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[1,3], 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], Zhe Wang[1], 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]Center for Excellence in Urban Atmospheric Environment, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen, 361021, China
[3]College of Earth Sciences, University of Chinese Academy of Sciences, Beijing, 100049, 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: Spatio-temporal variations of ozone (O_3) and nitrogen oxide (NO_x) mixing ratios from fourteen state-of-the-art chemical transport models (CTMs) are intercompared and evaluated 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 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 to the previous phase of MICS-Asia (MICS-Asia II), the evaluation with observations was extended to one-full year across China and the western Pacific Rim from four months. In general, model performance levels for O_3 varied widely by region and season. Most models captured key patterns 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 Pearl River Delta. A significant overestimation of surface O_3 was evident from May-September/October and from January-May over the North China Plain, western Pacific Rim and Pearl River Delta. Comparisons drawn from observations show that the considerable diversity in O_3 photochemical production partly contributed to this overestimation and to high levels of intermodel variability in O_3 for North China. In terms of O_3 soundings, the ensemble average of models reproduced the vertical structure for the western Pacific, but overestimated O_3 levels to below 800 hPa in the summer. In the industrialized Pearl River Delta, the ensemble average presented an overestimation for the lower troposphere and an underestimation for the middle troposphere. The ensemble average of 13 models for O_3 did not always exhibit a superior performance compared to certain individual models in contrast to its superior value for Europe. This finding suggests that the spread of ensemble-model values does not represent all uncertainties of O_3 or that most MICS-Asia III models missed key processes. This study improved the performance of modeling O_3 in March at Japanese sites than MICS-Asia II. However, it overpredicted surface O_3 concentrations for western Japan in July, which was not found 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 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 days in megacities (e.g. Beijing, Shanghai Guangzhou in China) values exceed air quality standard of World Health Organization (100 $\mu g/m^3$) for 8-hour average surface O_3 concentration (Wang et al., 2017). High O_3 concentrations have received more attention from the public and 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 under which a limit on the 8-hour O_3 maximum was set for the first time. However, these measures do not prevent the persistent increase of the ground-level O_3 in East Asia. The averaged 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 main method used for the detailed evaluation of 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 and have been widely used to simulate the O_3 formation process and to evaluate strategies for its control (Streets et al., 2007; 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., 2010). 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., 2010; He et al., 2017), have assessed the contributions of international and regional transport (Streets et al., 2008; Li et al., 2008), and have predicted O_3 mixing ratios under 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 emissions, chemistry, wet and dry deposition, and transport

can hardly be addressed using 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 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, to explore the potential sources of model uncertainties regarding sulfur, O_3, nitrogen compounds and aerosols (Carmichael et al., 2002, 2008). The study shows that the predicted temporal variations of surface O_3 in eight regional CTMs generally tended to be lower than 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 vary greatly in southern China. Inconsistencies in horizontal grids, emissions and meteorological inputs used among models have rendered explaining intermodel variability in MICS-Asia II results more difficult. More importantly, model evaluations for industrialized China have not been conducted due to a lack of observations, which has been detrimental to efforts made to improve O_3 model performance levels.

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, gas-phase chemistry mechanisms of Models 3-Community Multiscale Air Quality (CMAQ) have been evolved into CBM05 and SAPRC07 from CB04 and SAPRC99. It is critical to evaluate the updated models' abilities to simulate current air quality levels over East Asia. In 2010, MICS-Asia was expanded to Phase III wherein, 13 regional CTMs and 1 global CTM were run over one-full year by 14 independent groups from East Asia and North America, using a common reference model input dataset (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 for MICS-Asia II, new observational data from China were made available for MICS-Asia III and 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 for one full year had never before been performed,

creating a broader database to use for comparisons. The completeness of MICS-Asia III is therefore unique.

In this paper, we mainly evaluate the capacities of participating models in MICS-Asia III to simulate concentrations of O_3 and its related species within the MICS-Asia III framework. The following questions are addressed: (1) How well do various air quality models 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 simulation accuracy? This paper is expected to provide valuable insights into the capacities and limitations of CTMs when applied to East Asia.

2. Models and data

2.1 Experimental set up

In this study, all participating models were run for the year 2010 and provided gridded monthly mean diurnal mixing ratios of O_3 and its precursors in the lowest model layer. For O_3, monthly three-dimensional data were also submitted.

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 literatures and have been widely applied in modeling studies of East Asia. Table 1 does not list model names to maintain each model's anonymity. Similar behavior was observed from 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 is 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 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 (10 hPa). Standard meteorological fields were applied by the majority of groups. Several other models were employed to perform 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 considered the feedback of pollutants to meteorological fields. Consequently, their meteorological fields 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 varied for each group, likely resulting in a diversified set of model outputs.

MICS-Asia III provided a set of monthly anthropogenic emission inventories for the year 2010 called 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 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., 2011; Lu and Streets, 2012), and the official Korean emission inventory from the Clean Air Policy Support System (CAPSS; Lee et al., 2011). Biogenic emissions were 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 regridding 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 MICS-ASIA III emission group directly prepared a gridded inventory according to the configuration of each CTM. NMVOC emissions were 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 emission group. M13 and M14 make the projections by themselves. More information can be found in Li et al. (2017) and Gao et al. (2018).

MICS-Asia III also provided two sets of chemical concentrations for the top and lateral boundaries of the model domain, which were derived from 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 with 32 vertical layers by Nagoya University. Some models applied boundary conditions depending on their own past experiences.

2.3 Observational data for O_3

In this study, East Asia was divided into three subregions as shown in Fig. 1. The selection of subregions was based on emission, climate and observation data coverage. The North China Plain (EA1) and Pearl River Delta (EA2) represent highly industrialized regions of the mid-latitudes. EA1 is characterized by a temperate and tropical continental monsoon climate with marked seasonality. EA2 is located in southern China and is less affected by continental air masses. EA3 covers 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 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 includes 19 surface stations covering an area of 500 × 500 km^2 across North China Plain (EA1 sub-region; Ji et al., 2012). The stations are set up according to United States Environmental Protection Agency method designations. Half of them are remote, rural, suburban and clear urban sites. Nine sites are located within meteorological stations or on campuses of universities in urban regions, with little influence from local sources and sinks. The PRD RAQMN was jointly

established by the governments of Guangdong Province and the Hong Kong Special Administrative Region and consists of 16 automatic air quality monitoring stations located across the EA2 subregion (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 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. 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 subregion) were selected to evaluate model performance levels 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 provides detailed site description. Our comparisons of NO_x and VOCs emission rates conducted on grids for these stations at 45 km and 3 km resolution emission inventories suggest that our selected stations have rarely received local emissions.

O_3 was 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-NO_2-NO_x Analyzer with chemiluminescence technology at 40 sites in all three networks (CERN, PRD-RAQMN and EANET). NO_x measurements exhibited sometimes biases (especially for stations located far from sources) when using molybdenum converter devices since all nitrogen oxides were measured. This bias was found to be dependent on chemical conditions. A one-month continuous measurement collected in August with a chemiluminescence analyzer and Aerodyne Cavity Attenuated Phase Shift Spectroscopy (CAPS) from an urban site in Beijing shows that this bias from the chemiluminescence analyzer was minor when NO_2 concentrations exceeded 10-15 parts per billion by volume (ppbv), ranging from 10% to 30% under low NO_2 conditions (<10 ppbv) (Ge et al., 2013). Measurements collected from a rural site in South Korea revealed 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 performances for annual O_3, NO and NO_2 for three sub regions of East Asia. A box-and-whisker representation was used to show the frequency distribution of monthly concentrations measured from stations in each subregion. 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 cases (Table 2). Therefore, we only present multi-model mean ensemble results (Ense). In general, the majority of models significantly overestimated annual surface O_3 relative to observations in EA1, EA2 and EA3. Ense overestimated surface O_3 by 10-30 ppbv for these subregions. Ense NO_2 levels closely reflected observations to within ±20% across all subregions. In EA1 and EA2, Ense NO levels were found to be 5-10 ppbv lower than observations while exhibiting reasonable levels for EA3.

Of the models, M11 for subregions EA1 and EA2 and M7 in for EA2 and EA3 more closely reflected O_3 observations. M11 simulated O_3 with RMSEs of 9.5 ppbv and 13.3 ppbv for EA1 and EA2, respectively (Table 2). The models' performance in simulating O_3 was found to be closely related to their performance 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 models' performance regarding O_3 measurements to vary greatly for EA3, though M8 exhibited 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 subregions by 40-50%. The highest O_3 titration level observed in M8 may have generate lower O_3 levels than those indicated by other models and observations.

3.2 Monthly variations of surface O_3, NO and NO_2

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

For individual models, M11 achieved the highest degree of model reproductivity for monthly mean O_3 levels in EA1. Most of the other models overestimated O_3 by 100-200% for May-October. The largest levels of model bias and intermodel variability for NO and NO_2 appeared in the winter, and likely came from the NO_x vertical diffusion and heterogeneous chemistry (Akimoto et al., 2019). In EA2, M7 seems to have achieved the highest levels of O_3 reproducibility. Most of the models (except for M11 and M12) showed high O_3 concentrations for March-May and September-November. Observed O_3 values show that the highest concentrations appeared from October-November. M11 captured the observed January-May O_3 value due to relatively high NO concentrations. However, NO

was overestimated by M11 from May-September, leading to an underestimation of O_3 levels. In EA3, spatially averaged O_3 concentrations often differed by more than 20 ppbv in individual models. The highest levels of intermodel variability in O_3 values appeared from May-October, overestimating O_3 levels relative to observations by 10-40 ppbv. Interestingly, although M8, M9 and M14 exhibited similar magnitudes with observations for 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

Subregional O_3 diurnal variations are shown in Fig. 4. In general, model results for three subregions exhibited a larger spread with a magnitude of 10-50 ppbv across the diurnal cycle than those observed in Europe and North America (Solazzo et al., 2012). Summer Ense O_3 levels exhibited a systematic pattern of overestimation (20 ppbv) throughout the diurnal cycle in EA1.This indicates that the models had difficulty in estimating summer O_3 levels for the North China Plain. Compared to summer conditions, only a slightly systematic overestimation of Ense O_3 levels was observed for the other seasons (3-5 ppbv). In EA2, Ense O_3 levels generally agreed with summer, autumn and winter observations. In particular, the O_3 maximum occurring at around noon was reasonably reproduced. Only a 3-5 ppbv overestimation was observed from 16:00-23:00 and in early morning (6:00-10:00). In the spring, a systematic overestimation of Ense O_3 values was observed across the whole diurnal cycle (5-10 ppbv). In EA3, Ense captured the minor diurnal variations in O_3 across all four seasons, but significantly overestimated observations for the summer and autumn (5-20 ppbv). In the spring and winter, differences between Ense and observations fell within 5 ppbv.

Of all of the models, M11 exhibited the best model performance level in measuring peak daily O_3 concentrations of 60 ppbv from 14:00-16:00 in EA1, but it still overestimated nighttime O_3 levels by 10 ppbv. Compared to their performance in simulating summer patterns, the models performed significantly better in simulating winter conditions due to the weak intensity of photochemical reactions except in the case of M2, M10 and M8. Differences between observations and most simulations for both the nighttime and daytime fell within 5 ppbv. These differences in the models' performances between the summer and winter imply that the variety of chemistry parametrizations applied to different models partly explain the intermodel variability of simulated O_3 levels in EA1

(North China Plain). For EA2, the majority of models agreed well with diurnal variations occurring in the summer and autumn. However, most models exhibited a tendency to overestimate the O_3 concentrations for both the daytime and nighttime in the spring. The overestimated magnitudes exceeded 10 ppbv and 25 ppbv (of observed values of 20-35 ppbv) for the nighttime and daytime, respectively. M11 reproduced observed O_3 levels for the spring but underestimated O_3 levels for the summer and autumn. For EA3, significant levels of intermodel variability persisted throughout the year. Amplitudes of intermodel variability except for those of M8 and M14 reached approximately 20 ppbv and 10 ppbv in the spring-summer and autumn-winter, respectively. M8 and M14 generated the lowest O_3 values of the models for the whole year.

3.4 Error statistics on surface concentrations

In this section, we present statistics on the models' performance based on monthly values. Values are calculated with equations shown in Appendix A. On a yearly basis, all models showed the highest (0.8-0.9) and lowest (0.1-0.6) correlation coefficients for O_3 for EA1 and EA2, respectively (Table 2). High correlations were found in EA1 mainly because the summer-maximum and winter-minimum seasonal cycles are typical of polluted regions represented in all of the participating models. In general, Ense performed better than the individual models in representing NO_2 for East Asia, reproducing the observed seasonal cycles and magnitudes. However, Ense did not always exhibit a superior performance in simulating O_3 levels over individual models for East Asia, which stands in contrast to its performance for Europe (Table 2). M7 and M11 agreed well with observations for EA1 and EA2 while ENSE tended to overestimate O_3 concentrations for May-September in EA1 and for January-September in EA2. Loon et al. (2007) indicated that Ense exhibits a superior performance level only when the spread of ensemble-model values is representative of O_3 uncertainty. This indicates that most models do not reflect this uncertainty or miss key processes of MICS-Asia III.

Considerable overestimations made by most of the models for May-September led to high NMB (0.25-1.25) and RMSE (10-33 ppbv) values for EA1. M11 generated the lowest NMB (0.09) and RMSE (9.46 ppbv) values of the examined models. For EA2, M9 and M10 generated stronger correlations than the other models. However, their corresponding NMB and RMSE values were also the highest. These findings imply that systematic model biases are present in these two models. M7

exhibited lower NMB and RMSE values than the other models, but its correlation was measured as only 0.29. For EA3, correlations exhibited the largest degree of intermodel variability across all subregions, ranging from -0.13-0.65. M7 generated the lowest NMB and RMSE likely due to cancelling effect of its overestimation for the summer and underestimation for other seasons (Fig. 3).

For NO, model correlations for EA1 ranged from 0.57-0.68, showing that all of models effectively reproduced spatial variability in NO for this subregion (Table 3). NMBs indicated that underestimations by the models except in the case of M8, mostly occurred for the winter. This underestimation can be partly attributed to the coarse model horizontal resolution (45 km) used in 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 for all three subregions. It is noted that NO observations for EA3 were too low (<0.3 ppbv) to be discussed in this study.

Table 4 shows statistics on the models' performance in measuring NO_2 levels. In general, most of the models performed better in representing NO_2 than O_3 and NO for EA1. NMBs ranged from -0.28-0.32, falling far below those measured for O_3 (0.48-1.25). Correlations of 0.54-0.66 were recorded, implying the models' reliable performance in reproducing spatial and monthly variability of NO_2 for EA1. Similar to those for O_3 and NO, correlation coefficients for NO_2 in EA2 remained low. Thus, a dedicated investigation of O_3, NO and NO_2 levels in EA2 is urgently needed, but falls beyond the scope of this study. In EA3, correlation coefficients ranged from 0.43-0.72. NMBs and RMSEs except for those of M8 ranged from -0.23-0.46 and 0.90-1.79 ppbv, respectively.

3.5 Vertical profiles of O_3

Fig. 5 shows the vertical profiles of observed and simulated O_3 levels for East Asia for the summer and winter. Ensemble means (Ense) showed underestimations and overestimations of EA2 O_3 levels for the middle (500-800 hPa) and lower (below 900 hPa) troposphere, respectively. In the winter, underestimations extended to 200 hPa. Magnitudes of underestimations and overestimations reached 10-40 ppbv and 10-20 ppbv, respectively. For EA3, Ense reproduced the vertical structure of ozone for both the summer and winter. An overestimation of less than 800 hPa, with a magnitude of 10-20 ppbv was observed for the summer.

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

In the lower troposphere, intermodel variability in the summer was generally higher than that in the winter. In polluted regions (EA1), 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 planetary boundary layer (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 considerably to discrepancies between M1, M6 and M11. In EA2, vertical structures of O_3 among models were found to be consistent, but concentrations differed more than those in EA1. SD values covered 22% of mean concentrations.

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

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

Fig. 6 shows 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 CV value, the lower the degree of consistency among the models. For the summer, Ense predicted an elevated O_3 concentration belt in the middle-latitudes (30°-45°N). A region of O_3 in excess of 60 ppbv stretched across the North China Plain and East China Sea, far exceeding MICS-Asia II (45-50 ppbv) values for 2001 (Han et al., 2008). In other seasons, the O_3 distribution showed higher O_3 over the ocean than in eastern China, reflecting O_3 titration from high

NO_x emissions. Due to the stratospheric injection, surface O_3 over the Tibet Plateau remained at high levels throughout the year, ranging from 50 to 65 ppbv. The seasonal cycle of surface O_3 levels determined from Ense via MICS-Asia III agreed with that 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 high values were found in EA1 in the winter. These high values in the low-latitude western Pacific (10ºS-15ºN) and Indian Oceans were likely caused by the treatment of lateral boundaries in the models. For MICS-Asia III, M7, M8 and M9 employed the default model configurations, and the others employed outputs of the GEOS-Chem/CHASER/MOZART-GOCART global model. Compared to those of MICS-Asia II, CVs for the Asian continent except for the winter remained at 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 (30º-45ºN), the magnitudes of enhanced O_3 levels varied between the models (Fig. 7). M5 predicted the highest O_3 concentrations of 60-90 ppbv 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 supplementary section) whereas M8 predicted the lowest levels of 35-50 ppbv. Overhangs of 30 ppbv contour lines extending into the northwestern Pacific along the Asian continent outflow plume differed considerably between the models. A plume of 30 ppbv or higher O_3 levels was simulated in M1-M6, M13 and M14, reaching further south and east of Japan (135ºE, 20ºN), than those of M8, M10 and M11 (120ºE, 30ºN). From MICS-Asia II and HTAP, differences in the frequency of marine air masses from the western Pacific Ocean were thought to be a possible cause of O_3 discrepancies observed over oceans between the models due to different meteorological drivers (Han et al., 2008). For MICS-Asia III, wind fields employed by the models were similar due to the use of the same or similar meteorological fields (Fig. S2 in the supplementary section). These inconsistencies between the models resulted from the combined effects of a series of factors, including the diversity of condensed gas-chemical mechanism and heterogeneous chemistry. Li et al. (2016) found chemical productions to be the dominated controlling factor of O_3 along outflow pathways near the North China Plain in the summer rather than

lateral and top boundary conditions. Impacts of aerosols on O_3 in these regions were frequently reported in Olson et al. (1997) and Li et al. (2018) to alter photolysis rates and heterogeneous chemistry patterns. Detailed comparisons of parameterizations of these processes in models are needed in future intermodel comparison projects focused on Asia.

In the winter, distribution patterns of O_3 were quite similar between the models with high concentrations observed over parts of western China, northeastern India and the western Pacific from the East China Sea to southern Japan (Fig. S3 in the supplementary section). In the spring and autumn (Fig. S4 and Fig. S5 in the supplements), O_3 concentrations were generally higher than they were in the winter across the whole model domain due 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 winter. Increases of O_3 in further north of Japan were comparable with winter.

4.2 Comparison with MICS-Asia II

From MICS-Asia II, model evaluation on O_3 were conducted on sites in the western Pacific. Fig. 8 presents the simulated and observed surface O_3 levels at these monitoring sites 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 distributions of O_3 mixing ratios at most sites in both MICS-Asia II and III. ENSE results were consistent for March and December of 2001 and 2010. Underestimations of O_3 levels in March at Japan sites (site 4: Oki, site 5: Hedo and site 6: Banryu) in Phase II were largely remedied in Phase III. However, surface O_3 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 was not found in Phase II, for which differences from observations were valued at approximately 5 ppbv. Rural sites in western Japan are located in the upwind regions of Japanese domestic emissions, and are subjected to the impacts of Asian continent outflows. Overestimated O_3 values for North China Plain (EA1) in Phase III contributed considerably to enhanced concentrations simulated for sites of western Japan in July 2010. This indicates that transboundary transport from the Asian continent according to MCIS-Asia III was likely overestimated relative to that measured from MICS-Asia II.

5. Discussions

In reference to MICS-Asia II, Han et al. (2008) hypothesized that variations in meteorological fields, dry deposition, PBL, model treatments of chemistry and other physical processes had contributed to model biases in relation to observations and intermodel variability. Quantifying the contributions of these processes can help explain model biases through sensitivity simulations. However, this task comes with tremendous computational costs when applied to 14 models. The qualitative analysis of potential causes of these processes based on comparisons of models and observations is essential to selecting sensitivity simulating scenarios for the next phase of MICS-Asia. In MICS-Asia III, common input data (emission and meteorology) were effectively used in this qualitative analysis based on model parameterizations. We evaluated the models on dry deposition, PBL and chemistry by collecting their observations (dry deposition velocity and PBL height). This work was not conducted under MICS-Asia II and is intended to help model developers improve model performance for East Asia.

5.1 Dry deposition

Previous studies show that dry deposition processes serve as the key net sink of O_3, accounting for 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):

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

where F and C represent the simulated dry deposition flux and surface O_3 concentrations, respectively. We determined spatial mean dry deposition velocities from stations in each subregion.

Fig. 9 presents the simulated and observed monthly spatial mean dry deposition velocities of O_3. For EA1, ensemble mean values overestimated observed dry deposition velocities of O_3 (v_d) for August-September, but still fell 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, the lower dry deposition velocities in May-July for M1, M2, M4 and M6 than that of M11 partly explained higher May-July surface O_3 from those simulations than that 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). 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 North 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.

For EA2, similar features as those of EA1 were found. M1, M2, M4 and M6 were quite consistent with each other, with a seasonal cycle and a spring minimum. M11, M12 and M14 show no obvious signs of seasonal variability with a magnitude of 0.1-0.2 cm/s. Seasonal patterns in M13 are considerably different from 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 remained high under higher dry deposition velocities.

In EA3, most stations were located in remote oceanic sites, and few dry deposition observations were made. Thus, we collected observations from other oceanic sites to evaluate model performance (Helmig et al., 2012). Ense values for v_d agreed reasonably well with observations (Fig. 9). Both observations and simulated v_d values showed a July-September maximum with a magnitude of 0.02-0.03 cm/s. Park et al. (2014) found surface O_3 levels in EA3 to be more sensitive to dry deposition parameterization schemes in CTMs. O_3 measured from oceans differed by 5-15 ppbv in East Asia due to the use of various dry deposition parameterization schemes. Thus, more observations are needed over oceans in EA3 to mitigate O_3 simulation uncertainties.

5.2 Relationships between surface NO_x and O_3

In general, surface O_3 mainly comes from photochemistry processes involving NO_x and VOCs in polluted regions. Examining O_3-NO_x relationships is effectual to investigating sources of intermodel variability and model errors concerning O_3 chemistry in East Asia. Fig. 10 presents 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 for the stations shown in Fig. 1.

For EA1 (North China Plain), observations clearly show that O_3 concentrations decreased with an increase in NO_x concentrations. O_3 concentrations mostly remained at high levels (40-60 ppbv) when NO_x was less than 20 ppbv. 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 results were in relative agreement with observations. The slope and intercept (-1.01 ppbv/ppbv, 63.23 ppbv) reflected the observations. The other models showed a higher degree of model bias and intermodel variability in relationships between O_3 and NO_x. Their slopes mostly ranged 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, far exceeding observations (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 were relative agreement with observations.

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 low NO_x conditions (< 30 ppbv) but failed to capture low O_3 under high NO_x conditions (30-40 ppbv), accounting for overestimations of these models for O_3 in May-September. By contrast, M8 and M11 produced excessively high NO_x values, resulting in their underestimations of O_3 values. 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 stands in contrast to observations. This finding suggests that more attentions are needed 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, showing that local chemistry was not a key factor shaping O_3 formation. Thus, we did not examine the simulated O_3-NO_x relationship further.

5.3 Other factors

Previous studies show that O_3 precursors are mostly constrained within the boundary layer (Quan et al., 2013). The planetary boundary layer height (PBLH) model evaluation is essential for the interpretation of model biases with observations. Unfortunately, this evaluation was not applied in MICS-Asia II. In 2016, Guo et al. (2016) 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 for EA1 (Fig. S6 in the supplementary section), capturing the main climatological pattern of PBLH observations (Guo et al., 2016). The Ense on PBLH only overestimated radiosonde measurements by 100-200 m (~10-15%) likely due to sampling bias between the models and measurements. The simulation was recorded as the mean value of 12 hours (08:00-20:00), while the average of the measurements was calculated based on a 3-hour period (08:00, 14:00 and 20:00). For EA2, the observed PBLH did not vary as much as that for EA1, and differences between seasons ranged within 100 m. This pattern was captured by the models. As was observed from EA1, the simulated PBLH for EA2 exceeded the measurements by100-200 m. Few measurements of remote oceanic sites in East Asia were collected. Thus, 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 PBLH.

6. Summary

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

A model ensemble was produced and evaluated. In general, the model ensemble captured key patterns of monthly and diurnal O_3, NO and NO_2 in the North China Plain and western Pacific Rim. It failed to capture the observed seasonal cycle of O_3 for the Pearl River Delta of China. For the North China Plain and western Pacific Rim, the model ensemble severely overestimated surface O_3 levels for May-September by 10-30 ppbv. This overestimation systematically appeared in both daytime and nighttime. Similarly, the model ensemble tended to overestimate spring daytime and nighttime O_3 concentrations for the Peral River Delta. Compared to MICS-Asia II, MICS-Asia III was less prone to underestimating surface O_3 in March for Japanese sites. However, it predicted excessively high surface O_3 concentrations for western Japan in July, which was not the case 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 hPa in the summer. For the industrialized Pearl River Delta, the ensemble average presented an overestimation of O_3 levels for the lower troposphere and underestimations in the middle troposphere. We find that the ensemble average of 13 models for O_3 does not always perform better than individual models for East Asia in contrast to their performance for Europe. This suggest that the spread of ensemble-model values does not represent all uncertainties in O_3 levels or that most MICS-Asia III models missed key processes. In contrast to performance levels for O_3, Ense performed better than individual models for NO_2 in East Asia.

MICS-Asia II outlines potential causes of variability among models. Quantifying the contributions of these processes to O_3 concentrations serves as an effective way to explain model biases through sensitivity simulations. However, this would incur tremendous computational costs when applied to 14 models. In this study, we conducted a qualitative analysis of potential causes by

comparing models and observations for these processes to identify sensitivity simulating scenarios for the next phase of MICS-Asia. Our comparisons show that the ensemble model overestimated observed dry deposition velocities of O_3 for August-September in North China Plain, showing that other factors rather than dry deposition may contribute to the overestimation of simulated O_3 concentrations in the summer. For the western Pacific, simulated v_d values agreed with observations reasonably well. Photochemical treatment in models may contribute to O_3 overestimations in North China Plain. The studied models captured major climatological pattern of PBLH observations for three subregions of East Asia. More evaluations of turbulent kinetic energy in PBL are needed to assess 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, 2018YFC0213205). 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)

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

Root mean square error (RMSE):

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

Normalized Mean Bias (NMB)

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

References:

Ackermann, I. J., Hass, H., Memmesheimer, M., Ebel, A., Binkowski, F.S., and Shankar, U.: Modal aerosol dynamics model for Europe: Development and first applications, *Atmos. Environ.*, 32（17), 2981-2999, 1998.

Akimoto, H., Mori, Y., Sasaki, K., Nakanishi, H., Ohizumi, T., and 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., and Ohizumi, T.: Long-term assessment of nitrogen deposition at remote EANET sites in Japan, *Atmos. Environ*., 146, 70-78, 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. and Schere, K. L.: Rewiew of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system, *Appl. Mech. Rev.,* 59, 51–77, 2006.

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., and 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., and 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.

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 attenuated phase shift spectroscopy (CAPS) and implications in ozone production efficiency and nitrate formation 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., and 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., and 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., and 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., and 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.

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.

Jung, J., Lee, J. Y., Kim, B. M., and 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.

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., and 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., and 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., and 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.

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., and 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. and 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., and 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., and 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., and Guo, Y.: An observation study of ozone dry deposition over grassland in the suburban area of Beijing, Chinese Journal of Atmospheric Sciences (in Chinese), 34(1), 120-130, 2010.

Park, R. J., Hong, S. K., Kwon, H. A., Kim, S., Guenther, A., Woo, J. H., and Loughner, C. P.: An evaluation of ozone dry deposition simulations in East Asia, *Atmos. Chem. Phys.*, 14(15): 7929-7940, 2014.

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., and 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.

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., and 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., and 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., and 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., and 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, X, Wang, Z., Zhu, J., Abaguidi, A., Wu, Q., Li, J., and 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, 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., and 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., and 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 (WHO), 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., and 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., and Whitten, G.: Updates to the Carbon Bond Chemical Mechanism: CB05 Final Report to the US EPA, RT-0400675, 2005.

Zaveri, R. A. and Peters, L. K.: A new lumped structure photochemical mechanism for large-scale applications. *J. Geophys. Res*.104, 30387-30415, 1999.

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

Zhong, L., Louie, P. K., Zheng, J., Wai, K. M., Ho, J. W. K., Yuan, Z., Lau A. K. H., Yue D. L., and 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 subregions marked in this study. Also shown are the locations of surface monitoring stations used in this study. The meteorological model used to provide meteorological fields 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 determined from 13 models and averaged for stations in EA1, EA2 and EA3 in time for 2010. n denotes the numbers of stations. The rectangle represents the interquantile range (25^{th} to 75^{th} percentiles). The small star identifies the mean, the continuous horizontal line within the rectangle identifies the median, and whiskers extend between 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 across three subregions of East Asia (EA1: top row, EA2: middle row, EA3: bottom row). Observations are denoted by the black line. n represents the number of stations. Gray lines represent NO_2, NO and O_3 levels simulated by models except M1, M2, M4, M6, M11, M12, M13 and M14.

Fig. 4 Seasonal mean diurnal cycle of surface O_3, in ppbv, as a function of hours, for all models and their ensembles, averaged across all observed stations in three subregions of East Asia (EA1: top row, EA2: middle row, EA3: bottom row). Observations are denoted by the black line. n represents the number of stations. Gray lines represent O_3 levels simulated by models except M1, M2, M4, M11, M12 and M14. Spring, summer, autumn and winter were defined as time periods of March-April-May, June-July-August, September-October-November, and December-January-February, respectively.

Fig. 5 Simulated O_3 profiles for the summer (June-July-August) and winter (December-January-February) of 2010, averaged over all observed stations across three subregions of East Asia (EA1: left column, EA2: middle column, EA3: bottom column). Ozonesonde data for 2010 were taken from World Ozone and Ultraviolet Radiation Data Centre (WOUDC) database

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

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

Fig. 8 Modeled and observed monthly mean concentrations of O_3 for EANET sites in the phase II (left panel) and III (right panel) of the MICS-ASIA project. The solid line represents the ensemble mean. Note that MCIS-ASIA II and III data refer to March, July and December of 2001 and 2010, respectively. IDs of the monitoring sites denote 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 M1, M2, M4, M6, M11, M12, M13 and M14 for three subregions of East Asia (EA1: top row, EA3: middle row, EA3: bottom row). TEX, STR, GGSEX and AMMA denote observations for 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 from Sorimachi et al. (2003), Pan et al. (2010), and Helmig et al. (2012).

Fig. 10 Scatter plots for monthly daytime (08:00-20:00) surface NO_x and O_3 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 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	0.14	11.90		0.43	0.15	10.97		0.72	2.38	4.46
M9		0.60	0.32	18.80		0.51	-0.37	12.66		0.49	0.05	1.66
M10		0.61	0.11	10.65		0.15	-0.08	12.81		0.63	0.06	1.33

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

a: n represents the numbers of observation stations

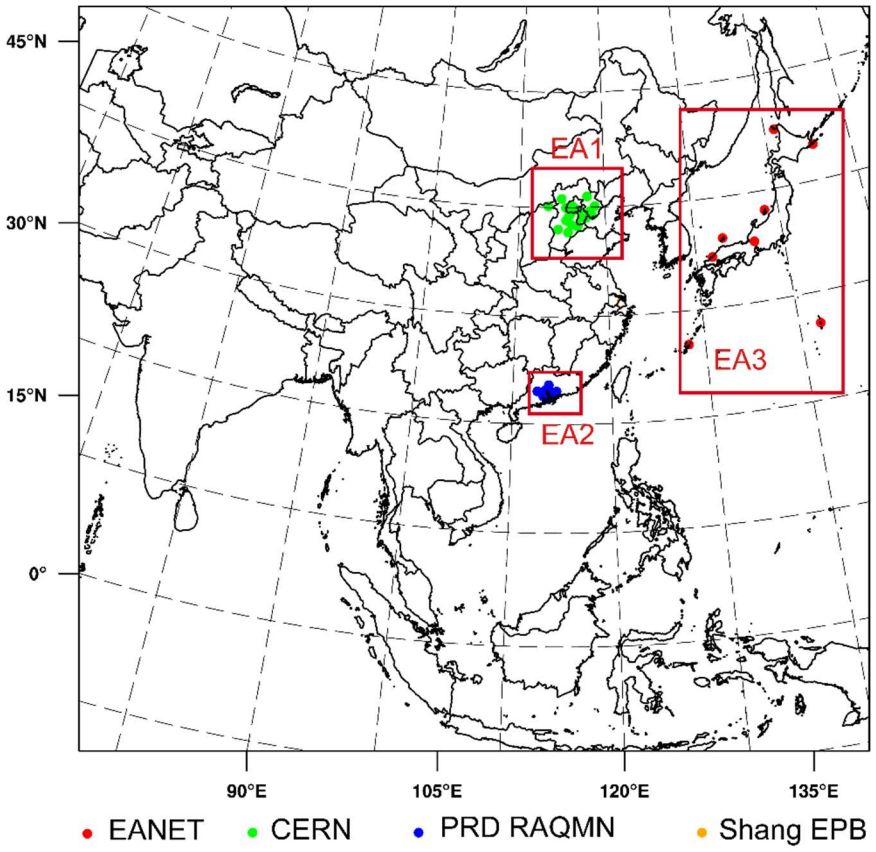

Fig.1 Li et al., 2018

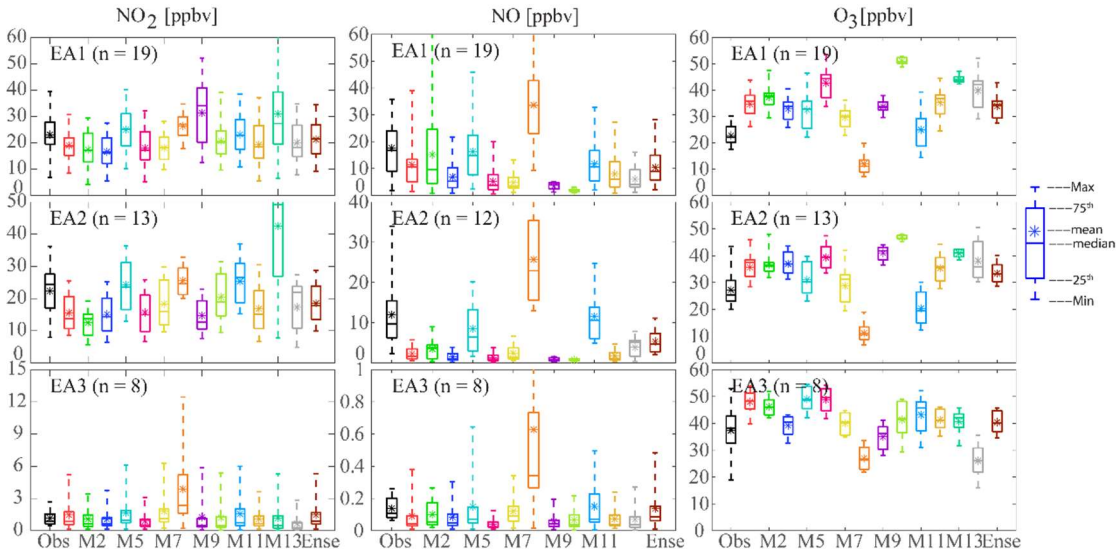

Fig.2 Li et al., 2018

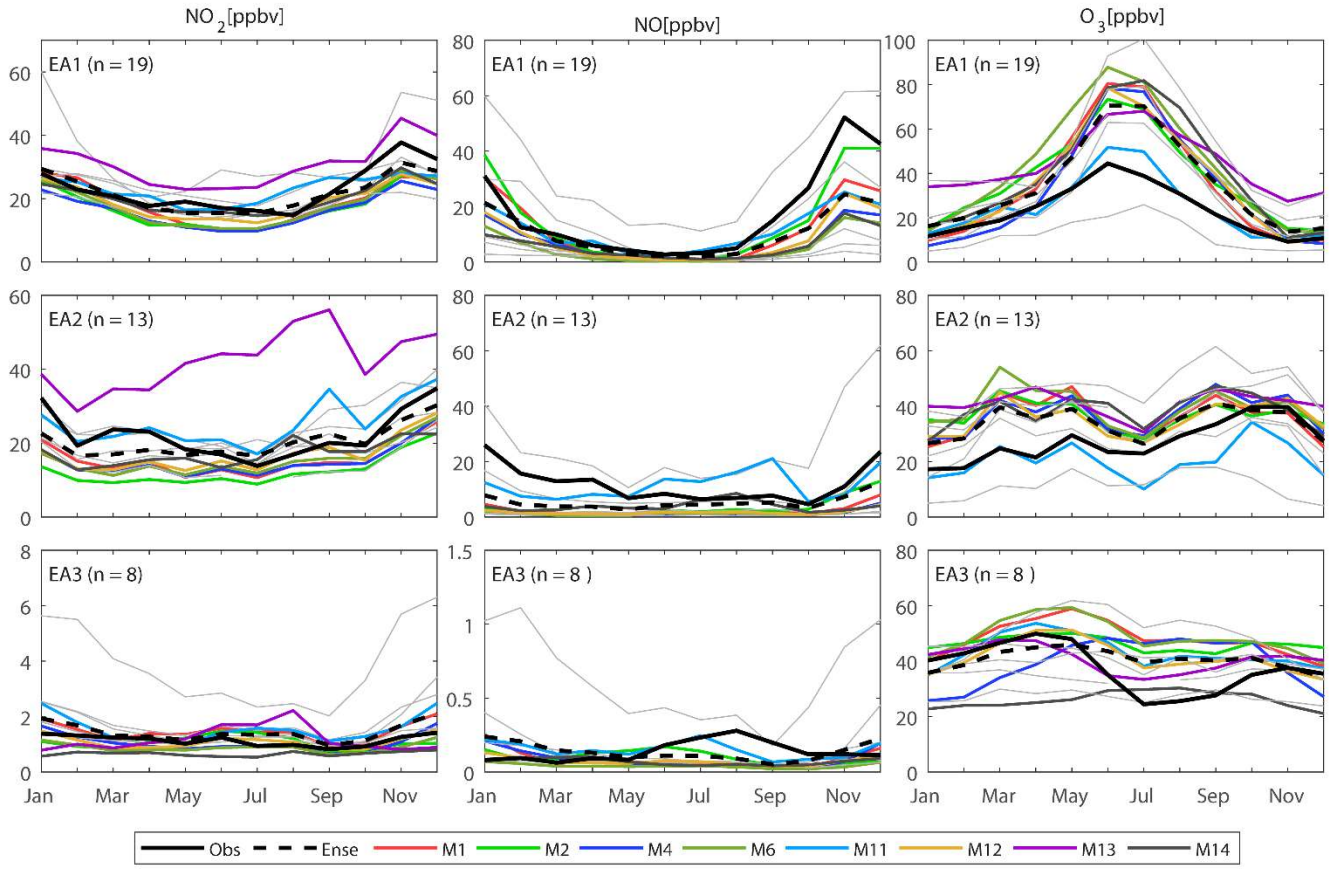

Fig.3 Li et al., 2018

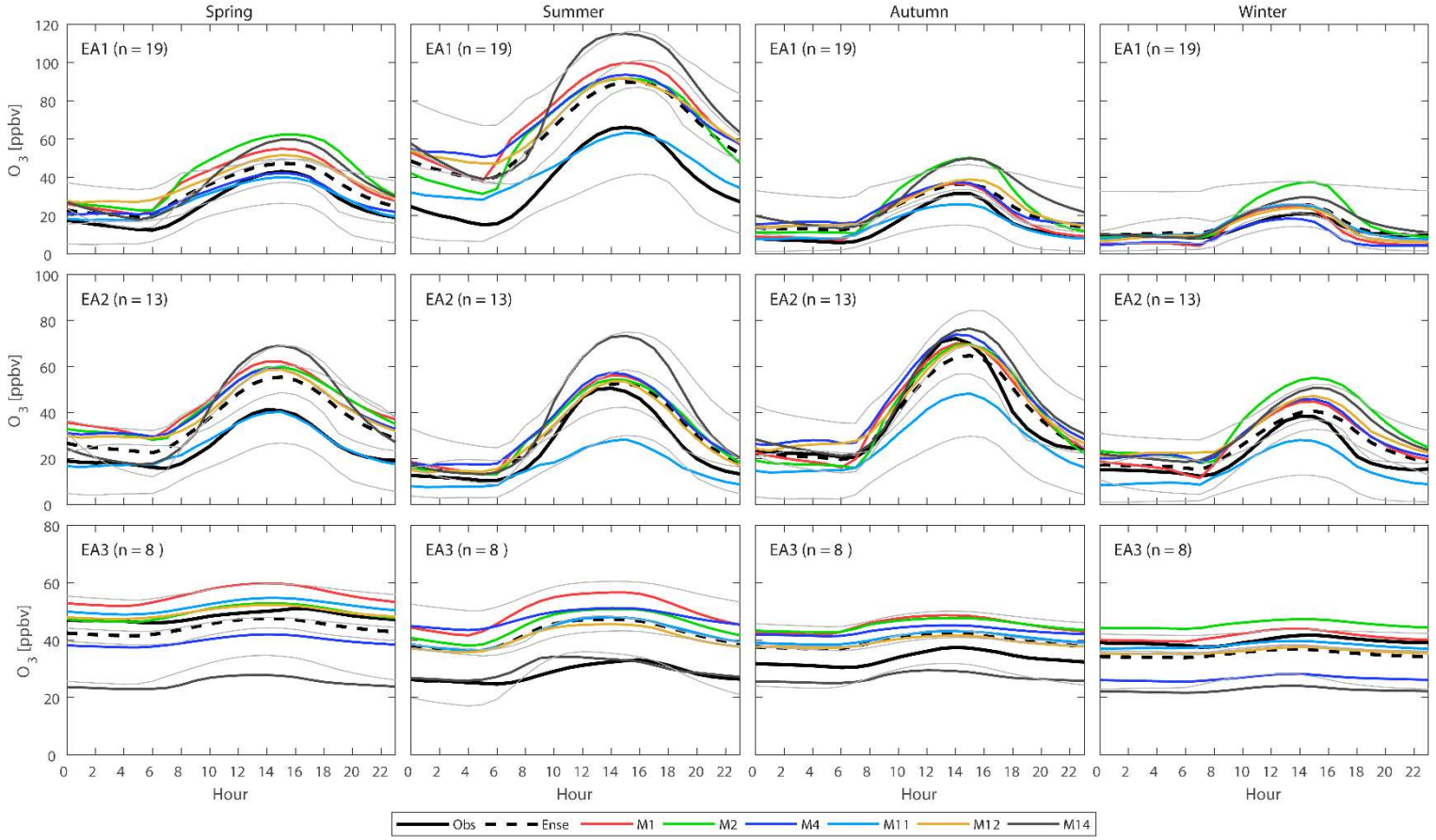

Fig.4 Li et al., 2018

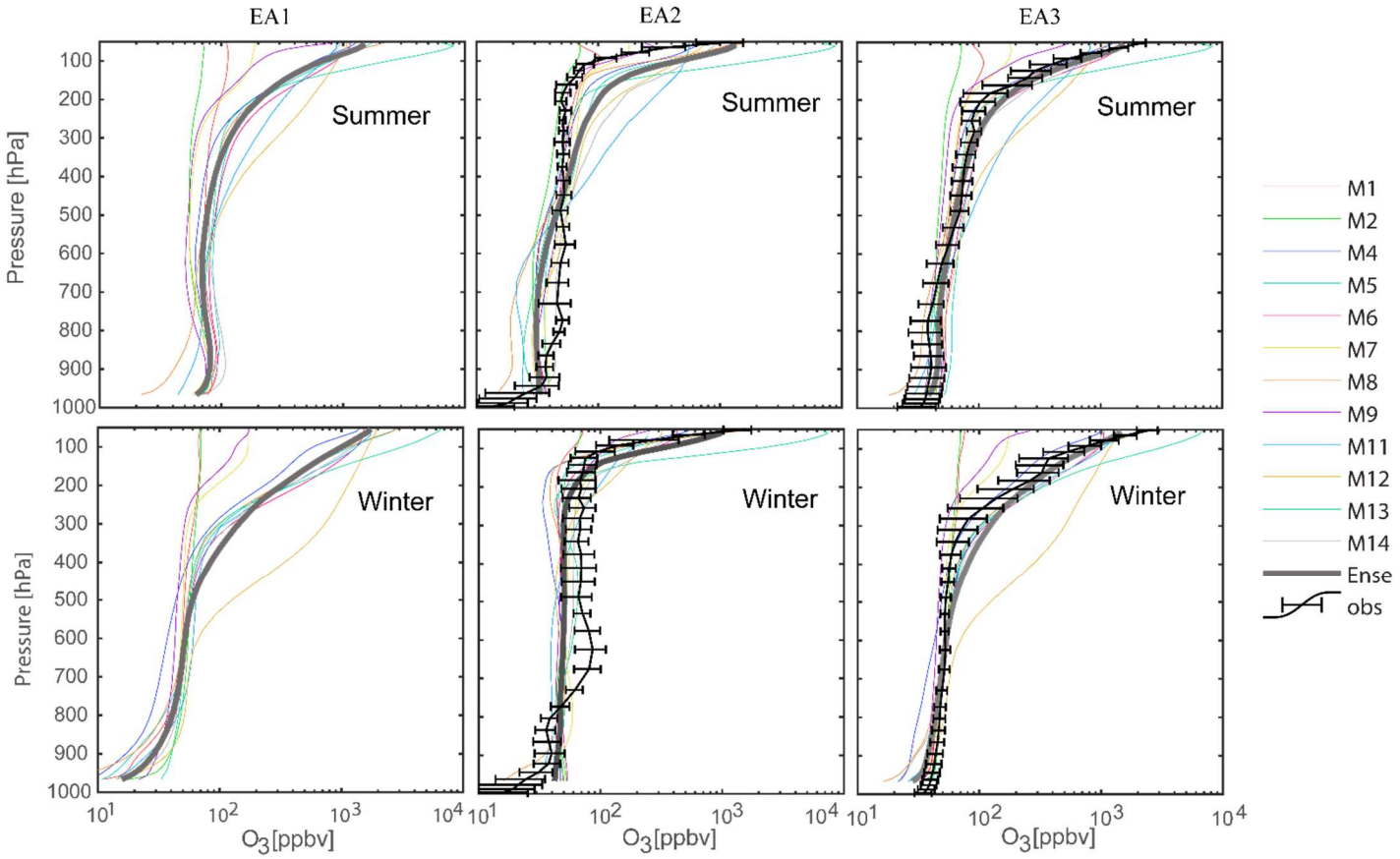

Fig.5 Li et al., 2018

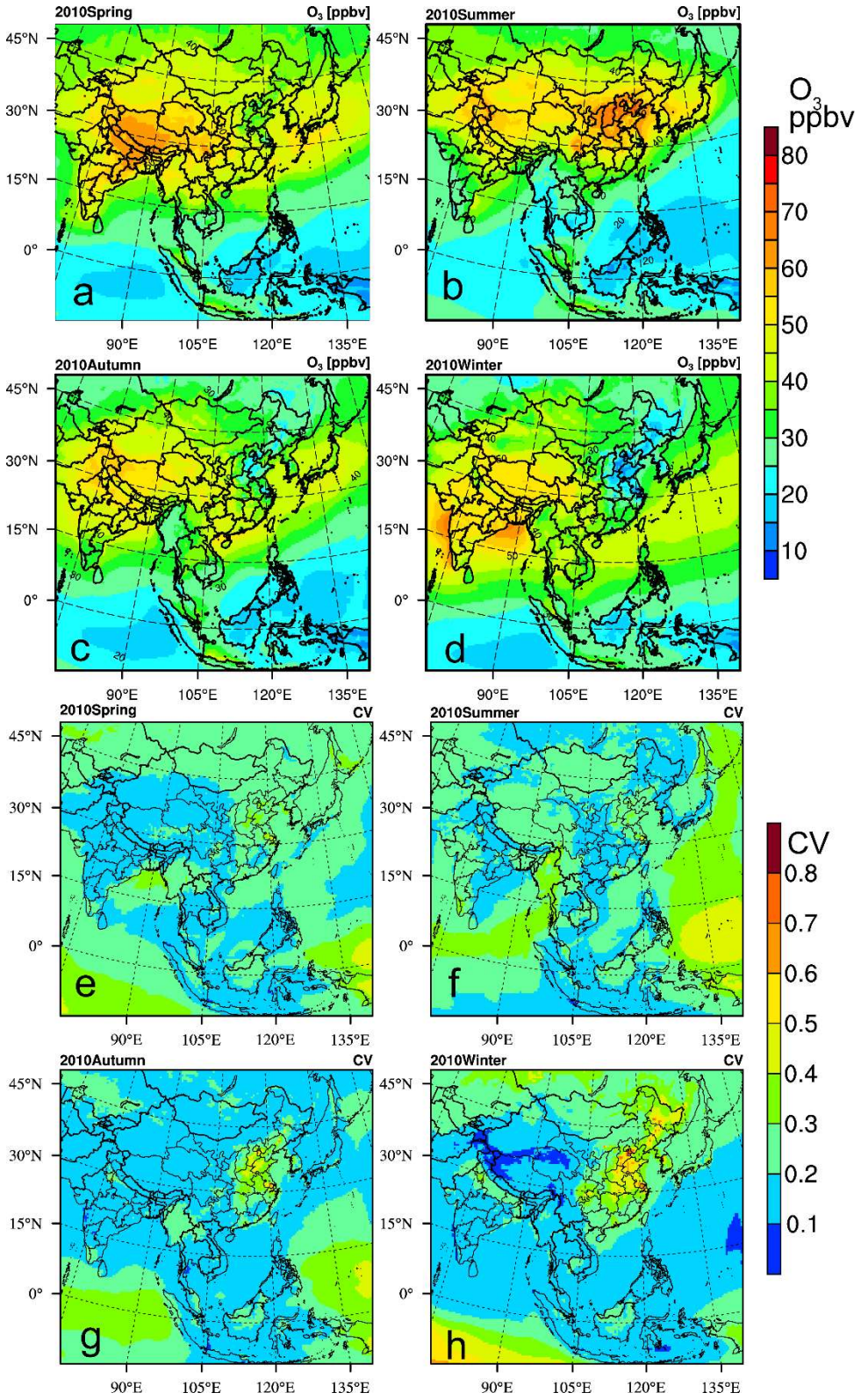

Fig.6 Li et al., 2018

.

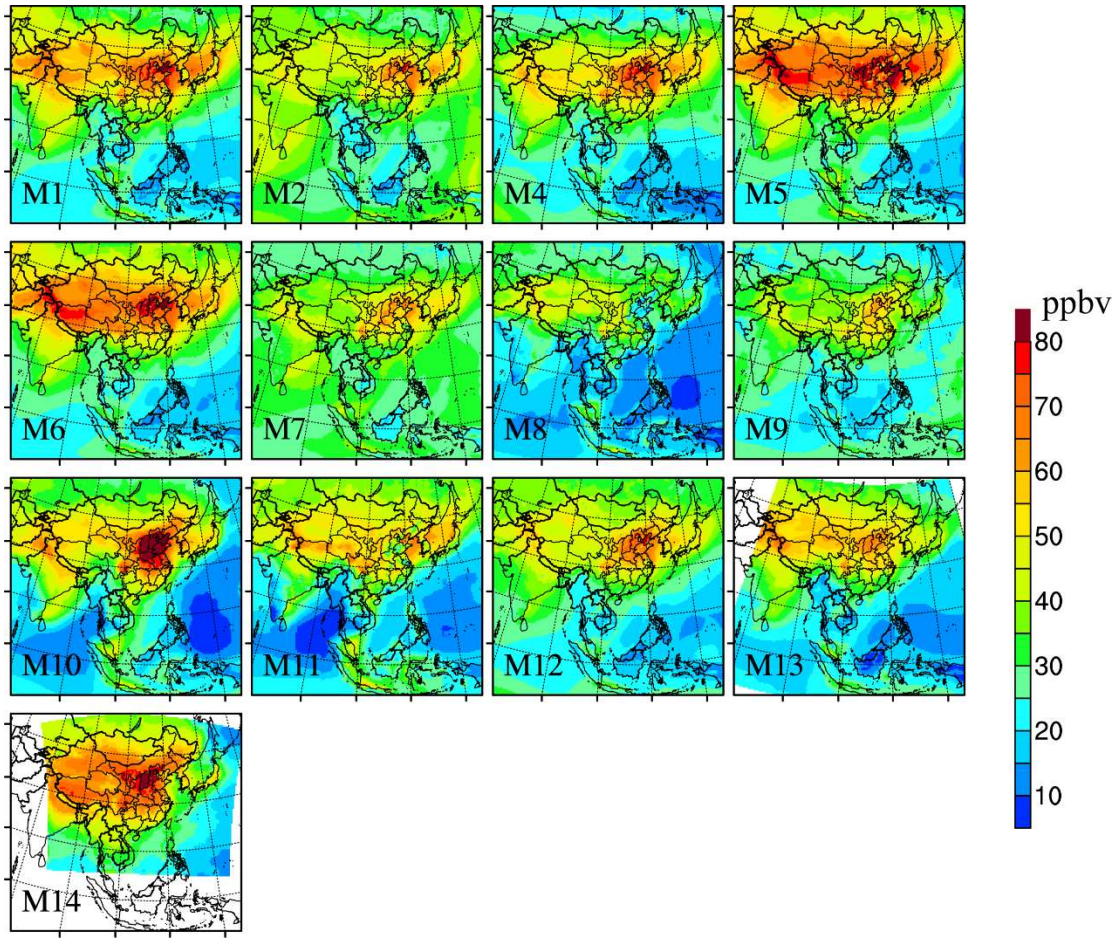

Fig.7 Li et al., 2018

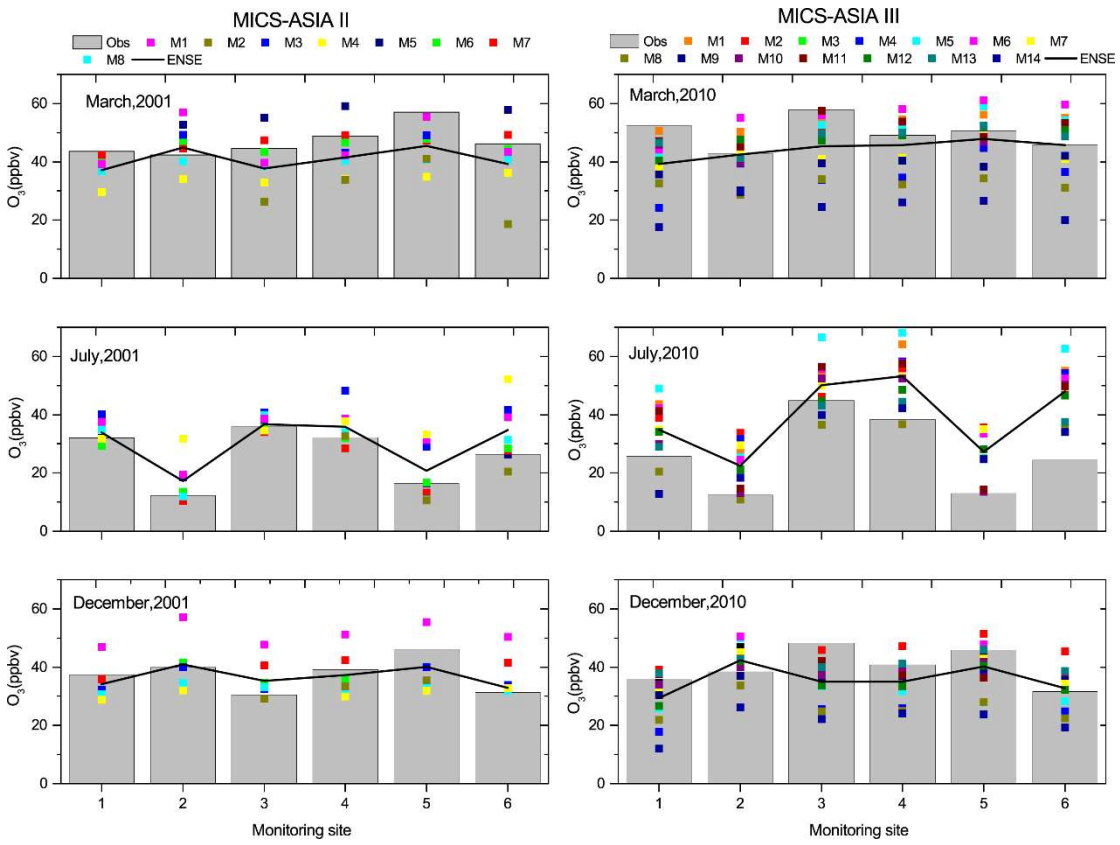

Fig.8 Li et al., 2018

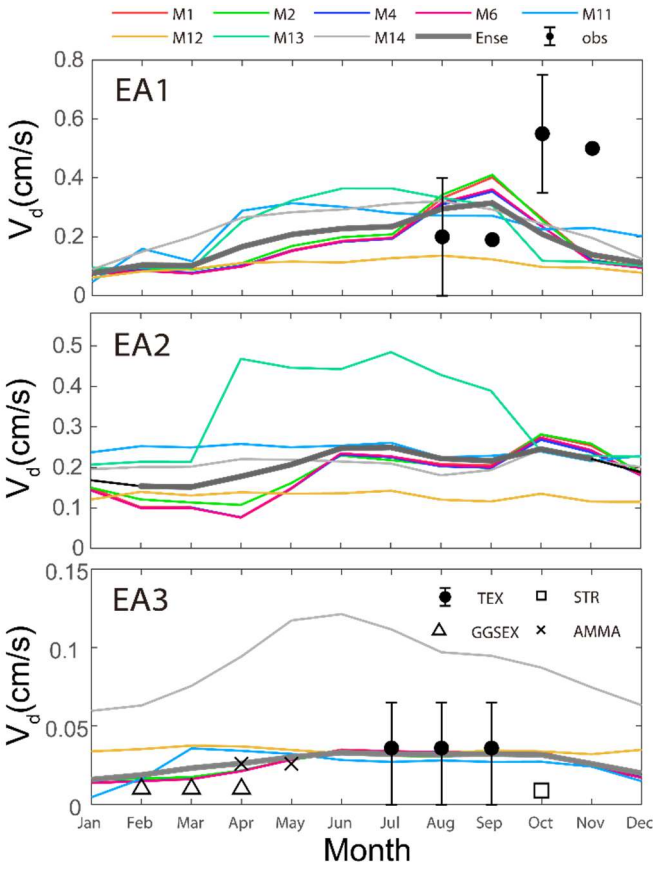

Fig.9 Li et al., 2018

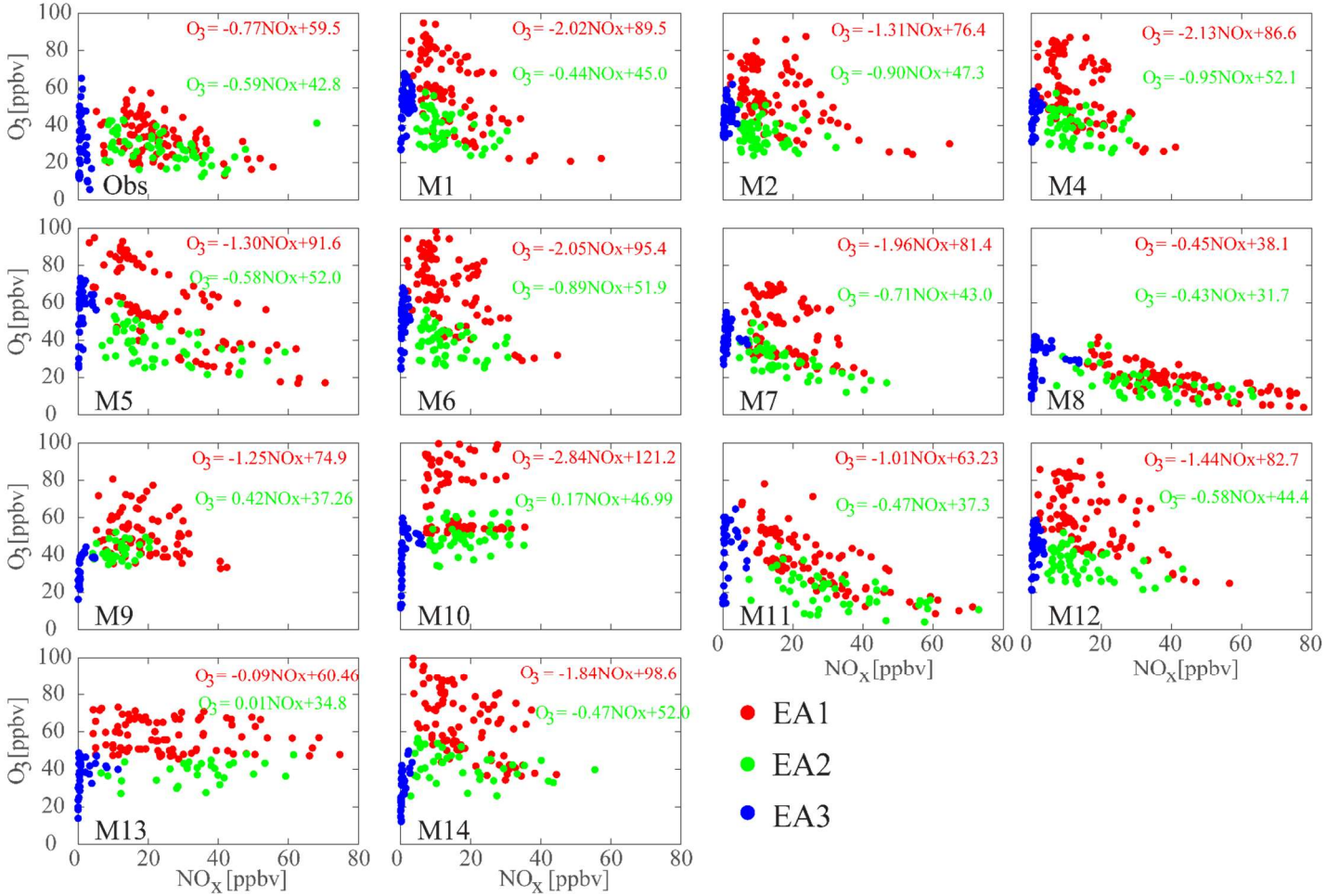

Fig.10 Li et al., 2018