefficients for~~of~~ NO_2 in EA2 remained low. ~~. .~~ Thus, a dedicated investigation of~~n~~ O_3, NO and NO_2 levels i~~i~~n EA2 is urgently needed, but falls beyond the scope of this study. In EA3, correlation coefficients ranged from 0.5-0.72. ~~The~~ NMBs and RMSEs except for those of M8 ranged from -0.42-0.46 and 0.91-1.79 ppbv, respectively.

3.5 Vertical profiles of O_3

Fig. 5 shows the vertical profiles of observed and simulated O_3 levels ~~in~~ for East Asia for~~in~~ the summer and winter. Ensemble means (Ense) ~~presented~~ show an underestimations and overestimations of~~for~~ EA2 O_3 levels for the ~~in~~ middle (500-800 h~~p~~Pa) and lower (below 900 hP~~p~~a) troposphere, respectively. In the winter, ~~the~~ underestimations ~~was even~~ extend~~ed~~ to 200 hP~~p~~a. M~~The m~~agnitudes of underestimations and overestimations reach~~ed~~ 10-40 ppbv and 10-20 ppbv, respectively. For~~In~~ EA3, Ense reproduced the vertical structure of ozone for ~~in~~ both the summer and winter. An overestimation of ~~existed~~ less than~~below~~ 800 ~~hpa~~hPa ~~in summer~~, with a magnitude of 10-20 ppbv was observed for the summer.

~~A large~~ High levels of intermodel variability ~~of~~ in O_3 exceeding~~above~~ 300 hPa is evident across all ~~in all~~ sub-regions, which is attributable to the ~~various~~ varied ~~different~~ top boundary conditions ~~among~~ applied by the models. However, this ~~large~~ considerable variability was not transmitted to the middle troposphere (400-600 hPa), in which O_3 concentrations were consistent ~~among~~ across the models. In the lower troposphere, a ~~small~~ minor level of intermodel variability of below 900 hP appeared in the winter ~~appeared below 900 hPa~~ in three sub-regions, and slowly decreased with height. ~~The m~~Mean standard deviations (SD) of models ~~(σ)~~ of below 900 hP~~p~~a were recorded as 7.6 ppbv, 6.9 ppbv and 6.0 ppbv ~~in~~ for EA1, EA2 and EA3, respectively, ~~which~~ cover~~inged~~ 18.3%, 15.0% and 15.4% of mean O_3 concentrations. In 700-900 hP~~p~~a, ~~σ~~ SD levels decreased to 5.4 ppbv, 4.4 ppbv and 4.8 ppbv ~~in~~ for EA1, EA2 and EA3, 12.2%, 9.4% and 10.8% of mean O_3 concentrations, respectively.

In the lower troposphere, ~~the~~ intermodel variability in the summer ~~were~~ was generally higher than ~~those~~ that in the winter. In polluted regions (EA1), ~~σ~~ SD levels reached 16.3 ppbv (20.8 % of mean

concentrations) in the summer, greatly exceeding those ~~,~~ in winter (6.2 ppbv, 15.2%). Various vertical structures of O_3 were found below 700 hPa in summer. O_3 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 O_3 transport from the upper boundary layer contributed ~~a lot to~~considerably to ~~the~~ discrepancies~~y~~ between M1, M6 and M11. In EA2, vertical structures of O_{3}~~x~~ among models were found to be consistent, but concentrations differed more than those in EA1. SD~~σ~~ values covered 22% of mean concentrations.

4. Multi-model ensemble O_3 and comparison with MICS-Asia II

4.1 Spatial distribution of single model and multi-model ensemble O_3

Fig. 6 shows ~~that~~ the spatial distributions of MICS-Asia III ensemble mean surface O_3 values (Ense) and the coefficient of variation (CV). The CV is defined as the standard deviation of the modeled O_3 divided by the average. The larger the ~~value of~~ CV value, the lower the degree of consistency among the models. ~~In~~ For the summer, ~~ENSE~~ Ense predicted ~~the~~ an elevated O_3 concentration belt in the middle-latitudes (30º-45ºN). A region of O_3 in excess of 60 ppbv stretched across the North China Plain and East China ~~East~~ Sea, ~~which was much~~far ~~higher than values in~~ MICS-Asia II (45-50 ppbv) values ~~for the year of~~ for 2001 (Han et al.,2008). In other seasons, the O_3 values distribution shows higher O_3 over the ocean than in eastern China, reflecting ~~the~~ O_3 titration from high NO_x emissions. Due to the stratospheric injection, surface O_3 over the Tibet P~~p~~lateau remained at ~~a~~ high level~~s~~ ~~in the whole~~throughout the year, ranging from 50 to 65 ppbv. The seasonal cycle of surface O_3 levels determined from ~~in~~ Ense ~~in~~ via MICS-Asia III agree~~d~~s with that ~~in~~ observed from MICS-Asia II, but O_3 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 the winter. These high values in the low-latitude western Pacific (10ºS-15ºN) and Indian Ocean~~s~~ were likely caused by the treatment of lateral boundaries in the models. ~~In~~ For MICS-Asia III, M7, M8 and M9 employed the default model configurations ~~of models~~, and the others employed outputs of the GEOS-Chem/CHASER/MOZART-GOCART global model. Compared ~~with~~ to those ~~MIC~~of MICS-Asia II, ~~the~~

CVs ~~in~~ for the Asian continent except for the winter remained at ~~a~~ similar levels in this study (0.1-0.3) (Carmichael et al.,2008).

Although all of the models similarly predicted the emergence of an elevated summer O_3 concentration belt in the middle-latitudes (30°-45°N), the magnitudes of ~~the~~ enhanced O_3 ~~were~~ levels varied between ~~different among~~ the models (Fig. 7). M5 predicted the highest O_3 concentrations of 60-90 ppbv ~~in~~ for the North China Plain (EA1) and for its outflow pathways including the ~~,~~ Bohai Sea, East China Sea, Korea, Japan and the Sea of Japan (Locations are shown in Fig. S1 in the ~~supplements~~supplementary section)~~,~~ whereas M8 predicted the lowest levels of35-50 ppbv. Overhangs of 30 ppbv contour lines extending into the n~~N~~orthwestern Pacific ~~in~~ among the Asian continent outflow plume differed considerably ~~among~~ between the models. ~~The~~ A plume of 30 ppbv or higher O_3 levels was observed from in M1-M6, M13 and M14, reaching~~ed~~ further south and east of Japan (135°E, 20°N), than those of M8, M10 and M11 (120°E, 30°N). ~~In~~ From MICS-Asia II and HTAP, differences ~~of~~ in the frequency of marine air masses from the western Pacific Ocean were thought to be a possible cause of O_3 discrepancies observed ~~y~~ over oceans ~~among~~ between the models ~~because of~~due to different meteorological drivers (Han et al., 2008). For~~In~~ MICS-Asia III, ~~the~~ winds fields reflected ~~in~~by the models were similar ~~because~~due to the use of the ~~models the~~ same or similar meteorological fields (Fig. S2 in the supplementary sections~~)~~. ~~Hence,~~ T~~t~~hese~~is~~ inconsistencies~~y~~ ~~among~~ between the models ~~have~~ resulted from the combined ~~influence~~ effects of a series of factors~~,~~ ~~that~~ includ~~ed~~ing the diversity ~~in~~ of condensed gas-chemical mechanism and heterogeneous chemistry. Li et al. (2015) found ~~that the~~ chemical production ~~was~~ to be the dominated controlling factor of O_3 along ~~the~~ outflow pathways near the North China Plain in the summer~~,~~ rather than lateral and top boundary conditions. Impact of aerosols on ozone in these regions ~~were~~ are frequently reported in Olson et al. (1997) and Li et al. (2018)~~, by~~ to alter~~ing~~ photolysis rates and heterogeneous chemistry patterns. D~~The d~~etailed comparisons ~~on~~ of parameterizations of these processes in models are needed in future intermodel comparison projects focused on ~~in~~ Asia.

In the winter, ~~the~~ distribution patterns of O_3 were quite ~~alike~~ similar between the~~among~~ models~~,~~ with high concentrations observed over parts of western China, northeastern India and the western Pacific from the East China Sea to southern ~~of~~ Japan (Fig. S3 in the supplementary sections~~)~~. In spring and

autumn (Fig. S4 and Fig. S5 in the supplements), O_3 concentrations were generally higher than they were in the winter acrossin the whole model domain because ofdue to the enhancement of solar radiation or stratosphere-troposphere exchanging fluxes of O_3. All of the models exhibited an enhancement of O_3 levels over southern Tibet, northeastern India and the western Pacific, generally echoing patterns observed in the winterA major feature consistently produced by all models was the enhancement of O3 over southern Tibet, northeastern India and the western Pacific, which was generally similar to that in winter. . Increases inThe position of O3 observed enhancement further north of Japan arewas comparable with winter.

4.2 Comparison with MICS-Asia II

In From MICS-Asia II, model evaluation on O_3 were conducted in onlyon sites in the western Pacific. Fig. 8 presents the simulated and observed surface O_3 levels at these monitoring sites in derived from the phase II and III of the MICS-Asia project. Note that different models were employed in two phases. In general, most of the models captured the major distributions of O_3 at most sites in both MICS-Asia II and III. ENSE results are showed a good consistent cy forin March and December of 2001 and 2010. The Uunderestimations of O_3 levels in March at Japan sites (site 4: Sado-seki, site 5: Oki and site 6: Banryu) in Phase II wereas largely improved remedied in Phase III. However, the surface O_3 at observed in 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 was not been found in Phase II, in for which the differences with from observations was were valued at approximately 5 ppbv. Rural sites in western Japan were are located in the upwind regions of Japanese domestic emissions, and and are subjected to usually used to capture the impacts of Asian continent outflows. The oOverestimated O_3 ivalues nfor North China Plain (EA1) in Phase III contributed considerablya lot to the enhanced concentrations measured for at sites of western Japanese sites in July 2010. This indicatesd that the transboundary transport from the Asian continent according toin MCIS-Asia III was likely overestimated compared relative to with that in measured from MICS-Asia II.

5. Discussions

In reference to MICS-Asia II, Han et al. (2008) ~~guessed~~ hypothesized that ~~the diversity of~~variations in ~~–~~meteorological fields, dry deposition, PBL, model treatment of chemistry and other physical processes had contributed to model biases ~~with~~ in relation to observations and ~~the~~ intermodel variability. Quantifying the contributions of these processes ~~is one effective way to~~can help explain model biases ~~by~~ through sensitivity simulations. ~~But~~ However, this task comes with ~~required a~~ tremendous ~~amount of~~ computational costs when applied to ~~for~~ 14 models. The~~A~~ qualitative analysis of~~n~~ potential causes of these processed based on~~by~~ comparisons ~~between~~ of models and observations ~~on these processes~~ is essential to ~~narrow~~ selecting sensitivity simulating scenarios for the next phase of MICS-Asia. In MICS-Asia III, common input data (emission and meteorology) ~~provide a good chance~~be effectively used in ~~for this qualitative analys~~this qualitative analysis based~~is~~ 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~~ under MICS-Asia II and is ~~believed~~ intended to ~~be~~ help~~ful~~ ~~for~~ model developers ~~to~~ improve model performance ~~in~~ for East Asia.

5.1 Dry depositions

Previous studies ~~revealed~~ show that dry deposition processes ~~are~~ serve as the key net sink of O_3, accounting for ~~about~~ roughly 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 O_3 were compared. Simulated deposition velocities were calculated from Eq. (1):

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

w~~W~~here F and C represent the simulated dry deposition flux and surface O_3 concentrations, respectively. We determined ~~the~~ spatial mean dry deposition velocities ~~at~~ from stations in each sub-~~r~~egion.

Fig. 9 presents the simulated and observed monthly spatial mean dry deposition velocities of O_3. ~~In~~ For EA1, ensemble mean values ~~underestimated~~ overestimated observed dry deposition velocities of O_3 (v_d) ~~in~~ for August-September, but still fell ~~into~~ within the range of the observed standard deviation. This shows that other factors rather than dry deposition could play important roles in overestimations of

August-September O_3 values in EA1. In October-November, simulated v_d apparently underestimated observations by 30-50%. Among the models, ~~This underestimation could contribute to the overestimation of O_3 concentrations in summer discussed in section 3.2. T~~the lower dry deposition velocities in May-July ~~from~~ for M1, M2, M4 and M6 than that of M11 partly explained higher ~~summer~~ May-July surface O_3 from those simulations than that ~~from~~ in 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 EA1. In October November, simulated v_d apparently overestimated observations by 30 50%.~~Notably, our observations were made on grassland, which covers ~20% of the land area in EA1. There are few v_d observations on agriculture crops (50% of the land area) in China. Hardacre et al. (2015) reported O_3 dry deposition measurements on crops in Europe and simulated O_3 dry deposition in 15 global models. Both observations and simulations showed that O_3 dry deposition velocities on agriculture crop class were quite similar to those of grassland, showing uncertainties related to be the representativeness of measurement sites used in this study did not affect our conclusions.

~~In~~For EA2, similar features ~~with~~ as those of EA1 were~~are~~ found. M1, M2, M4 and M6 were quite consistent with each other, with a seasonal cycle and a ~~of~~ spring minimum. M11, M12 and M14 show~~had~~ no obvious signs of seasonal variability~~,~~ with a magnitude of 0.1-0.2 cm/s. S~~The s~~easonal patterns in M13 are ~~was~~ considerably different from ~~the~~ those of 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 EA3, most stations were located in remote oceanic sites, and few dry deposition observations were ~~conducted~~made. Thus~~So~~, we collected observations ~~in~~ from other oceanic sites to evaluate model performance (Helmig et al., 2012). Ense values ~~of~~for v_d agree reasonably~~d~~ well with ~~with~~ observations ~~reasonably~~ (Fig. 9). Both observations and simulated v_d values show~~ed~~ a July-September maximum with

a magnitude of 0.02-0.03 cm/s. Park et al. (2014) ~~revealed~~ found~~that~~ surface O_3 levels in EA3 ~~were~~ to be more sensitive to dry deposition parameterization schemes in CTMs. O_3 measured from ~~on~~ oceans differed by 5-15 ppbv in East Asia ~~resulting from~~due to the use of ~~different~~ various dry deposition parameterization schemes. Thus, more observations are needed over oceans in EA3 to ~~decrease~~ mitigate ~~the uncertainties on~~ O_3 simulation~~s~~ uncertainties.

5.2 Relationships between surface NO_x and O_3

In general, surface O_3 mainly comes from ~~the~~ photochemistry processes 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 (EA2) 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~~effectual to investigating sources of intermodel variability and model errors concerning ~~on~~ O_3 chemistry in East Asia. Fig. 10 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~~ for the stations shown in Fig. 1.

~~In~~ For EA1 (North China Plain), observations clearly ~~revealed~~ show that O_3 concentrations decreased with ~~the~~ an increase in NO_x concentration. O_3 concentrations mostly remained at 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.~~ The slope and intercept of the regression line between observed O_3 and NO_x were measured as -0.77 ppbv/ppbv and 59.5 ppbv, respectively. Among the models, M11 ~~were~~ results are in relative agreement with observations~~, reasonably~~. The slope and intercept (-1.01 ppbv/ppbv, 63.23 ppbv) ~~were~~ reflect the ~~close to~~ observations. The ~~O~~other models show~~ed~~ a higher degree of model bias and intermodel variability i~~o~~n relationships between O_3 and NO_x. Their slopes mostly range~~d~~ from -1.25 ppbv/ppbv to -2.13 ppbv/ppbv, amounting to 1.3-2.8 times of observed slope. Their intercepts were 74.9 -121.2 ppbv, ~~much higher than~~far exceeding observation~~s~~ (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 O_3 rather than those in M1 and M6. This finding is consistent with

the low slope in M11. To reduce the impact of O_3 buildup and transport by consuming NO_x, relationships between O_x (NO_2+O_3) with NO_x was compared (Fig. S7 in the supplementary section). Observed O_x increases with the increase of NO_x levels, with coefficient of determination (R^2) of 0.61. Most of the models (except for M8, M11 and M13) failed to reproduced observed positive correlations between O_x and NO_x, and their R^2 only ranged from 0.01-0.08. The slope, intercept and R^2 of M8 and M11 are relative agreement with observations. ~~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 O_3 than M1 and M6. This is consistent with the low slope in M11. Interestingly, M13 maintained a similar O_3 level at all NO_x levels (Slope: -0.09), which was different from other models and previous theoretical results.~~

~~In~~ For EA2, all models reproduced observed key patterns in which O_x positively correlated with NO_x. For O_3-NO_x relationships, M1, M2, M4 and M6 reproduced observed O_3 levels under~~in~~ low NO_x conditions(< 30 ppbv) but failed to capture ~~the~~ low O_3 under high NO_x conditions (30~40 ppbv), accounting for~~. This explained the~~ overestimations of these models for O_3 in May-September. By contrast, M8 and M11 produced excessively high NO_x values, ~~which~~ resulting~~ed~~ in their underestimations of~~for~~ O_3 values. ~~In~~ For 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 M10, which ~~is~~ stands in contrast to observations. This finding suggests that more attention is needed ~~when policy makers designate the O_3 regime (VOCs limited or NOx 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~~showing ~~the~~ that local chemistry ~~appeared to~~were not ~~be~~ a key factor shaping ~~of~~ O_3 formation. Thus, we did not ~~discuss~~ examine the simulated O_3-NO_x relationship further ~~in this study~~.

5.3 Other factors

Previous studies ~~revealed~~ show that O_3 precursors are mostly constrained within the boundary layer (Quan et al., 2013). ~~The model evaluation on~~ PBLH model evaluation is essential for the interpretation of model biases with observations. Unfortunately, this ~~evaluation~~ evaluation was not applied~~conducted~~ in MICS-Asia II. In 2016，Guo et al. (1996) calculated the PBLH using the bulk Richardson number (Ri) method from the radiosonde network of the L-band sounding system of the

China Meteorological Administration (Vogelezang and Holtslag,1996). The system provides fine-resolution profiles of temperature, pressure relative humidity, wind speed and direction. In MICS-Asia III, all selected models exhibited the spring-maximum and winter-minimum season cycle ~~i~~for~~n~~ EA1 (Fig. S6 in the supplement~~s~~ary section), ~~which~~ capturinge~~d~~ the mainj~~or~~ ~~pattern of~~ climatological patter~~n~~ of PBLH observations (Guo et al.,2016). The Ense on PBLH only overestimated radiosonde measurements by 100-200 m (~10-15%) and ~~. This is~~ likely ~~caused by~~ due to sampling bias ~~the inconsistency of samples~~ between the models and measurements. The simulation recorded as ~~was~~ the mean value of 12 hours (08:00-20:00), while the average of the measurements was calculated based on a 3-hour period~~s~~ (08:00, 14:00 and 20:00). For~~In~~ EA2, the observed PBLH did not vary~~ied~~ as much as that for~~in~~ EA1, and differences between seasons ~~were~~ ranged within 100 m. This pattern was captured by the models. As was observed from ~~Similar as~~ EA1, the simulated PBLH for~~in~~ EA2 ~~was~~ exceeded the measurements by100-200 m ~~higher than measurements~~. Few measurements of~~n~~ remote oceanic site~~s~~ ~~were conducted~~ in East Asia were collected. Thus~~So~~, we compared simulations with European Centre for Medium-Range Weather Forecasts Reanalysis Data (von Engeln et al., 2013). Both showed a winter-maximum pattern for~~of~~ PBLH.

~~The East Asia monsoon played an important role in seasonal cycle of O_3 in subregions by the long-range transport. Besides local intensive photochemical productions, the O_3 summer maxima in EA1were 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 O_3 in September November was quite different from EA1 and EA3. This is largely attributed to the long range transport of O_3 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 O_3 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 O_3 by regional emissions, respectively. In the next studies, we will conduct the intermodel comparison on transport fluxes of O_3 between sub-regions over East Asia.~~

6. Summary

~~In the~~Under MICS-Asia III framework, the evaluation and intercomparison of 13 CTMs was~~ere~~ conducted ~~with~~ using a wide variety of observations covering two Chinese industrialized regions and the western Pacific, using long-term simulations for ~~the year~~ 2010 with a focus ~~. This study has focused~~ on O_3~~,~~, ~~,~~ NO and NO_2. In particular, surface O_3 levels in China~~, was evaluated,~~ which were~~as~~ ~~absent~~ neglected in ~~the~~ previous model-intercomparison projects~~, were evaluated~~. ~~Large~~ Considerable levels of intermodel variability ~~of~~ in O_3 ~~existed~~ were observed ~~in~~across all subregions ~~over~~ of East ~~Asia in this study~~, with model concentrations varying by ~~a~~ factors of 2 to 3 between different models.

A model ensemble was ~~conducted~~ produced and evaluated. In general, the model ensemble captured ~~the~~ key patterns of monthly and diurnal O_3~~,~~~~,~~$_3$~~,~~ NO and NO_2 in the North China Plain and ~~the~~ western Pacific Rim. It failed to capture the observed seasonal cycle of O_3 for the~~in~~ Pearl River Delta of China. ~~In~~ For the North China Plain and western Pacific R~~r~~im~~,~~~~.~~ t~~T~~he model ensemble severely overestimated surface O_3 levels ~~in~~for May-September by 10-30 ppbv. This overestimation systematically appeared in both daytime and nighttime. Similarly, the model ensemble tende~~d had a predominate tendency~~ to overestimate ~~the~~ spring daytime and nighttime O_3 concentrations for the ~~in spring in~~ Peral River Delta. Compared to MICS-Asia II, MICS-Asia III was less prone to underestimating~~ion of~~ surface O_3 in March ~~at~~ for Japanese sites. However, it predicted ~~too enhanced~~excessively high surface O_3 concentrations for~~at~~ western Japan in July, which was not the case ~~in~~ for MICS-Asia II. In term of O_3 soundings, the ensemble model used in this study reproduced the vertical structure in the western Pacific, but overestimated O_3 below 800 h~~p~~Pa in the summer. For the~~In~~ industrialized Pearl River Delta, the ensemble average presented an overestimation ~~for~~ of O_3 levels for~~in~~ the lower troposphere and underestimations in the middle troposphere. ~~This study revealed~~We find that the ensemble average of 13 models for~~on~~ O_3 d~~id~~oes not always ~~exhibit a superior~~ performance better than ~~to certain~~ individual models ~~in~~ for East Asia in~~,~~ ~~which~~ contrast~~ed~~ to~~with its~~ their performance for~~in~~ Europe. This suggest~~ed~~ that the spread of ensemble-model values ~~had~~ does not represent~~ed~~ all uncertainties in~~of~~ O_3 levels or that most ~~models in~~ MICS-Asia III models missed key processes. ~~Unlike~~ In contrast to ~~the~~ performance levels for O_3, ENSE ~~demonstrated superior~~ performance ~~level~~ better than individual models for NO_2 in East Asia.

MICS-Asia II ~~guessed~~ outlines ~~some~~ potential ~~reasons~~ causes of variability~~ies~~ among models. Quantifying the contributions of these processes to O_3 concentrations serves as an ~~is one~~ effective way to explain model biases through~~by~~ sensitivity simulations. However, ~~But~~ this would incur ~~required a~~ tremendous ~~amount of~~ computational costs ~~for~~ when applied to 14 models. In this study, we conducted a qualitative analysis ~~on~~ of potential causes by comparing~~son~~ ~~between~~ models and observations for ~~on~~ these processes to identify~~narrow~~ sensitivity simulating scenarios for the next phase of MICS-Asia. ~~The~~ Our comparisons show ~~revealed~~ that the ensemble model over~~under~~estimated observed dry deposition velocities of O_3 for~~in~~ August-September in North China Plain, ~~which~~ showing that other factors rather than dry deposition ~~could~~ may contribute to the overestimation of simulated O_3 concentrations in the summer. For the~~In~~ western Pacific, simulated v_d values agreed with observations reasonably well. Photochemical treatment in models may contribute~~d~~ to ~~the~~ O_3 overestimations in North China Plain. The studied ~~M~~models captured ~~the~~ major pattern of climatologicaly~~y~~ pattern of ~~of~~ PBLH observations for~~in~~ three subregions of~~ver~~ East Asia. More evaluations of~~on~~ turbulent kinetic energy in PBL ~~is urgent~~are needed to ~~for~~ assess ~~the~~ vertical mixing in future studies.

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_{ij} and Obs_{ij} modeled and observed the i^{th} monthly concentrations of air pollutants at the j^{th} station, having mean value $\bar{y}$ and $\overline{obs}$. m and n represent the numbers of stations and months.

Correlation coefficient (R)

$$R = \frac{\sum_{j=1}^{m}\sum_{i=1}^{n}(y_{ij}-\bar{y})(obs_{ij}-\overline{obs})\sum_{i=1}^{n}(y_i-\bar{y})(obs_i-\overline{obs})}{\sum_{j=1}^{m}\sum_{i=1}^{n}(y_{ij}-\bar{y})^2\sum_{i=1}^{n}(y_i-\bar{y})^2\sum_{i=1}^{n}((obs_i-\overline{obs})^2\sum_{j=1}^{m}\sum_{i=1}^{n}((obs_{ij}-\overline{obs})^2} \quad \text{(A1)}$$

Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i}^{n}(y_i-Obs_i)^2\sum_{j=1}^{m}\sum_{i}^{n}(y_i-Obs_{ij})^2}{n}} \quad \text{(A2)}$$

Normalized Mean Bias (NMB)

$$NMB = \frac{\sum_{i=1}^{n}(y_i-Obs_i)\sum_{j=1}^{m}\sum_{i=1}^{n}(y_{ij}-Obs_{ij})}{n\times\bar{y}\times\overline{Obs}} \quad \text{(A3)}$$

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.

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.

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.

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.

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.: 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.

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, 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, 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.

Fiore, A. M., Dentener, F. J., Wild, O., Cuvelier, C., Schultz, M. G., Hess, P., Textor, C., Schulz, M., 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.

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, *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 model evaluation, *Atmos. Chem. Phys.*, 18, 4859-4884, https://doi.org/10.5194/acp-18-4859-2018, 2018.

Ge, B., Sun, Y., Liu, Y., Dong, H., Ji, D., Jiang, Q., Li, J., and Wang, Z. : Nitrogen dioxide measurement by cavity attenuatedphase shift spectroscopy(CAPS) and implications in ozone productionefficiency and nitrateformation in Beijing, China, *J. Geophys. Res. Atmos.*, 118, 9499–9509, doi:10.1002/jgrd.50757,2013.

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.

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.

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., 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.

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.: 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.

Jung, J., Lee, J. Y., Kim, B. M., Oh, S. H.: Seasonal variations in the NO_2 artifact from chemiluminescence measurements with a molybdenum converter at a suburban site in Korea (downwind of the Asian continental outflow) during 2015–2016, *Atmos. Environ.*, 165, 290-300,2017.

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.

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.

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.

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.

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.

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.

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.

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.

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 O_3-NO_x-VOC chemistry and NOx-VOC indicators, *J. Geophys. Res*.*-Atmos.*, 107(D22), 4659, doi:10.1029/2001JD001123, 2002.

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.

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 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.

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., Abaguidi, 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.

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, Policy Document, 15/08,2008.

Vogelezang, D. H. P. and Holtslag, A. A. M.: Evaluation and model impacts of alternative boundary-layer height for mulations, *Bound.-Lay. Meteorol.*, 81, 245–269, doi:10.1007/BF02430331, 1996.

von Engeln, 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 Teylor, G. R.: A theoretical method for computing vertical distributions of acidity and sulfate production within cumulus clouds, *J. Atmos. Sci.*, 43, 339–355, 1986.

Wang, S., Ackermann, R., and Stutz, J.: Vertical profiles of O_3 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 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.

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.

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.

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): 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 joint air quality management. *Aero. Air Qual. Res.,* 13(5), 1582-1597, 2013.

Table and Figure captions:

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

Table. 2 Statistical analysis for surface O_3 in three subregions over East Asia (R: correlation coefficient; NMB: Normalized Mean Bias; RMSE: Root Mean Square Error)

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 NO_2 in three subregions over East Asia (R: correlation coefficient; NMB: Normalized Mean Bias; RMSE: Root Mean Square Error)

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

Fig. 2 Box-plots of observed and simulated annual NO_2 (left column), NO (middle column) and O_3 (right column) frequency distribution by determined from 13 models and, averaged forin stations over in EA1, EA2 and EA3 , and in time for the whole 2010 year. n represents denotes the numbers of stations. The rectangle represents the inter-quantile range (25^{th} to 75^{th} percentiles). The small star identifies the mean, the continuous horizontal line withininside the rectangle identifies the median, andthe whiskers extend between the minimum and maximum values.

Fig. 3 Time series of monthly NO_2, NO and O_3 levels simulated by all models and their ensembles (Ense), in ppbv, averaged over all observed stations acrossin three subregions ofover East Asia (EA1: top row, EA2: middle row, EA3: bottom row). Observations are denoted also shown by the black line. n represents the numbers of stations.

Fig. 4 Seasonal mean diurnal cycle of surface O_3, in ppbv, as a function of hours, for all models and their ensembles, averaged acrossover all observed stations in three subregions ofover East Asia (EA1: top row, EA2: middle row, EA3: bottom row). Observations are also showndenoted by the black line. n represents the numbers of stations

Fig. 5 Simulated O_3 profiles in for the summer and winter of 2010, averaged over all observed stations inacross three subregions over of East Asia (EA1: left column, EA2: middle column, EA3: bottom column). The Oozonesonde data observe infor 2010 were as taken from the data base stored by World Ozone and Ultraviolet Radiation Data Centre (WOUDC) database

Fig. 6 EThe ensemble mean seasonal surface O_3 concentrations and CV values for the different seasons. CV is defined as the standard deviation of the modeled fields divided by the average, for the different seasons

Fig. 7 Surface O_3 spatial distribution derived from 13 models for summer 2010 (unit: ppbv).

Fig. 8 The Mmodeled and observed monthly mean concentrations of O_3 at for EANET sites in the phase II (left panel) and III (right panel) of the MICS-ASIA project. The sSolid line represents the ensemble mean. Note that data in MCIS-ASIA II and III data are refer to in the period of March, July and December of 2001 and 2010, respectively. IDs of the mMonitoring sites representsdenote the following: 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)

Fig. 9 Simulated and observed monthly O_3 dry deposition velocities (V_d) for~~of~~ M1, M2, M4, M6, M11, M12, M13 and M14 for~~in~~ three subregions of~~ver~~ East Asia (EA1: top row, EA3: middle row, EA3: bottom row). TEX, STR, GGSEX and AMMA ~~represents~~ denote observations ~~i~~for~~n~~ TexAQS06 (7 July–12 September 2006; north-western Gulf of Mexico), STRATUS06 (9–27 October 2006; the persistent stratus cloud region off the coast of 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). Observational data were taken~~is~~ from Sorimachi et al. (2003), Pan et al. (2010), and Helmig et al. (2012).

Fig. 10 Scatter plots ~~between~~ for monthly daytime (08:00-20:00) surface NO_x and O_3 for~~at~~ each station in~~over~~ EA1 (red), EA2 (green)and EA3 (blue) in May-October, for observations (obs) and models. Also shown are the linear regression equations ~~between~~ for NO_x and O_3 in ~~–~~EA1 (red) and EA2 (green).

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	45 km	45 km	45 km	45 km	45 km	45 km	45 km	45 km	45 km	45 km	45 km	45 km	0.5 ° ×0.667°	45 km
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	58 m	58 m	58 m	58 m	58 m	58 m	29 m	58 m	16 m	44 m	48 m	27 m		100 m
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(Colella and Woodward 1984)	PPM	Yamo	5th order monotonic	5th order monotonic	5th order monotonic	5th order monotonic	Walcek and Aleksic (1998)	Walcek and Aleksic (1998)	PPM	PPM
Vertical diffusion	ACM2 (Pleim,2007)	ACM2	ACM2	ACM2	ACM2	ACM2	3th order Monotonic	3th 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,2 000)	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	ISORROPIAv1.7(Nenes et al.,1998)	Kajino et al. (2012)	ISORROPIAv1.7	ISORROPIAv1.7
Boundary conditions	GEOS-Chem global model (Martin et al.,2002)	Gipson (1999)	GEOS-Chem global model	CHASER global model (Sudo et al., 2002a, 2002b)	CHASER global model	CHASER global model	Liu et al. (1996)	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]Standard 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.

[c]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 O_3 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	EA2 (n=13)	0.29	-0.08	11.11	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 (n=19)	0.60	-0.71	23.36	EA2 (n=13)	0.11	-0.89	15.94	EA3 (n=8)	0.15	-0.70	0.16
M7		0.63	-0.75	24.91		0.04	-0.78	15.32		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 NO_2 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	EA2 (n=13)	0.23	-0.30	13.86	EA3 (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

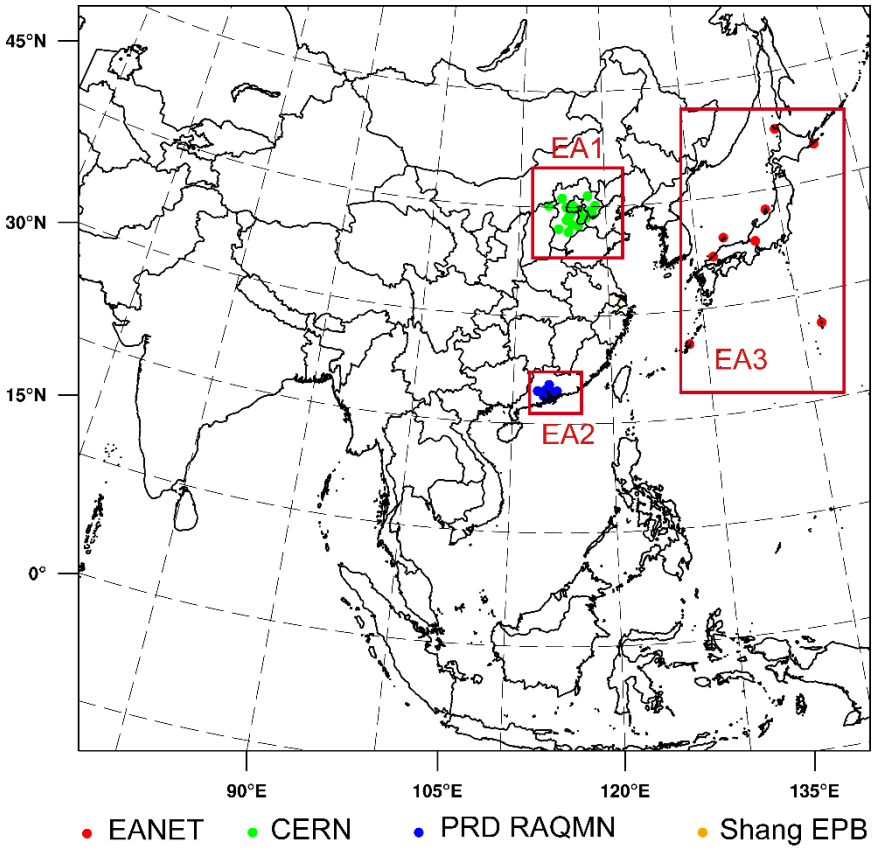


Fig.1 Li et al., 2018

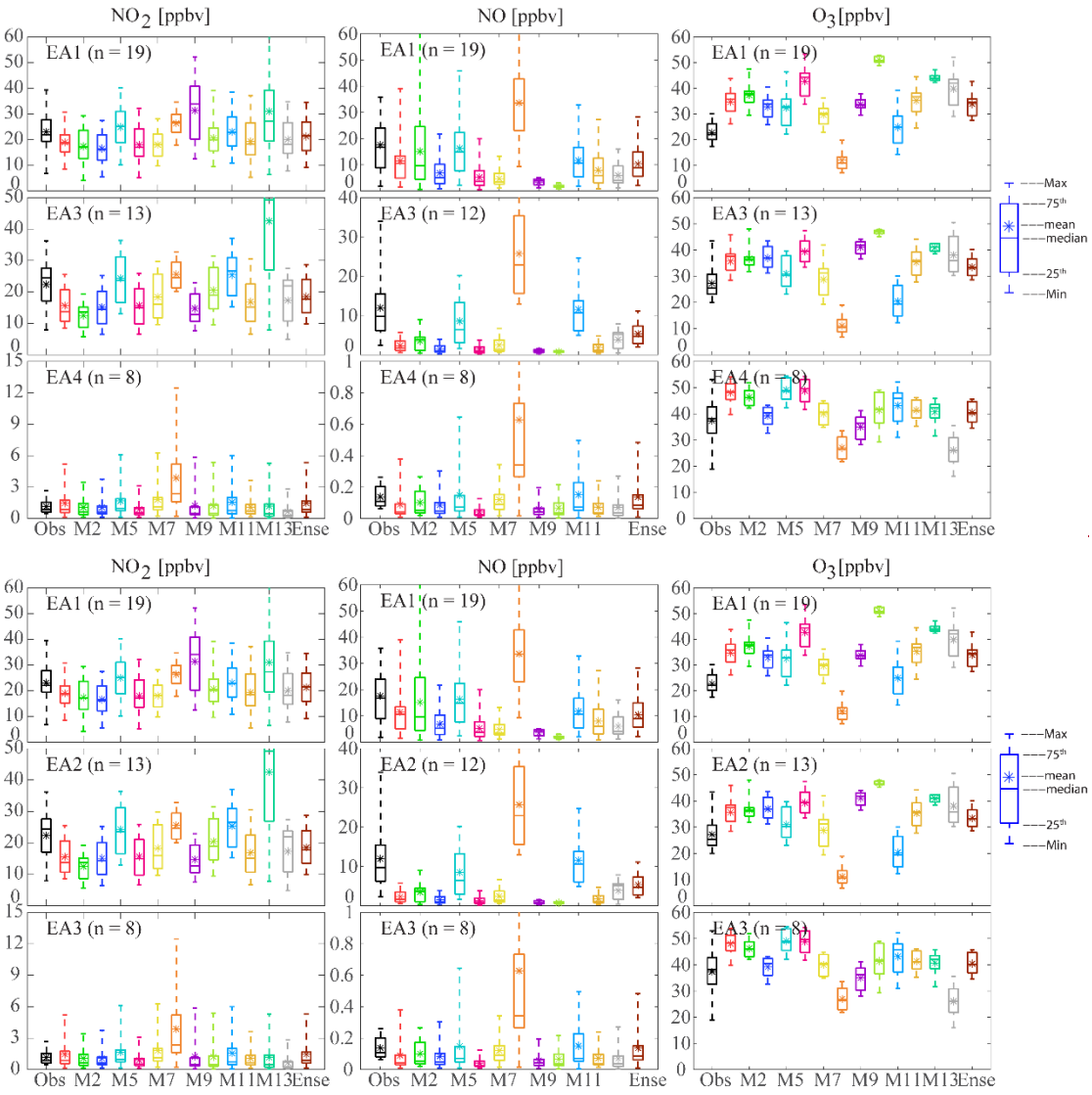
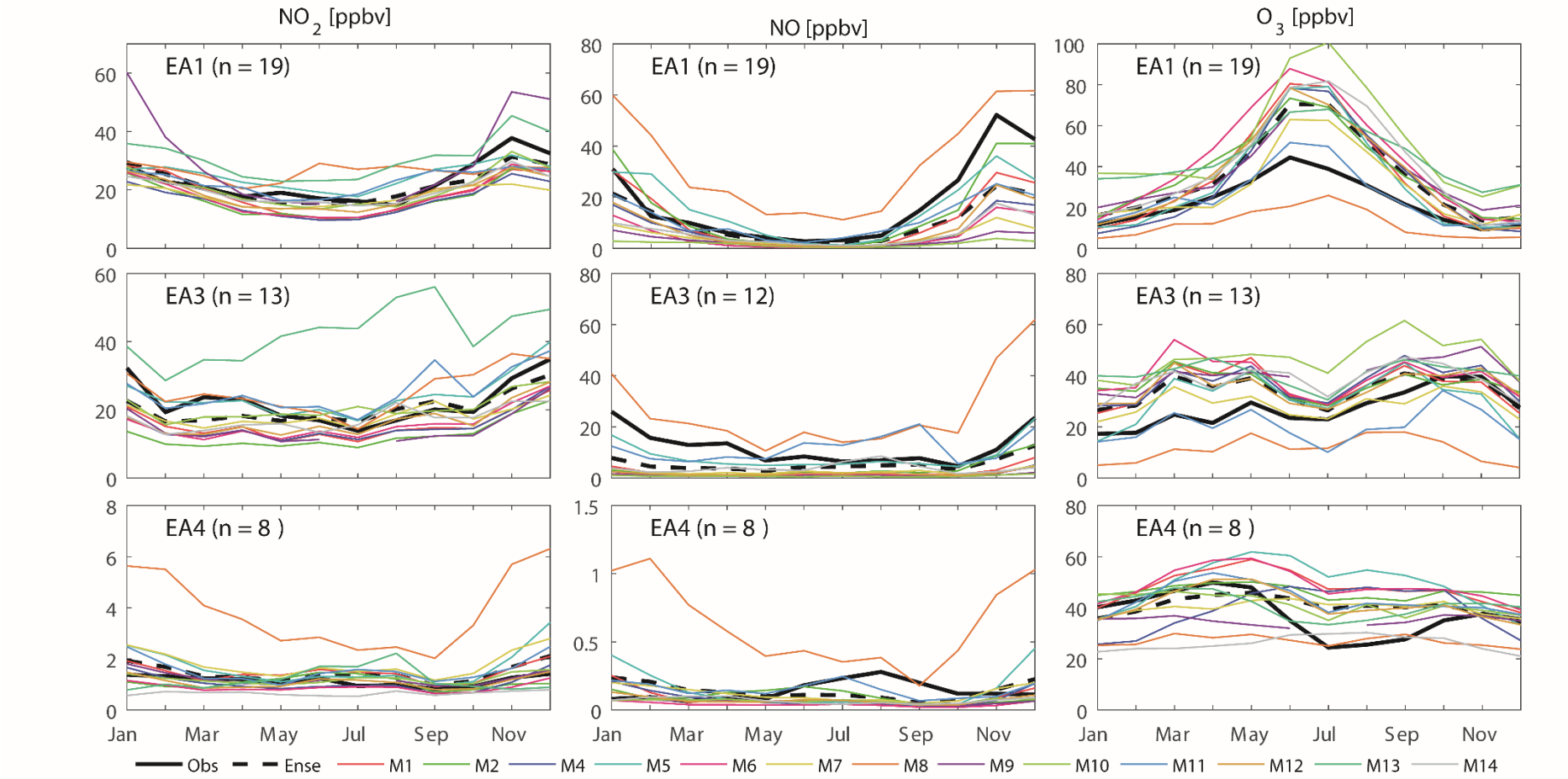


Fig.2 Li et al., 2018


NO2 [ppbv]
NO [ppbv]
O3 [ppbv]
EA1 (n = 19)
EA1 (n = 19)
EA1 (n = 19)
EA3 (n = 13)
EA3 (n = 12)
EA3 (n = 13)
EA4 (n = 8)
EA4 (n = 8)
EA4 (n = 8)
Jan
Mar
May
Jul
Sep
Nov
Obs
Ense
M1
M2
M4
M5
M6
M7
M8
M9
M10
M11
M12
M13
M14

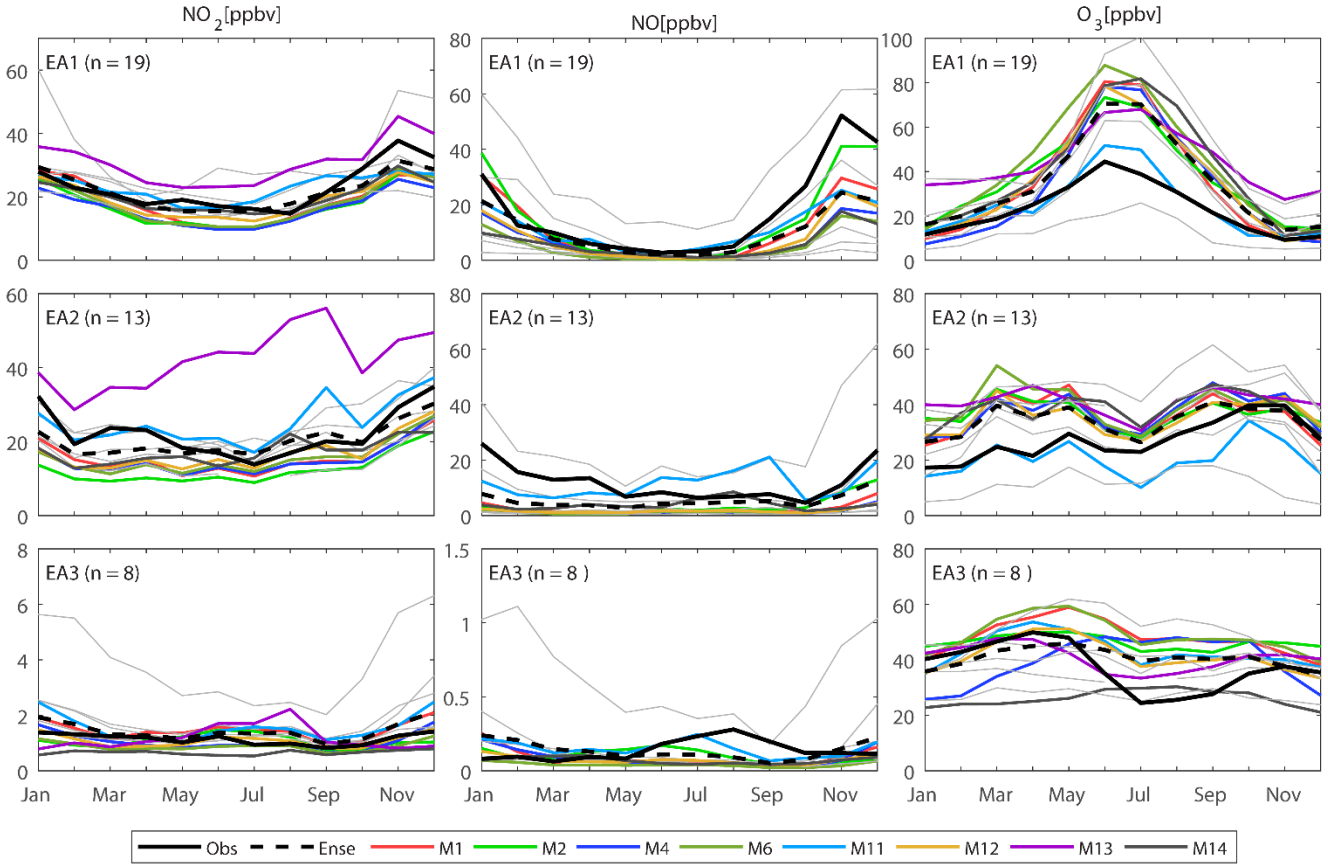
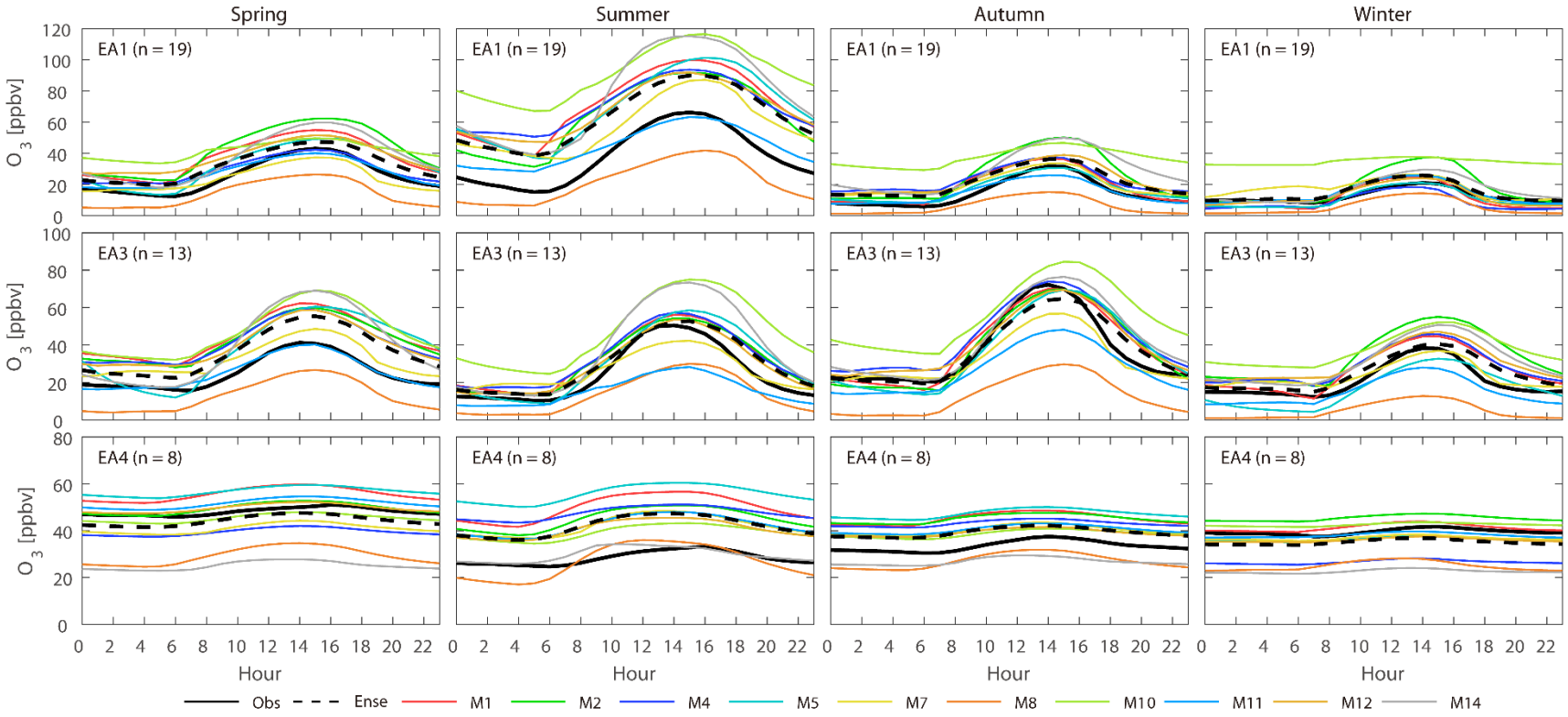


Fig.3 Li et al., 2018


Spring
Summer
Autumn
Winter
EA1 (n = 19)
EA3 (n = 13)
EA4 (n = 8)
O3 [ppbv]
Hour
Obs
Ense
M1
M2
M4
M5
M7
M8
M10
M11
M12
M14

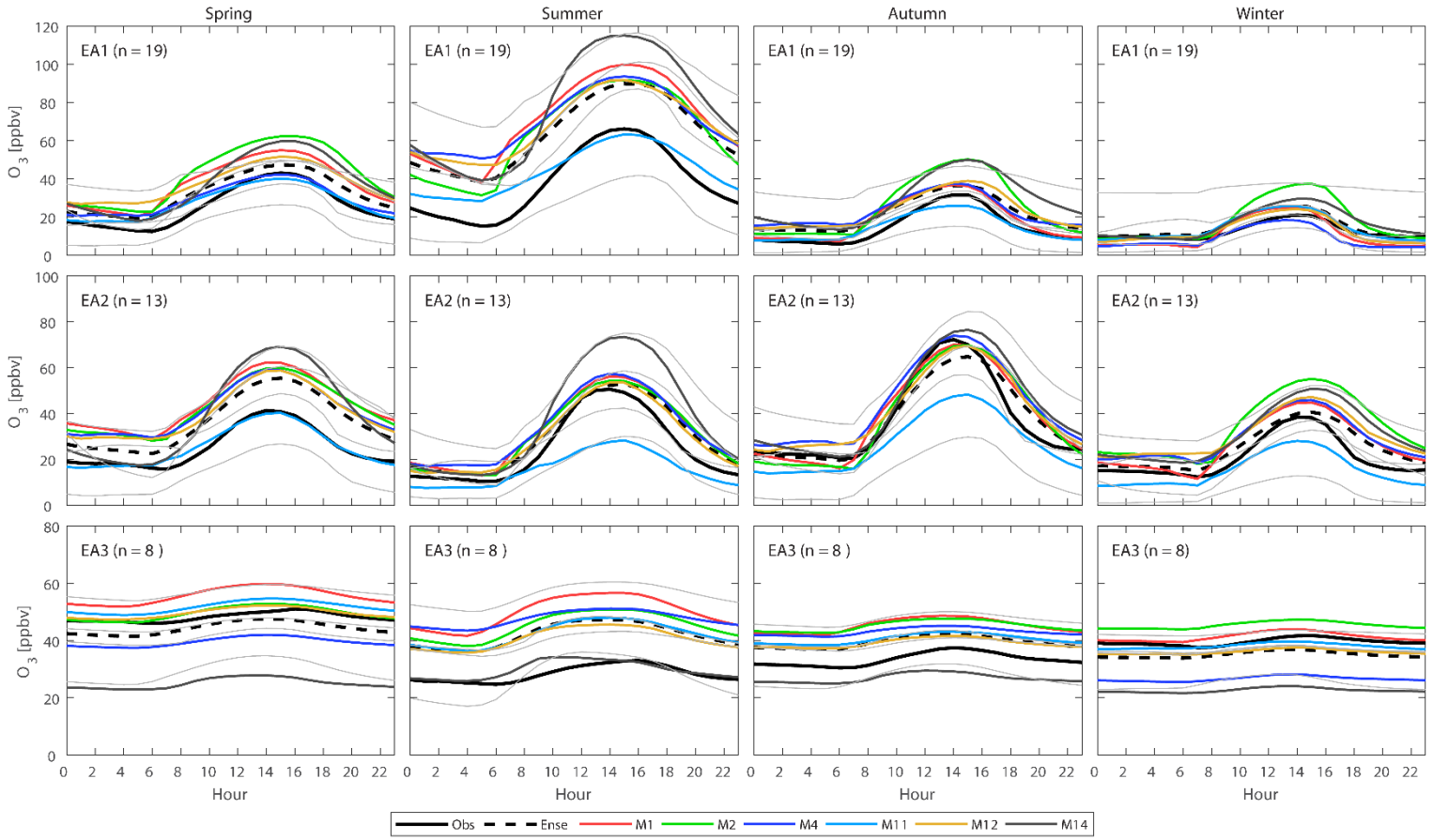
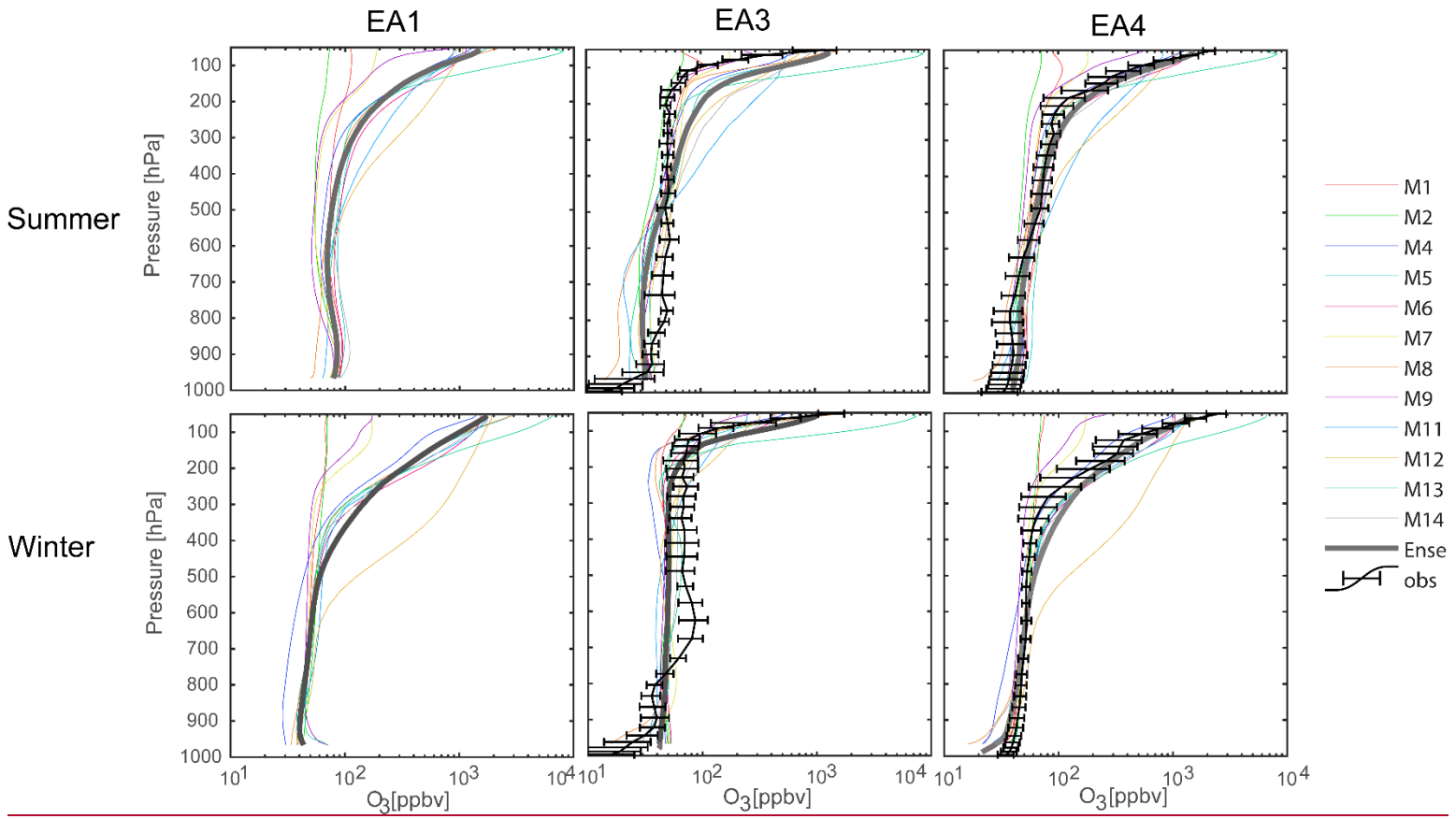

Spring
Summer
Autumn
Winter
EA1 (n = 19)
EA2 (n = 13)
EA3 (n = 8)
EA3 (n = 8)
O3 [ppbv]
Hour
Obs
Ense
M1
M2
M4
M11
M12
M14

Fig.4 Li et al., 2018


EA1
EA3
EA4
Summer
Winter
Pressure [hPa]
100
200
300
400
500
600
700
800
900
1000
10^1
10^2
10^3
10^4
O_3[ppbv]
M1
M2
M4
M5
M6
M7
M8
M9
M11
M12
M13
M14
Ense
obs

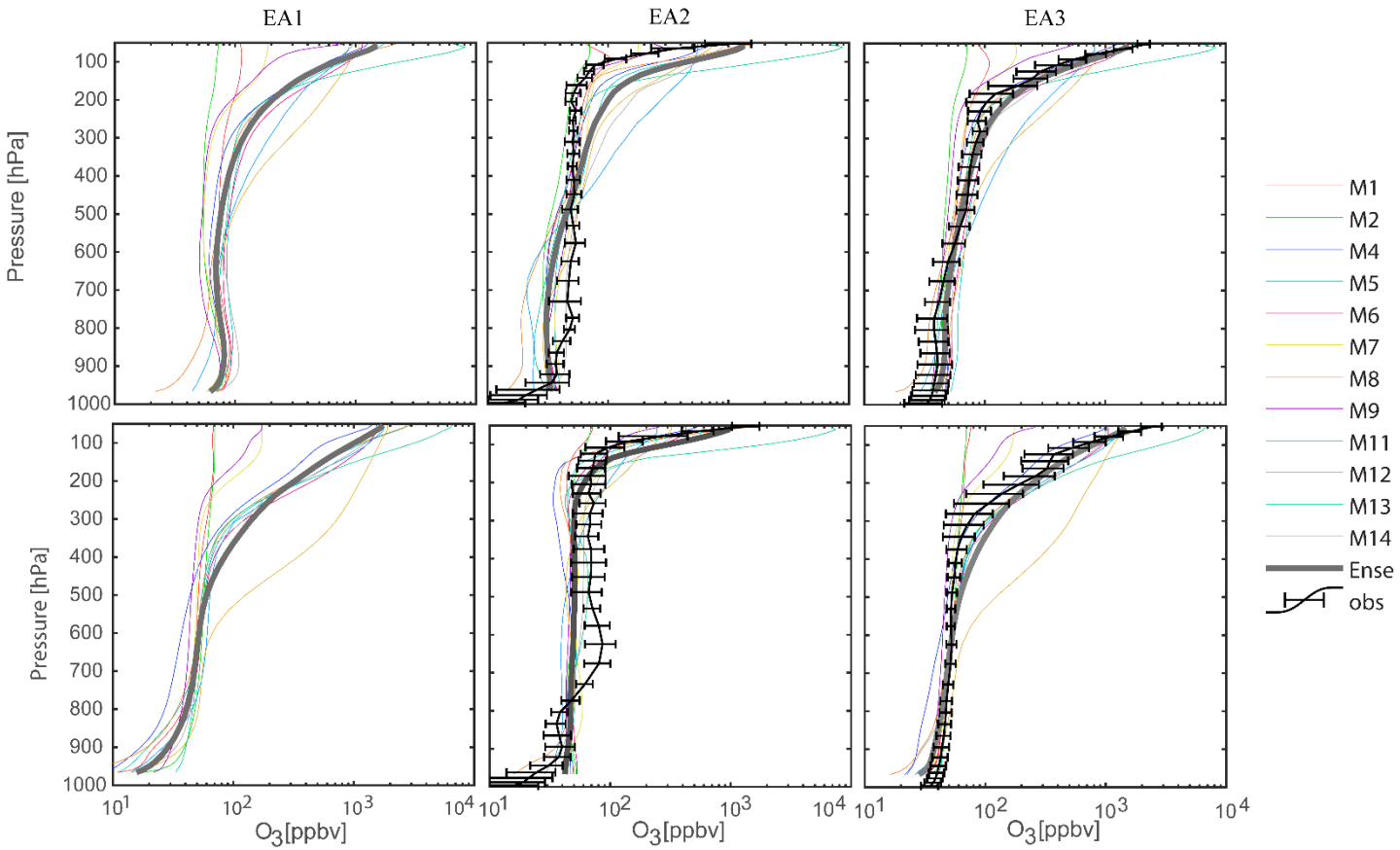


Fig.5 Li et al., 2018

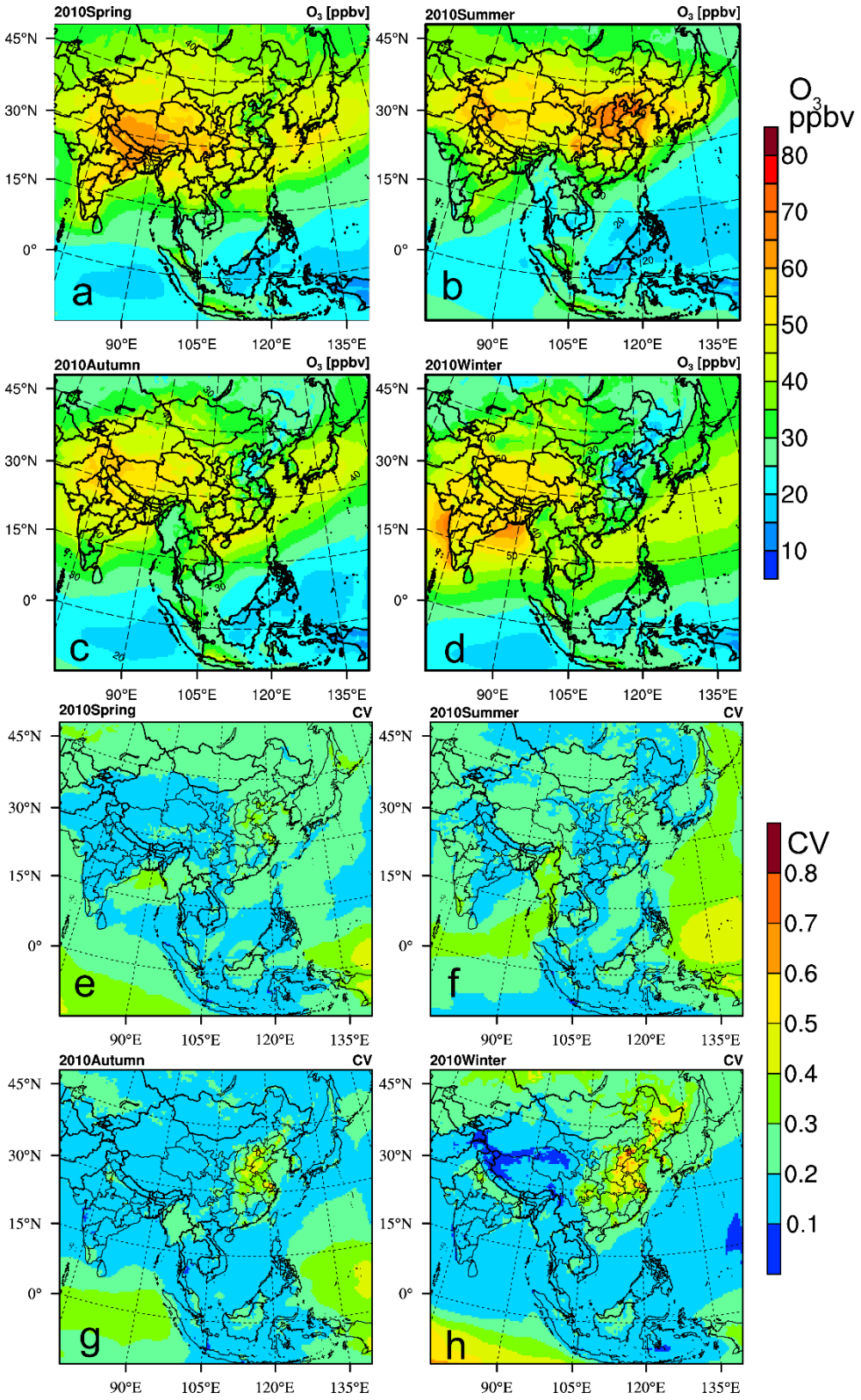


Fig.6 Li et al., 2018

.

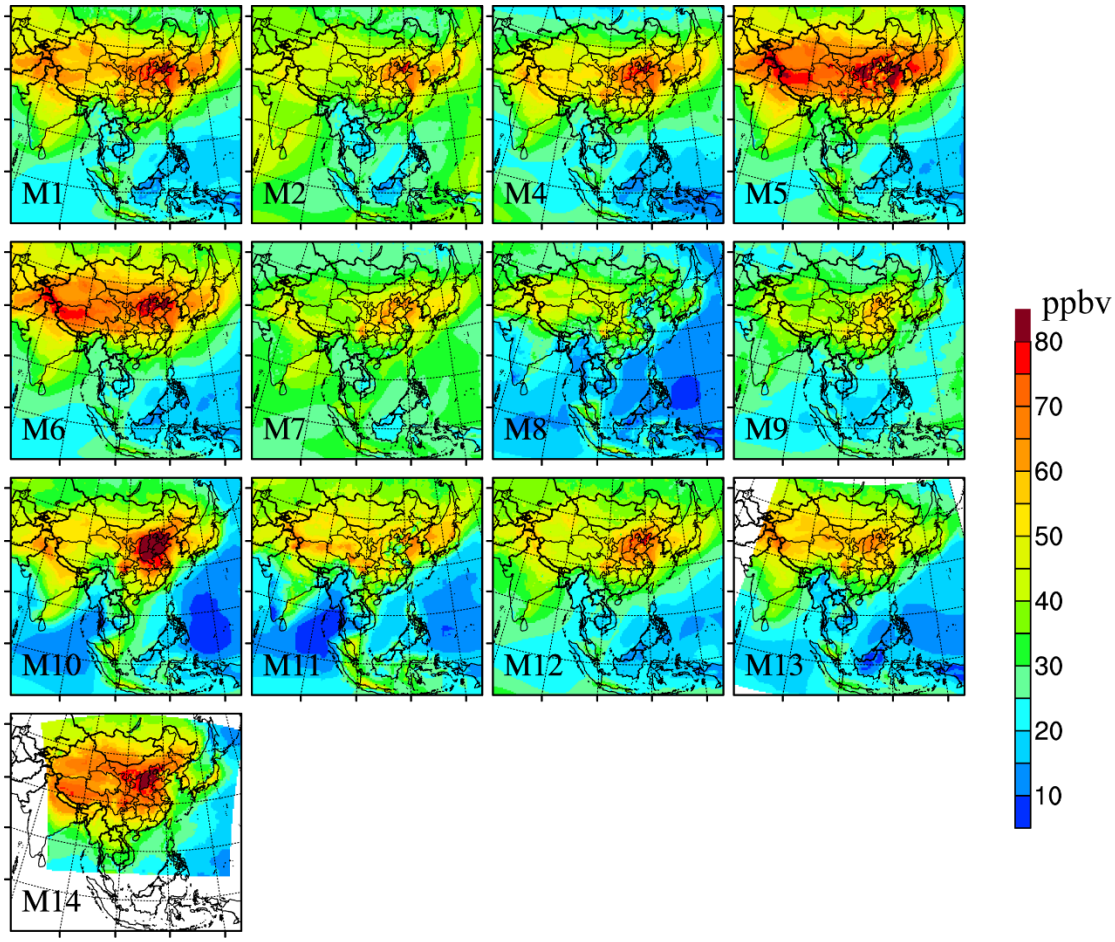


Fig.7 Li et al., 2018

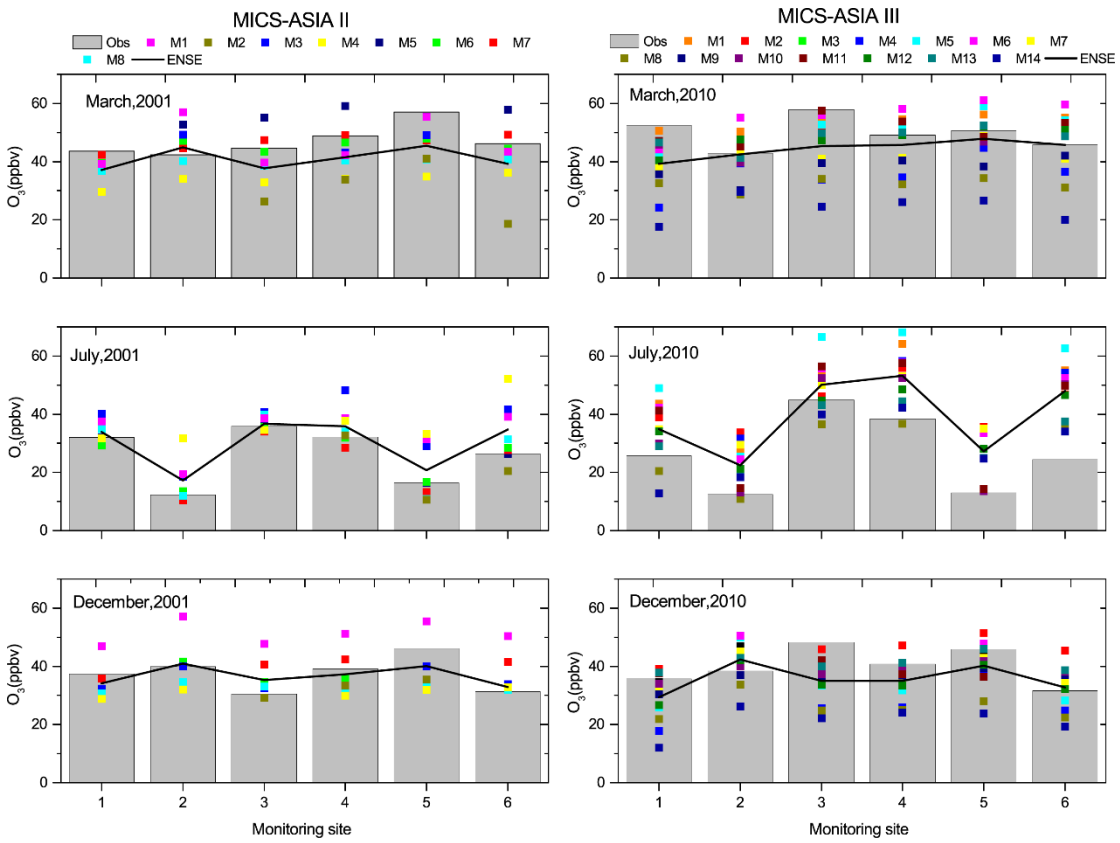
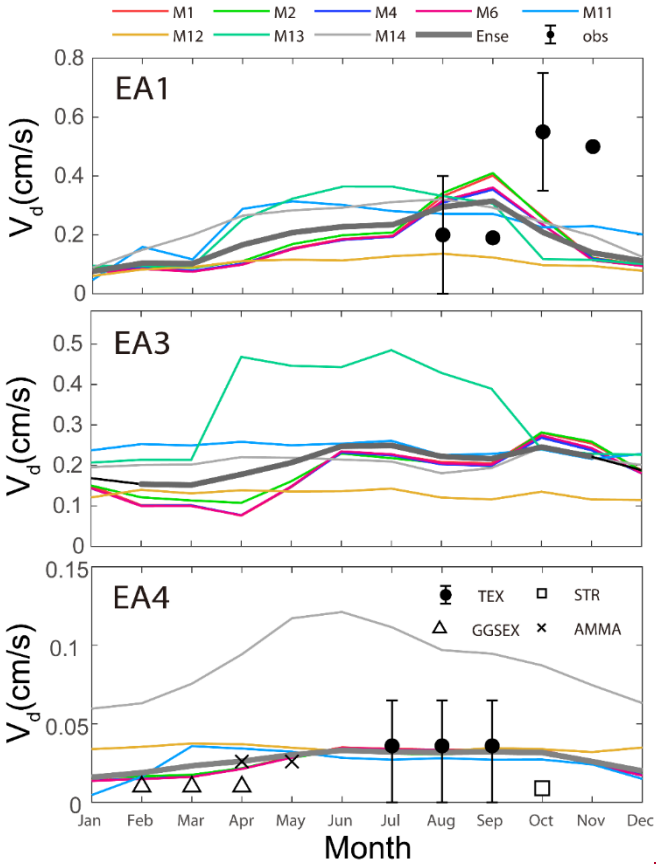


Fig.8 Li et al., 2018


M1
M2
M4
M6
M11
M12
M13
M14
Ense
obs
EA1
EA3
EA4
TEX
STR
GGSEX
AMMA
V_d(cm/s)
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month

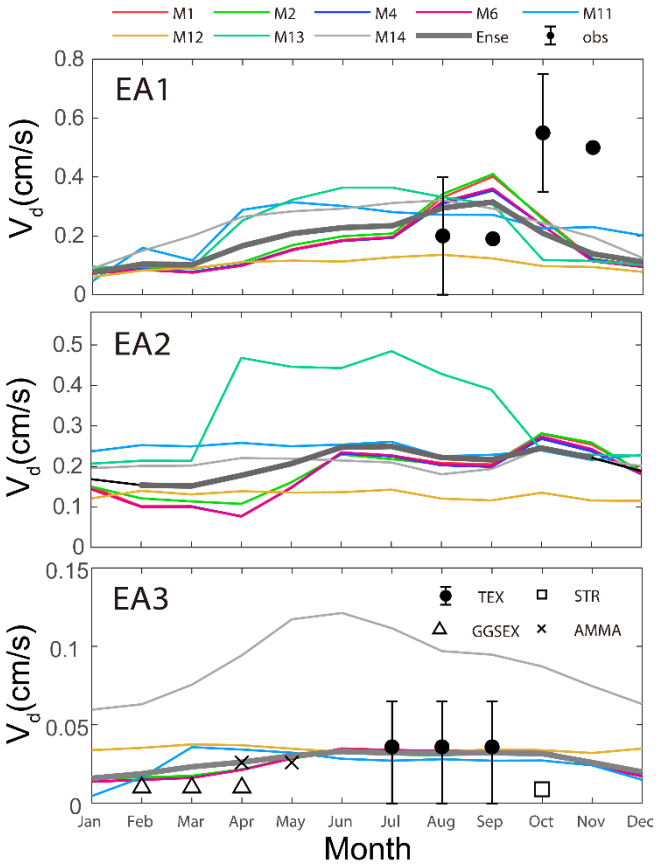
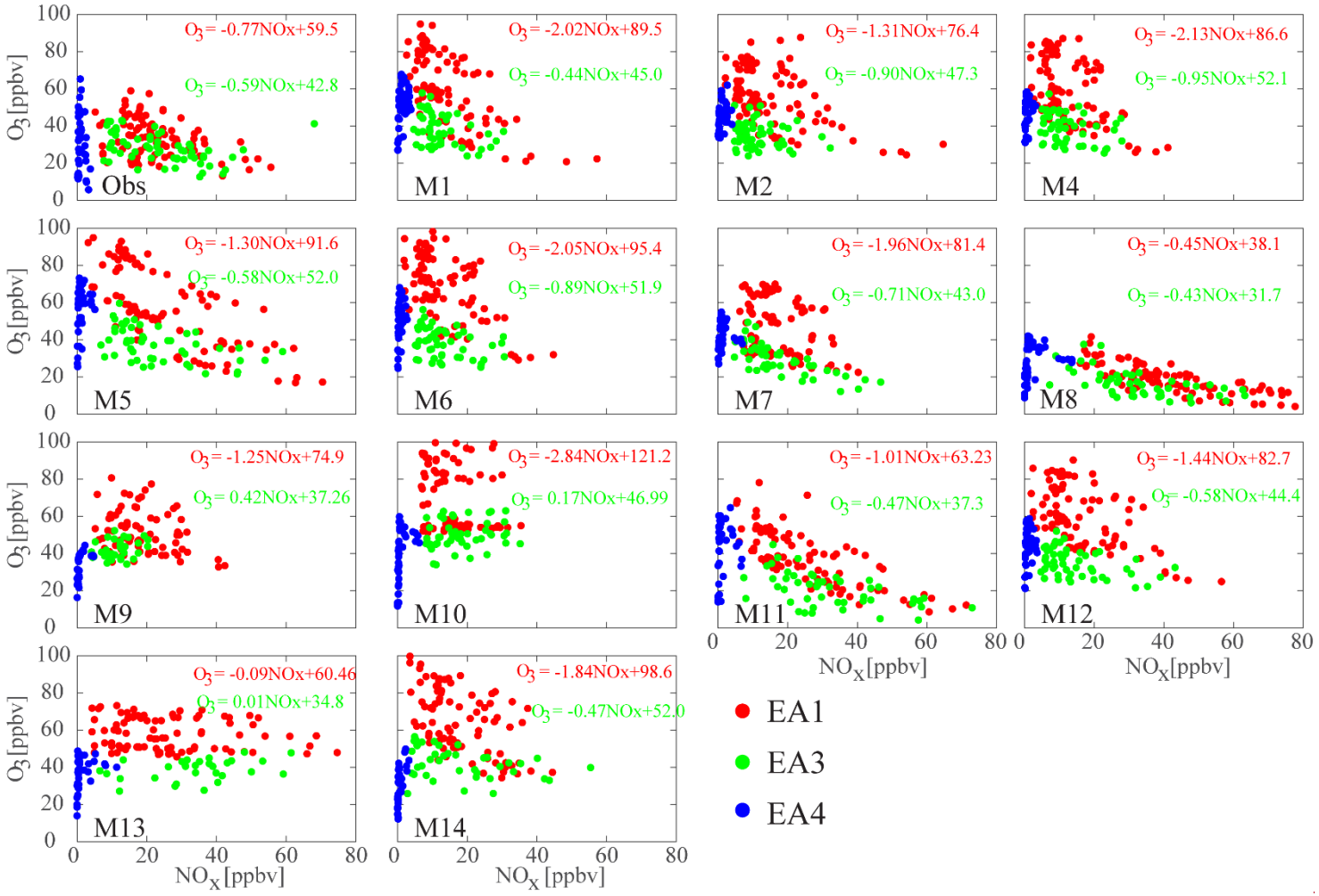

M1
M2
M4
M6
M11
M12
M13
M14
Ense
obs
EA1
EA2
EA3
TEX
STR
GGSEX
AMMA
V_d(cm/s)
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month

Fig.9 Li et al., 2018


Obs
O3= -0.77NOx+59.5
O3= -0.59NOx+42.8
M1
O3= -2.02NOx+89.5
O3= -0.44NOx+45.0
M2
O3= -1.31NOx+76.4
O3= -0.90NOx+47.3
M4
O3= -2.13NOx+86.6
O3= -0.95NOx+52.1
M5
O3= -1.30NOx+91.6
O3= -0.58NOx+52.0
M6
O3= -2.05NOx+95.4
O3= -0.89NOx+51.9
M7
O3= -1.96NOx+81.4
O3= -0.71NOx+43.0
M8
O3= -0.45NOx+38.1
O3= -0.43NOx+31.7
M9
O3= -1.25NOx+74.9
O3= 0.42NOx+37.26
M10
O3= -2.84NOx+121.2
O3= 0.17NOx+46.99
M11
O3= -1.01NOx+63.23
O3= -0.47NOx+37.3
M12
O3= -1.44NOx+82.7
O3= -0.58NOx+44.4
M13
O3= -0.09NOx+60.46
O3= 0.01NOx+34.8
M14
O3= -1.84NOx+98.6
O3= -0.47NOx+52.0
O3[ppbv]
NOX[ppbv]
EA1
EA3
EA4

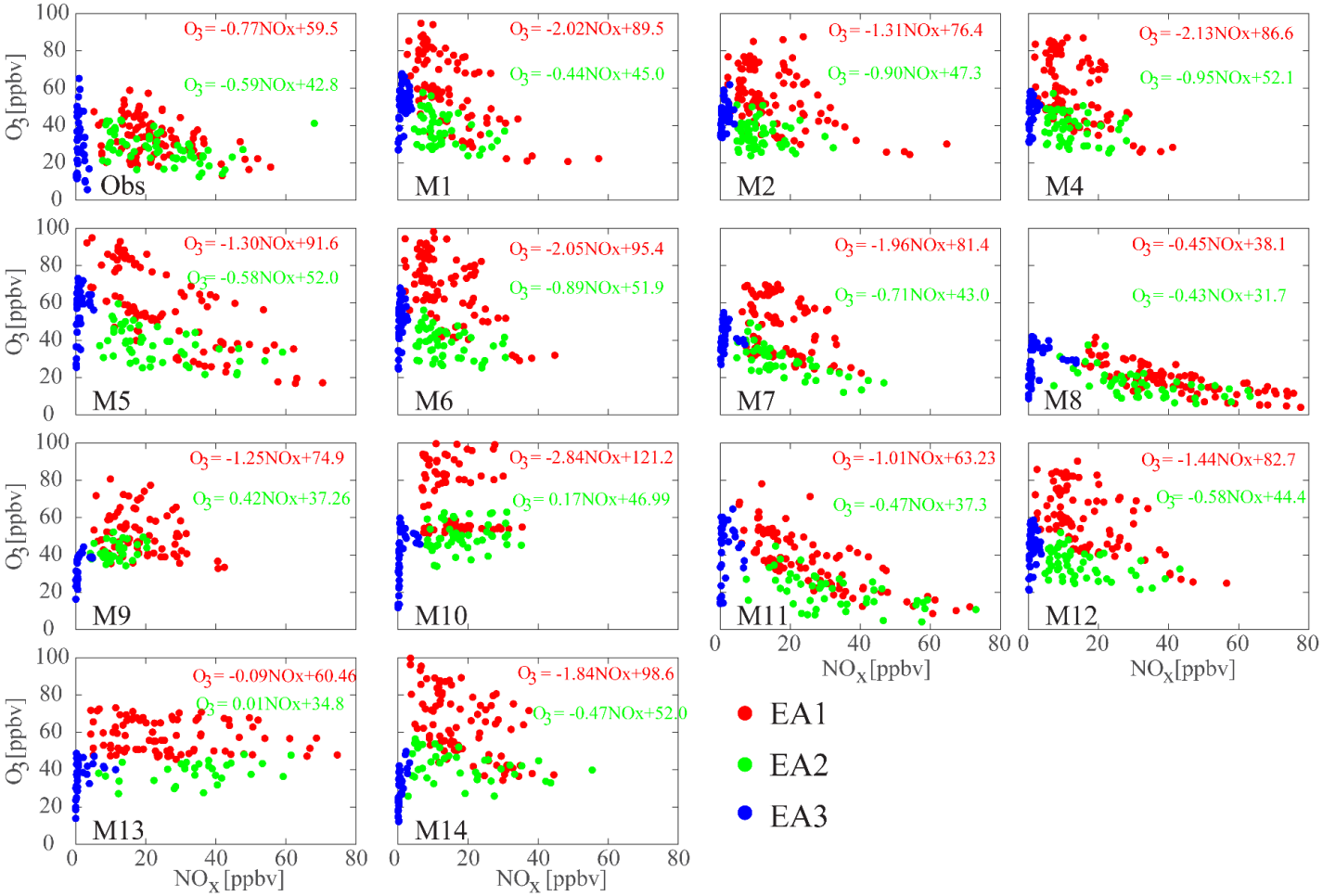


Fig.10 Li et al., 2018