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

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

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

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

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

In section 3.1,

"The O_3 NMB and RMSE of ensemble mean were significantly less than the ensemble median in most situations (Table 1). Therefore, we only presented the results of multi-model mean ensemble (Ense). In general, the majority of models significantly overestimated annual surface O_3 compared with the observations in EA1, EA3 and EA4 (Fig. 2). Ense overestimated surface O_3 by 10-15 ppbv in these subregions. Ense NO_2 was generally close to the observations to within ±20% in all subregions. In EA1 and EA3, Ense NO was 5-10 ppbv lower than observation, and showed a reasonable

performance in EA4."

In section 3.2,

"From the perspective of monthly variation, the overestimation of O_3 mostly appeared in May-September in EA1. Ense O_3 was 10-30 ppbv higher than observations, 30-70% of observed values. In the same period (May-September), Ense NO and NO_2 appeared to be consistent with observations, attaining mean biases of < 3 ppbv. This suggests that the intercomparison on O_3 production efficiency per NO_x with observations is needed. In EA3, Ense O_3 agreed well with observed high autumn O_3, but overestimated from January to September by 5-15 ppbv (15-60% of observations). This maximum of overestimation appeared in March-April (15ppbv), which led to a spring peak in simulated O_3 which was not found in observations. This overestimation was partly related to the underestimation of NO in the same months, which decreased the titration effect. For NO_2, Ense agreed well with observed values in June-December, and slightly underestimated observations in January-May. In EA4, a significant overestimation of O_3 and underestimation of NO existed in June-October. Both observations and Ense NO were lower than 0.5 ppbv, so impact of by NO underestimation on O_3 are needed to be further explored. The ensemble NO_2 was generally close to the observations to within ±0.5 ppbv."

In section 3.3,

"In general, model results for three sub-regions exhibited a larger spread with a magnitude of 10-50 ppbv throughout the diurnal cycle than that in Europe and North America (Solazzo et al., 2012). The Ense O_3 in summer exhibited a systematic overestimation (20 ppbv) throughout the diurnal cycle in EA1. This indicated that models had difficulty dealing with O_3 in North China Plain. Compared with summer, there was only a slight systematic overestimation of Ense O_3 in other seasons (3-5 ppbv). In EA3, Ense O_3 generally agreed with the observations in summer, autumn and winter. In particular, the O_3 maximum around noon was reproduced, reasonably. There was only a 3-5 ppbv overestimation during 16:00-23:00 and early morning (6:00-10:00). In spring, a systematic overestimation of Ense O_3 exited in the whole diurnal cycle (5-10 ppbv). In EA4, Ense captured the small diurnal variation of O_3 in four seasons, but significantly overestimated observations in summer and autumn (5-20 ppbv). In spring and winter, differences between Ense and observations were within 5 ppbv."

In section 3.4,

"In general, Ense performed a better performance level than individual models for representing NO_2 in East Asia, reproducing the observed seasonal cycle and magnitudes. However, Ense did not always exhibited a superior performance for O_3 over certain individual model in East Asia, which was in contrast to its performance in Europe . M11 and M7 agreed well with observations in EA1 and EA3, while ENSE tended to overestimate O_3 concentrations in May-September in EA1 and January-September in EA3. Loon et al. (2007) indicated that ENSE exhibited a superior performance level

only when the spread of ensemble-model values was representative of the uncertainty of O_3. This indicated that most models did not reflect this uncertainty or missed key processes in MICS-Asia III."
In section 3.5,

"In general, ensemble means (Ense) presented an underestimation and overestimation for EA3 O_3 in middle (500-800 hpa) and lower (below 900 hpa) troposphere, respectively. In winter, the underestimation even extended to 200hpa in winter. The magnitudes of underestimation and overestimation reached 10-40 ppbv and 10-20 ppbv. In EA4, Ense reproduced the vertical structure of ozone in both summer and winter. An overestimation existed below 800 hpa, with a magnitude of 10-20 ppbv."

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

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

"The East Asia monsoon played an important role in seasonal cycle of 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 EA3, a late maximum of O_3 in September-November was quite different from EA1 and EA4. 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 EA4, the seasonal change of O_3 concentrations was characterized by two peaks in spring and autumn. The first and second peak in March–April and May-June were mainly influenced by the inflow from outside of East Asia and chemically produced 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."

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

Reply: We agree. In this study, stations are taken from from 1) Chinese Ecosystem Research Network (EA1); 2) Pearl River Delta Regional Air Quality Monitoring Network (PRD RAQMN) (EA2); 3) the Acid Deposition Monitoring Network in East Asia (EANET) (EA3). Observations were rarely affected by the very local emissions around sites, and were used to represent the regional air quality.

- As listed in Table R1 in this reply, most stations are located in rural, remote and

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

Table R1site descriptions in Chinese Ecosystem Research Network

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

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

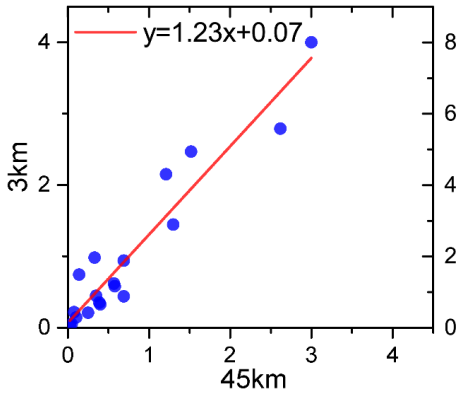

Fig.R1 Scatter plots of NO emission rates (μg/m^2/s) at observation sites in EA1 in 45km and 3km resolution emission inventory

- Pearl River Delta Regional Air Quality Monitoring Network (PRD RAQMN) was

jointly established by the Guangdong Provincial Environmental Monitoring Centre (GDEMC) and the Environmental Protection Department of the Hong Kong Special Administrative Region (HKEPD) from 2003 to 2005. The PRD RAQMN was to probe the regional air quality, assess the effectiveness of emission reduction measures and enhance the roles of monitoring networks in characterizing regional air quality and supporting air quality management (Zhong et al.,2013). So sites are rarely affected by the local emissions near them. Fig. R2 showed the Spatial distribution of average concentrations of NO_2 and O_3 in the PRD-RAQMN Network. Obviously, concentrations of pollutants are smooth. The effect of very local emissions was not seen.

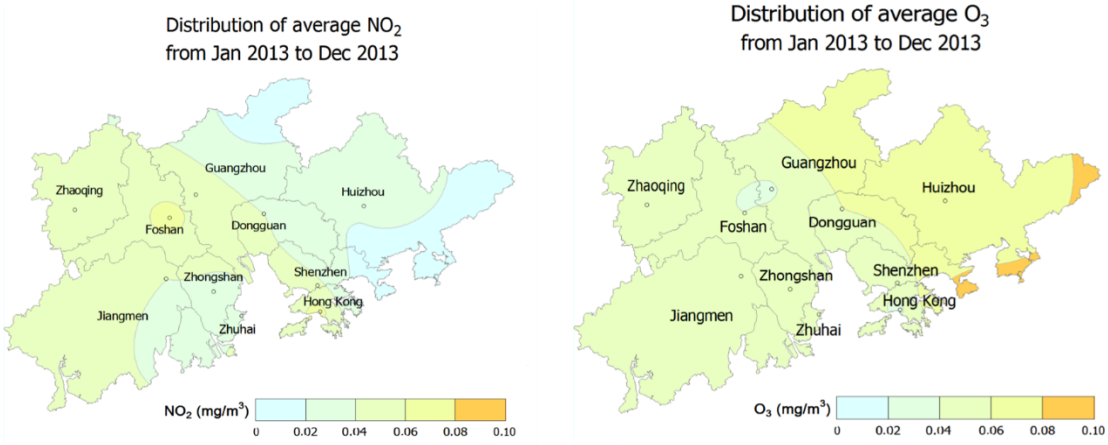

Fig.R2 Spatial distribution of average concentrations of NO_2 and O_3 in the PRD-RAQMN Network, figure is annual report of Pearl River Delta Regional Air Quality Monitoring Network in 2013 (https://www.epd.gov.hk/epd/sites/default/files//epd/english/resources_pub/publications/files/PRD_2013_report_en.pdf)

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

As for NO_2 measurements, we agree that molybden convertors devices may cause errors. Ge et al. (2013) compared the measurements at an urban site in Beijing in summer by commercially standard chemiluminescence-based (called CL hereafter) instruments and Aerodyne Cavity Attenuated Phase Shift Spectroscopy (CAPS). The CAPS NO2 monitor directly measures the absorption of NO_2 at the wavelength of 450 nm and requires no conversion of NO_2 to other species.

Fig. R3-R4 presents the comparison between instruments. Generally, the biggest discrepancy appeared in 12:00-16:00, with a magnitude of 10-20%. In other periods, NO2 by CL and CAPS were similar. On average, discrepancies between CL and CAPS were less than 10%. The linear fitting slope reached 0.999 between CL and CAPS.

As shown in Fig. R4, observations between CL and CAPS agreed well with each other with hourly NO_2>15 ppbv. In low hourly NO_2(<10 ppbv), CL NO_2 overestimated CAPS by 10-30%. This is consistent with the statement by the reviewers, which

reported NO2 exist sometimes biases for stations far from sources in the measurements.

In this study, we compared observed monthly mean NO_2 with models, instead of daytime NO_2. This partly decreased the impact of errors from CL instrument. What's more, the observed NO2 in EA1 and EA3 were 20 ppbv or more. In these high NOx emission regions, biases from CL instruments may not bring too much impact on model validation. In EA4, most stations are located in islands or remote regions, with ~ 2 ppbv NO_2. The CL NO_2 will overestimated NO_2 concentrations.

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

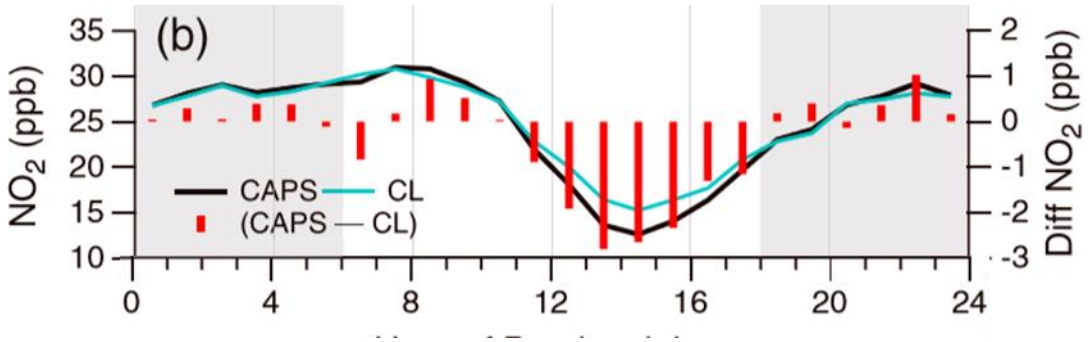

Fig. R3Observed mean diurnal variation of NO_2 in summer in Beijing by chemiluminescence-based (CL) instruments and CAPS in Beijing. Also shown is the difference of two instruments.

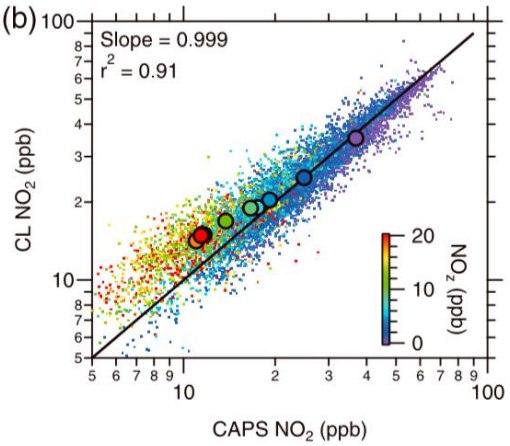

Fig. R4 Comparison of NO_2 measured by the CL NOx analyzer and CAPS.

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

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

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

previous manuscript.

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

Reply: We totally agree. In the revised manuscript, we collected observation data as much as possible. The new observation data includes:1) vertical profiles of O_3 in EA3 and EA4; 2) PBLH in EA1 and EA3; 3) dry deposition velocities in EA1 and EA4. We also discussed the model performance against these observations.

Fig. R5 presents the simulated and observed O_3 profiles in subregions. Because there was lack of O_3 sounding in EA1 in 2010, only observations in EA3 and EA4 are show. In general, ensemble means (Ense) presented an underestimation and overestimation for EA3 O_3 in middle (500-800 hpa) and lower (below 900 hpa) troposphere, respectively. In winter, the underestimation even extended to 200hpa in winter. The magnitudes of underestimation and overestimation reached 10-40 ppbv and 10-20 ppbv. In EA4, Ense reproduced the vertical structure of ozone in both summer and winter. An overestimation existed below 800 hpa in summer, with a magnitude of 10-20 ppbv.

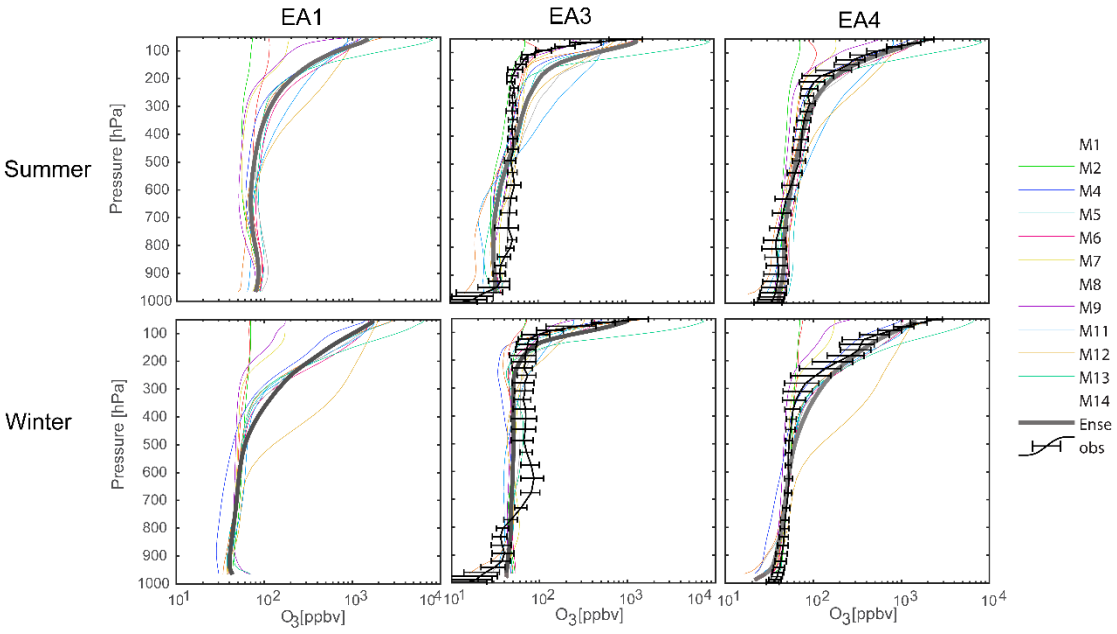

Fig. R5 Simulated and observed O_3 profiles in summer and winter of 2010, averaged over all observed stations in three subregions over East Asia (EA1: left column, EA3: middle column, EA4: bottom column).

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

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

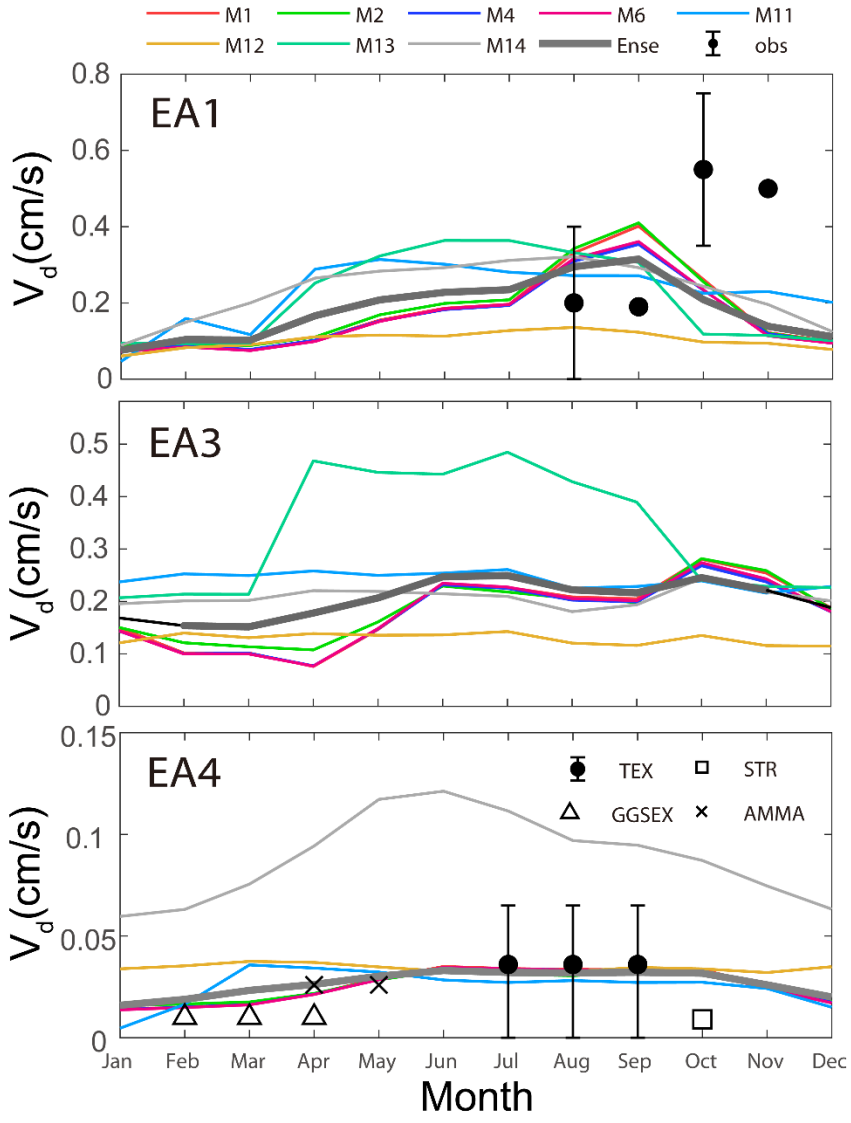

Fig. R6 simulated and observed monthly O_3 dry deposition velocities. Observations in

EA1 were from Sorimachi et al. (2003) and Pan et al. (2010). Observations in EA4 were from Luhar et al. (2017).

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

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

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

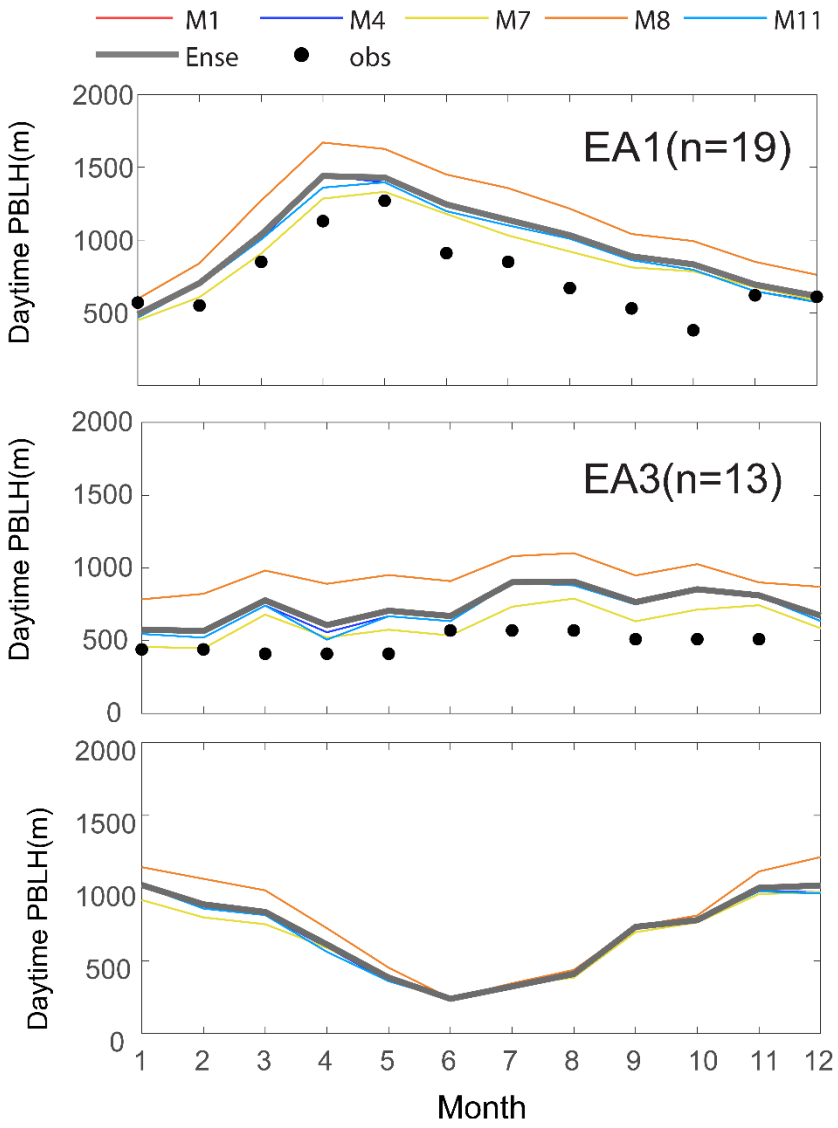

Fig. R7 Simulated daytime (08:00-20:00 LST) PBL height (m). Also shown are observed mean PBL height (m) at 08:00, 14:00 and 20:00 LST from Guo et al. (2016).

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

Other comments etc...

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

Reply: We added it (100 μg/m^3).

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

Reply: We deleted it.

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

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

“In general, model results for three sub-regions exhibited a larger spread with a magnitude of 10-50 ppbv throughout the diurnal cycle than that in Europe and North America (Solazzo et al., 2012). The Ense O_3 in summer exhibited a systematic overestimation (20 ppbv) throughout the diurnal cycle in EA1. This indicated that models had difficulty dealing with O_3 in North China Plain. Compared with summer, there was only a slight systematic overestimation of Ense O_3 in other seasons (3-5 ppbv)”

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

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

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

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

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

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

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

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

Reply: Yes, it is “von Engeln”.

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

Reply: We plotted VOCs (ethene) emissions (Fig. R8). Compared with NO, the

consistency on ethene is better. Only M2 showed a small underestimation and overestimation in EA1 and EA3, respectively.

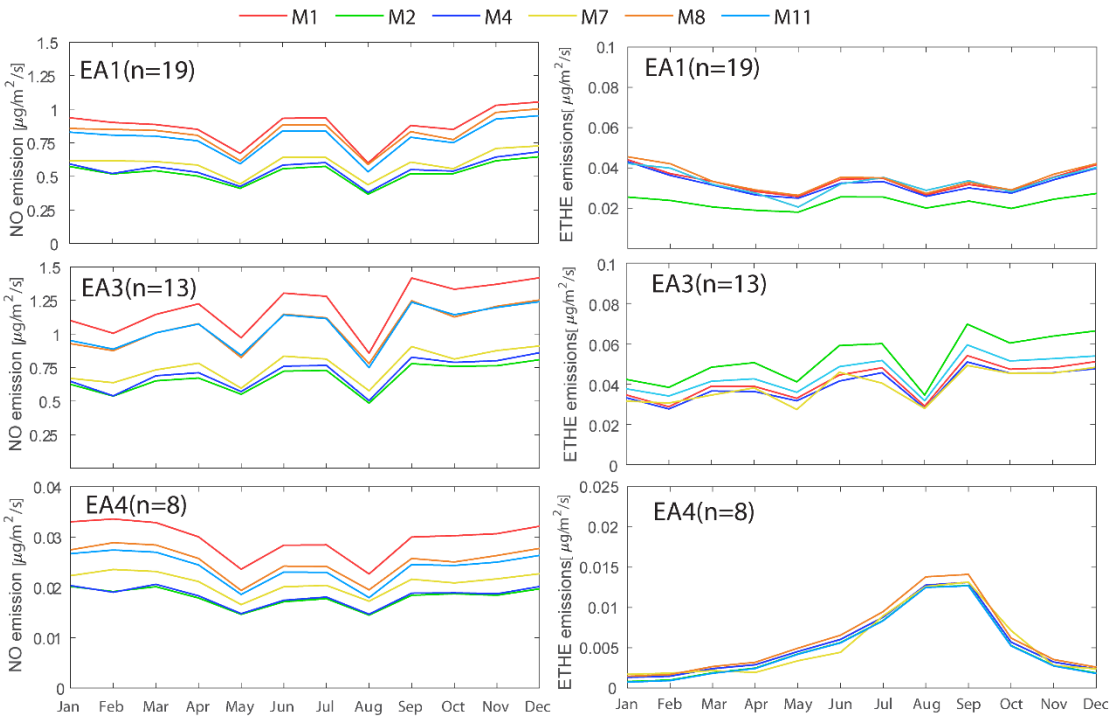

Fig.R8 NO (left) and ethene (right) emission fluxes on the first day in each month.

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

Reply: Thanks a lot.

"The difference in emissions allocations could contribute to the simulation variability. In the future, the projected gridded anthropogenic emissions should be provided to each group to eliminate the possibility that each group uses different mapping method. Interestingly, emissions in M1 and M8 exhibited similar levels, but their simulated NO_2, NO and O_3 presented a high intermodel variability in EA1 (Fig. 3 and Fig. 6). M1 simulated summer O_3 reached 80 ppbv while M8 was only 30 ppbv. This indicated that there were others causes to bring the intermodel variability on O_3."

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

Reply: We agree.

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

Reply: We agree. In the revised manuscript, we reworded this sentence. "Compared to other regions, surface O_3 in EA4 were more sensitive to dry deposition parameterization

schemes in CTMs (Park et al.,2014). Park et al. (2014) revealed that O_3 on oceans differed by 5-15 ppbv in East Asia resulting from different dry deposition parameterization schemes". We deleted "Ganzeveld et al. (2009) revealed that surface O_3 may differ by up to 60% when O_3 dry deposition velocity varied from 0.01 to 0.05 cm/s."

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

Reply: This assumption was taken from Park et al. (2014), in which the impact of O_3 dry deposition was examine over East Asia. They found that O_3 mixing ratios in EA4 were more sensitive to dry deposition parameterization schemes in CTMs than other regions. O_3 decrease as low as 5-15 ppbv at stations in EA4 in Wesely scheme than M3DRY scheme (1990). In EA1 and EA3, the changes of O_3 only ranged from 0-5 ppbv.

Comment 18: Page 17 – Line 1 – I observe that range of concentrations for O3 and NOx can be very different between models but it is not clear if slopes are that different.

Reply: We plotted the slopes between NO_x and O_3 in Fig. 8 in the revised manuscript. The slopes between NO_x and O_3 in EA1 ranged from -2.84 to -0.09 between models.

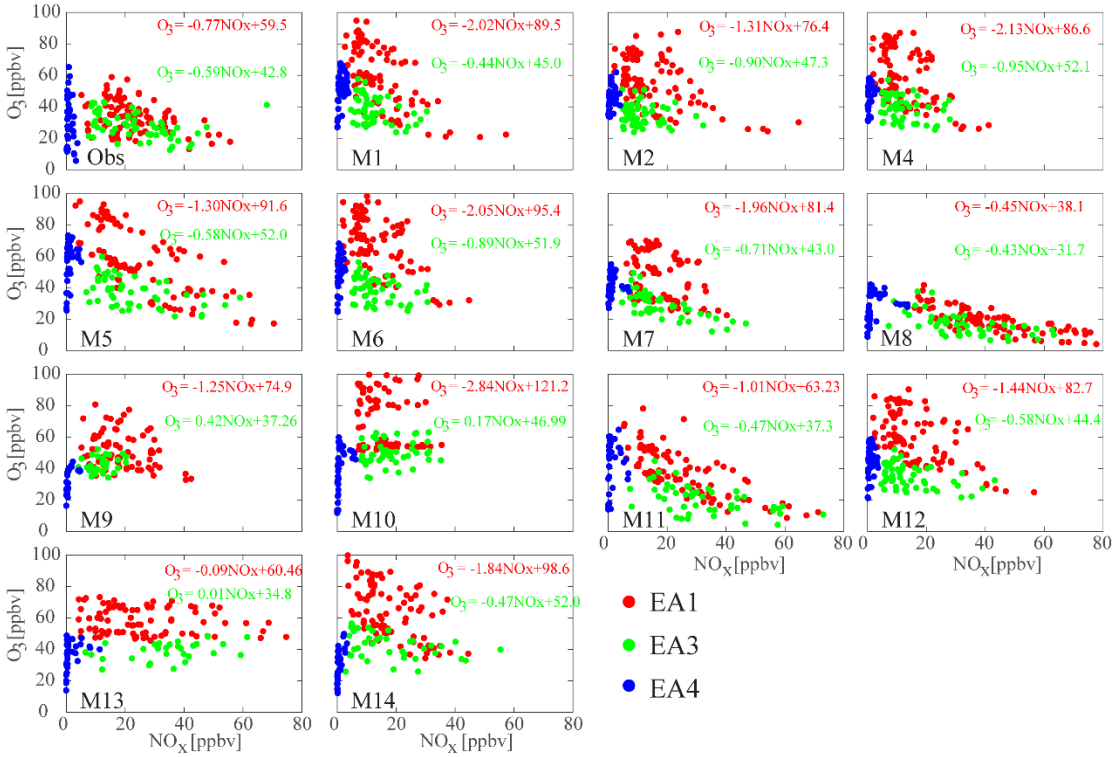

Fig. R9 Scatter plots between monthly daytime (08:00-20:00) surface NO_x and O_3 at each station over EA1(red), EA3(green)and EA4(blue) in May-October, for observations(obs) and models

Comment 19: Page 18 – Line 2 to 5 but also Line 7 to 20 – The variability authors are

mentioning is not clear from figure 9. Also for differences between winter and summer, we need to have numbers to better evaluate this variability.

Reply: Thanks. Line 2-5: "A small variability in winter appeared below 900 hPa in three sub-regions, and slowly decreased with height. The mean standard deviation (σ) below 900 hpa were 7.6 ppbv, 6.9 ppbv and 6.0 ppbv in EA1, EA3 and EA4, which covered 18.3%, 15.0% and 15.4% of mean O_3 concentrations. In 700-900 hpa, σ decreased to 5.4 ppbv, 4.4 ppbv and 4.8 ppbv in EA1, EA3 and EA4, 12.2%, 9.4% and 10.8% of mean O_3 concentrations".

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

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

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

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

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

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

Reply: Thanks a lot. We totally agree that an ensemble averages representing the uncertainty of O_3 is helpful. In MICS-ASIA III, the arithmetic means of all models is difficult meet this criteria, although it has been successfully in other regions. Potempski and Galmarini (2009) did some basic theoretical to find optimal linear combination of

model results with the help of complex mathematical tools. Solazzo et al. (2012) used this method for O_3 ensemble in Europe and North America. They found that the most skillful ensemble is not necessarily generated by including all available models, and suggested that the clustering technique could generate a better ensemble average, but needs further refinement. This is beyond the scope of this manuscript and will be the major topic of our next manuscript

We mentioned that most models did not reflect this uncertainty or missed key processes in MICS-Asia III. The parameterization of heterogeneous chemistry in models is possibly a key process. The manuscript by Akimoto et al. (2019) in this special issue found that the missing heterogeneous "renoxification" reaction of HNO_3 on soot in most models except NAQPMS would partly explain the overestimation of simulated O_3 mixing ratios. The treatment of O_3 vertical transport in models also affect the simulated results significantly in Akimoto et al. (2019).

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

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

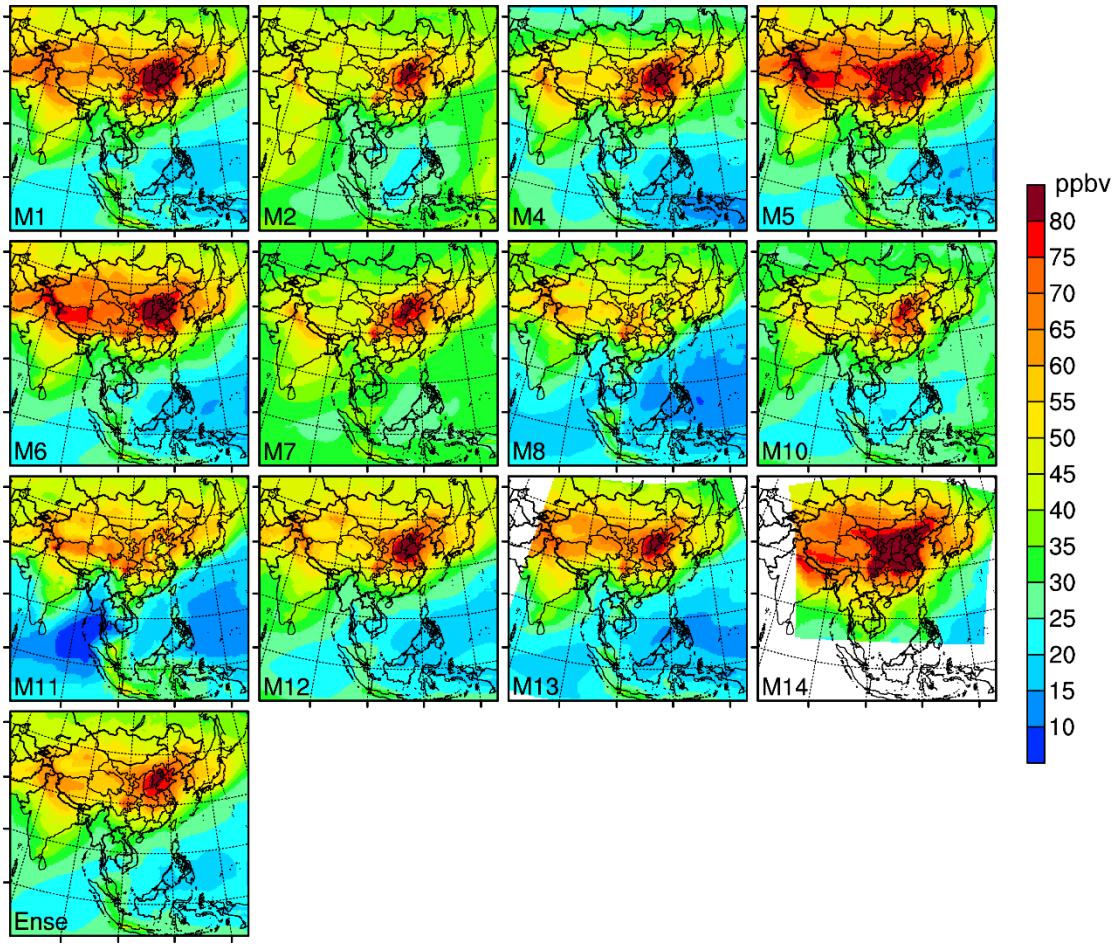

Fig. R10 500m O_3 spatial distribution from 13 models for summer 2010

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

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

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

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

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

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

Comment 25： Page 22 –Line 7 – " .. its relevant species .." I also see VOC or even radicals as relevant species for the tropospheric ozone cycle then it is better to mention 03 and NOx instead.

Reply: We agree and revised it.

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

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

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

Reply: We revised it.

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

Reply: We revised it.

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

Reply: We revised it.

References:

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

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

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

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

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

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

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

Liu, S. C., McKeen, S. A., Hsie, E-Y., Lin, X., Kelly, K. K., Bradshaw, J. D., Sandholm, S. T., Browell, E. V., Gregory, G. L., Sachse, G. W., Bandy, A. R., Thornton, D. C., Blake, D. R., Rowland, F. S., Newell, R., Heikes, B. G., Singh, H., and Talbot, R. W. : Model study of tropospheric trace species distributions during PEM-West A*, J. Geophys. Res.,* 101, 2073-2085,1996.

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

and Air Quality Research, 13: 1582–1597, 2013

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

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

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

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

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

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