Air Quality and Climate Change, Topic 3 of the Model Inter-Comparison Study for Asia Phase III (MICS-Asia III), Part I: overview and model evaluation

Meng Gao[1,2], Zhiwei Han[3,4], Zirui Liu[5], Meng Li[6, 13], Jinyuan Xin[5], Zhining Tao[7,8], Jiawei Li[4], Jeong-Eon Kang[9], Kan Huang[10], Xinyi Dong[10], Bingliang Zhuang[11], Shu Li[11], Baozhu Ge[5], Qizhong Wu[12], Yafang Cheng[13], Yuesi Wang[5], Hyo-Jung Lee[9], Cheol-Hee Kim[9], Joshua S. Fu[10], Tijian Wang[11], Mian Chin[8], Jung-Hun Woo[14], Qiang Zhang[6], Zifa Wang[4,5], Gregory R. Carmichael[1]

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

2 John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA

3 Key Laboratory of Regional Climate-Environment for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

4 University of Chinese Academy of Sciences, Beijing 100049, China

5 State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

6 Ministry of Education Key Laboratory for Earth System Modeling, Center for Earth System Science, Tsinghua University, Beijing, China

7 Universities Space Research Association, Columbia, MD, USA

8 NASA Goddard Space Flight Center, Greenbelt, MD, USA

9 Department of Atmospheric Sciences, Pusan National University, Busan, South Korea

10 Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN, USA

11 School of Atmospheric Sciences, Nanjing University, Nanjing, China

12 College of Global Change and Earth System Science, Beijing Normal University, Beijing, China

13 Multiphase Chemistry Department, Max Planck Institute for Chemistry, Mainz, Germany

14 Department of Advanced Technology Fusion, Konkuk University, Seoul, South Korea

Correspondence to: M. Gao (mgao2@seas.harvard.edu), Z. Han (hzw@mail.iap.ac.cn), and G. R. Carmichael (gcarmich@engineering.uiowa.edu)

Abstract

Topic 3 of the Model Inter-Comparison Study for Asia (MICS-Asia) Phase III examines how online coupled air quality models perform in simulating high aerosol pollution in the North China Plain region during wintertime haze events and evaluates the importance of aerosol radiative and microphysical feedbacks. A comprehensive overview of the MICS-ASIA III Topic 3 study design, including descriptions of participating models and model inputs, the experimental designs, and results of model evaluation, are presented. Six modeling groups from China, Korea and the United States submitted results from seven applications of online coupled chemistry-meteorology models. Results are compared to meteorology and air quality measurements, including the Campaign on Atmospheric Aerosol Research Network of China (CARE-China) network, and the Acid Deposition Monitoring Network in East Asia (EANET). The correlation coefficients between multi-model ensemble mean and the CARE-China observed near-surface air pollutants range from 0.51 to 0.94 (0.51 for ozone and 0.94 for $PM_{2.5}$) for January 2010. However, large discrepancies exist between simulated aerosol chemical compositions from different models, which is due to different parameterizations of chemical reactions. The

coefficient of variation (standard deviation divided by the mean) can reach above 1.3 for sulfate in Beijing, and above 1.6 for nitrate and organic aerosol in coastal regions, indicating these compositions are less consistent from different models. During clean periods, simulated Aerosol Optical Depths (AOD) from different models are similar, but peak values differ during severe haze events, which can be explained by the differences in simulated inorganic aerosol concentrations and the hygroscopic growth efficiency (affected by varied relative humidity). These results present how current online-coupled meteorology-chemistry models reproduce severe haze events, and provide some directions for future model improvements, such as new heterogeneous or aqueous pathways for sulfate and nitrate formation under hazy conditions, secondary organic aerosol (SOA) formation chemical mechanism with new volatile organic compounds (VOCs) precursors, yield data and approaches, and the dependence of aerosol optical properties on size distribution and mixing state.

1 Introduction

Air pollution in Asia, particularly in China and India, has been an increasing important research topic, and has attracted enormous media coverage since about 60% of the world population live and are exposed to extremely unhealthy air in this region. It is estimated that outdoor air pollution brings about 3.3 million premature deaths per year worldwide with most deaths occur primarily in Asia (Lelieveld et al., 2015). In addition, the impacts of regional and intercontinental transport of Asian pollutants on air quality and climate change have been frequently reported (Akimoto, 2003; Menon et al., 2002, Ramanathan and Carmichael, 2008). Chemical transport models have been developed and applied to study various air pollution issues in Asia. For

example, an Eulerian regional scale acid deposition and photochemical oxidant model was developed in the United States (Carmichael and Peters, 1984; Carmichael et al., 1986; Carmichael et al., 1991) and applied to study long-range transport of sulfur oxides (SO_x), dust and ozone production in East Asia (Carmichael et al., 1998; Xiao et al., 1997). A nested urban and regional scale air quality prediction modeling system was developed and applied to investigate ozone pollution in Taiwan (Wang et al., 2001). Although important advances have taken place in air quality modeling, large uncertainties still remain, which are related to inaccurate and/or incomplete emission inventories, poorly represented initial and boundary conditions and missing or poorly parameterized physical and chemical processes (Carmichael et al., 2008a).

Furthermore, many models used to study air quality in Asia have been developed in other regions (e.g., USA and Europe), and the assumptions and parameterizations included in these models may not be applicable to the Asian environment. In order to develop a common understanding of model performance and uncertainties in Asia, and to further develop the models for Asian applications, a model inter-comparison study was initiated, i.e., Model Inter-Comparison Study for Asia Phase I (MICS-Asia I), in 1998 during a workshop on Transport of Air Pollutants in Asia in Austria. The focus of MICS-Asia Phase I was to study long-range transport and deposition of sulfur within Asia in support of on-going acid deposition studies. Eight long-range transport models from six institutes in Korea, Japan, Denmark, the USA, and Sweden participated in MICS-Asia I. Multi-model results of sulfur dioxide (SO_2) and sulfate concentrations, and wet deposition amounts in January and May 1993 were compared with surface observations in East Asia (Carmichael et al., 2002). Source-receptor relationships and how model structure and parameters affect model performance were also discussed during this

89 phase (Carmichael et al., 2002). In 2003, MICS-Asia Phase II was initiated to include more
90 species, including nitrogen compounds, ozone and aerosols. The study period was expanded to
91 cover two different years and three different seasons, and global inflow to the study domain was
92 also considered (Carmichael et al., 2008b). Nine modeling groups from Korea, Hong Kong,
93 Japan, the USA, Sweden, and France participated in this phase. Seven topics (i.e., ozone and
94 related precursors, aerosols, acid deposition, global inflow of pollutants and precursors to Asia,
95 model sensitivities to aerosol parameterization, analysis of emission fields, and detailed analyses
96 of individual models) were discussed and published in a special issue of Atmospheric
97 Environment (Carmichael et al., 2008b).

98 In 2010, MICS-Asia phase III was launched and three topics for this phase were decided during
99 the first and second Workshop on Atmospheric Modeling in East Asia. Phase III aims to evaluate
100 strengths and weaknesses of current air quality models and provide techniques to reduce
101 uncertainty in Asia (Topic 1), to develop a reliable anthropogenic emission inventory in Asia
102 (Topic 2), and to evaluate aerosol-weather-climate interactions (Topic 3). Various multi-scale
103 models participated in this phase and the study periods range from year to month depending on
104 study topics. This phase uses data from the Acid Deposition Monitoring Network in East Asia
105 (EANET), in addition to new observations related to atmospheric chemistry in the region. A
106 detailed overview of MICS-Asia Phase III, including descriptions of different research topics and
107 participating models, will be published in a companion paper. An important advance to this
108 phase is the inclusion of multiple online-coupled chemistry-meteorology models to investigate
109 aerosol-weather-climate interactions, which is the target of topic 3. On-line coupled models are
110 playing important roles in air quality, meteorology and climate applications, but many important
111 research questions remain (Baklanov et al., 2017).

The influences of aerosols on meteorology, e.g., radiation, temperature, boundary layer heights, winds, etc. and $PM_{2.5}$ concentrations have been examined in previous studies using different online coupled models (Forkel et al., 2015; Gao et al., 2016a, 2016b, 2017a, 2017b; Han et al., 2012, 2013; Makar et al., 2015a, 2015b; San Jose et al., 2015; Tao et al., 2015, 2016; Wang et al., 2014; Zhang et al., 2010). In general, there are two ways of online coupling: online integrated coupling (meteorology and chemistry are simulated using the same model grid, and one main time step is used to integrate) and online access coupling (meteorology and chemistry are independent but data are exchanged on a regular basis) (Baklanov et al., 2014). These two different coupling ways can lead to uncertainties in the results of aerosol-weather-climate interactions. Even using the same coupling way, different parameterizations in different online models causes uncertainties as well. Thus, it is important to inter-compare how different online models simulate aerosol-weather-climate interactions, particularly in heavily polluted Asian region. Other ongoing related modeling frameworks include the Task Force on Hemispheric Transport of Air Pollution (TF HTAP) and the Air Quality Model Evaluation International Initiative (AQMEII). The TF HTAP was initiated to improve knowledge of the intercontinental or hemispheric transport and formation of air pollution, and its impacts on climate, ecosystems and human health (Galmarini et al., 2017; Huang et al., 2017). The AQMEII project specifically focuses on regional modeling domains over Europe and North America (Galmarini et al., 2017), within which aerosol meteorology interactions was studied (Forkel et al., 2015; Makar et al., 2015a, 2015b; San Jose et al., 2015) over Europe and North America.

This paper presents and overview of the MICS-ASIA III Topic 3, serving as the main repository of the information linked to Topic 3 simulations and comparisons. Specifically, this paper aims to archive the information of participating models, how the experiments, and results of model

evaluation. The results of the MICS-Asia Topic 3 experiments looking at the direct and indirect effects during heavy haze events will be published in a companion paper, part II. This paper is organized as follows: in Section 2, we provide the inter-comparison framework of Topic 3, including the participating models, emissions, boundary conditions, observational data, and analysis methodology. Section 3 presents the general descriptions of the study periods and Section 4 presents comparisons and discussions focused on the results related to the meteorological and air pollution conditions during the January 2010 heavy haze episode. The results of January 2013 haze episode and detailed analysis of the direct and indirect effects will be presented in a companion paper.

2 Inter-comparison framework

In North China, severe aerosol pollution frequently happens and attracts enormous interests from both public and scientific communities (Cheng et al., 2016; Gao et al., 2015, 2016a, 2016b, 2016c). Two winter months in which severe haze episodes happened in North China were selected as the study periods for Topic 3. During these two months, maximum hourly $PM_{2.5}$ concentration in urban Beijing reached ~500 $\mu g/m^3$ and 1000$\mu g/m^3$, respectively. Compared to the China Grade 1 24-h $PM_{2.5}$ standard (35$\mu g/m^3$), daily mean $PM_{2.5}$ concentrations in urban Beijing exceeded this standard for 20 days and 27 days within these two months, respectively. The dramatically high aerosol loadings during these two hazy months substantially affected radiation transfer, and provide a good opportunity to study the aerosol effects on weather, air quality and climate. In this study, the participants were required to use common emissions to predict air quality during these two months and submit requested model variables. The emissions

were placed on a publicly accessible website. Six modeling groups submitted results for Topic 3. In this section, we briefly describe these models and their configurations, introduce the emission inventories (including anthropogenic, biogenic, biomass burning, air and ship, and volcano emissions), observational datasets, and describe the analysis methodology.

2.1 Participating models

Table 1 summarizes the characteristics of the participating models. These models include: one application of the Weather Research Forecasting model coupled with Chemistry (WRF-Chem, Fast et al., 2006; Grell et al., 2005) by Pusan National University (PNU) (M1), one application of the WRF-Chem model by the University of Iowa (UIOWA) (M2), two applications (two domains: 45km and 15km horizontal resolutions) of the National Aeronautics and Space Administration (NASA) Unified WRF (NU-WRF, Peters-Lidard et al., 2015; Tao et al., 2013) model by the Universities Space Research Association (USRA) and NASA's Goddard Space Flight Center (M3 and M4), one application of the Regional Integrated Environment Modeling System with Chemistry (RIEMS-Chem, Han et al., 2010) by the Institute of Atmospheric Physics (IAP), Chinese Academy of Sciences (M5), one application of the coupled Regional Climate Chemistry Modeling System (RegCCMS, Wang et al., 2010) from Nanjing University (M6), and one application of the coupled WRF-CMAQ (Community Multiscale Air Quality) model by the University of Tennessee at Knoxville (UTK) (M7). These models are all online coupled, which enables aerosol-weather-climate interactions. Domain setting of each model application is shown in Figure 1. The domains of M2, M5, and M6 (UIOWA, IAP, and NJU in Figure 1) cover most areas of East Asia, including China, North Korea, South Korea, Japan, Mongolia, and north parts of Southeast Asia. M1, M3 and M7 domains (PNU, NASA D01 and UTK) include more countries in Southeast and South Asia. M4 (NASA D02) covers east China, Korea and Japan.

The descriptions of major model settings are listed below. More descriptions including microphysics, radiation, and boundary layer, are listed in Table 1.

(1) Model grids: The horizontal model resolutions of these applications range from 15km to 60km (Table 1). Model vertical resolutions vary from 16 to 60 layers (Table 1) and the set model top pressures range from 100mb to 20mb (Table 1).

(2) Gas phase chemistry: At PNU (M1), the RACM-ESRL (Regional Atmospheric Chemistry Mechanism, Earth System Research Laboratory) gas phase chemistry was used. RACM was developed based on Regional Acid Deposition Model (RADM2) to simulate regional atmospheric chemistry (Stockwell et al., 1997) (including 237 reactions) and the rate coefficients were updated in RACM ESRL version (Kim et al., 2009). At the University of Iowa (M2), CBMZ (Carbon-Bond Mechanism version Z) gas phase chemistry was used. CBMZ (Zaveri and Peters, 1999) extends the original CBM4 mechanism to function properly at larger spatial and longer timescales. The augmented CBMZ scheme includes 67 species and 164 reactions. The NU-WRF model (M3 and M4) uses RADM2 for gas phase chemistry. Both the RIEMS-Chem model from IAP (M5) and the RegCCMS model from NJU (M6) used CBM4 to calculate gas phase chemistry (Gery et al., 1989). The CBM4 version incorporated in RIEMS-Chem (M5) includes 37 species and 91 reactions. The version of CBM4 implemented in RegCCMS (M6) consists of 36 reactions (4 photolysis reactions) and 20 species (Wang et al., 2010). M7 applied SAPRC 99 to simulate gas phase chemistry. The SAPRC99 mechanism implanted within the CMAQ model has 88 species and 213 chemical reactions (Carter, 2000a, b).

(3) Aerosol modules: MADE/SORGAM (Modal Aerosol Dynamics Model for Europe/Secondary Organic Aerosol Model) aerosol module was coupled and used in M1. MADE

uses 3 log-normal modes (Aitken, accumulation, coarse) and simulates major aerosol compositions, including sulfate, ammonium, nitrate, sea-salt, black carbon (BC), and organic carbon (OC). M2 uses 8 bin MOSAIC (Model for Simulating Aerosol Interactions and Chemistry) aerosol module. MOSAIC considers major aerosol species at urban, regional and global scales, including sulfate, nitrate, ammonium, sodium, chloride, EC, and other unspecified inorganic species (such as inert minerals, trace metals, and silica) (Zaveri et al. 2008). The MOSAIC version used in M2 includes some aqueous reactions but no SOA formation. At NASA, the GOCART aerosol model (Chin et al., 2002) was coupled to RADM2 gas phase chemistry, and incorporated into the NU-WRF model (M3 and M4) to simulate major tropospheric aerosol species, including sulfate, BC, OC, dust, and sea-salt. In this aerosol model, 10% of organic compounds from the volatile organic compounds (VOCs) emission inventory are assumed to be converted to SOA (Chin et al., 2002). Aerosols in RIEMS-Chem include sulfate, nitrate, ammonium, BC, OC, SOA, 5 bins of soil dust, and 5 bins of sea salt (Han et al., 2012). ISORROPIA (Nenes et al., 1998) is coupled to RIEMS-Chem to treat thermodynamic equilibrium process and to simulate inorganic aerosols. SOA production from primary anthropogenic and biogenic VOCs is calculated using a bulk aerosol yield method according to Lack et al. (2004). RegCCMS also used ISORROPIA to calculate inorganic aerosols (Wang et al., 2010). For implementation of aerosol effects, sulfate radiative properties were treated following Kiehl and Briegleb (1993), OC were assumed to have the same properties as sulfate, and the wavelength-dependent radiative properties of BC follows Jacobson (2001). AE6 aerosol (the sixth-generation CMAQ aerosol module) mechanism is coupled with WRF. Compared to previous version of CMAQ aerosol modules, AE6 improves SOA treatments, adds a new heterogeneous N_2O_5 hydrolysis parameterization and ads a new gas-to-particle mass transfer for

coarse aerosols in sea-salt emissions (Yu et al., 2014). There are seven components including water soluble mass, water insoluble mass, elemental carbon, sea salt, water, diameters and standard deviations passed to WRF to directly change radiation calculations.

(4) Meteorological boundary and initial conditions: M1, M2, M5 and M7 use the National Centers for Environmental Prediction (NCEP) final analysis (FNL) data to drive the model; M3 and M4 use NASA MERRA reanalysis data and M6 uses NCEP-NCAR reanalysis 1 dataset.

(5) Soil dust: M1, M6 and M7 do not include soil dust calculation. M3 and M4 use GOCART dust module, and M2 uses a GOCART version that modified by AFWA (Air Force Weather Agency). M5 uses a dust module that described in Han et al. (2004).

(6) Mixing state: M6 assumes external mixing, while other models use internal mixing treatments for major aerosol compositions.

Many previous studies have underscored that the choice of gas phase mechanism and aerosol models are of great importance for simulating air pollutants (Knote et al., 2015). The different gas phase chemistry and aerosol modules used in the participating models are expected to yield notable differences in performances, which are shown later in section 4.

2.2 Emissions

The accuracy of air quality modeling results highly depends on the quality and reliability of emission inventory. Accordingly, a new Asian emission inventory was developed for MICS-III by integrating state-of-the-art national/regional inventories to support this model inter-comparison study (Li et al., 2017). This is the major theme of MICS-ASIA III Topic 2. These emissions, along with biogenic emissions, biomass burning emissions, emissions from air and

ship, and volcano emissions were used. This section offers some basic descriptions of these provided emissions.

2.2.1 Anthropogenic emissions

The state-of-the-art anthropogenic emission inventory for Asia (MIX) was developed by incorporating five inventories, including the REAS inventory for Asia developed at the Japan National Institute for Environmental Studies (NIES), the MEIC inventory for China developed at Tsinghua University, the high resolution ammonia (NH_3) emission inventory in China developed at Peking University, the Indian emission inventory developed at Argonne National Laboratory in the United States, and the CAPSS Korean emission inventory developed at Konkuk University (Li et al., 2017). This MIX inventory includes emissions for ten species, namely SO_2, nitrogen oxides (NO_x), carbon monoxide (CO), non-methane volatile organic compounds (NMVOC), NH_3, PM_{10}, $PM_{2.5}$, BC, OC, and carbon dioxide (CO_2). NMVOC are provided with CB-05 and SAPRC-99 speciation datasets. Emissions of these species were prepared for years 2008 and 2010 in monthly temporal resolution and 0.25 degree spatial resolution. Weekly/diurnal profiles were also provided. Five sectors were considered, namely industry, power generation, residential sources, transportation and agriculture. Figure 2 shows the spatial maps of these ten species for January 2010. Emissions of most of these species exhibit similar spatial patterns, with enhanced values in east China and lower values in north and south India. Emissions of NH_3 display a different spatial distribution, with pronounced values in India and lower values in north China (Figure 2). More detailed description of this emission inventory is documented in Li et al. (2017).

2.2.2 Biogenic emissions

Terrestrial ecosystems generate various chemical species, including volatile and semi-volatile compounds, which play important roles in atmospheric chemistry and are the largest contributor to global annual flux of reactive volatile organic compounds (VOCs) (Guenther et al., 2006). For MICS-ASIA III, hourly biogenic emissions were provided for the entire year of 2010 using the Model of Emissions of Gases and Aerosols from Nature (MEGAN) version 2.04 (Guenther et al., 2006). The variables that drive MEGAN include land cover information (plant function type, leaf area index) and weather condition, which includes solar transmission, air temperature, humidity, wind speed, and soil moisture. In the preparation of MEGAN biogenic emissions, land cover information is taken from the NASA MODIS products, and weather condition are calculated using WRF simulations. Figure S1 shows biogenic emissions of some selected species (isoprene and HCHO) for January 2010. High biogenic emissions are found in south Asia during winter, including India, south China, and Southeast Asia, where solar radiation, air temperature and vegetation covers are relatively higher than in northern regions. As shown in Table 1, M1 and M5 use prescribed biogenic VOCs emissions, other models except M6 use internal calculation.

2.2.3 Biomass burning emissions

Biomass burning in the tropics is a strong contributor to air pollutants, and extensive biomass burning in Asia, particularly Southeast Asia, exerts a great influence on air quality (Streets et al., 2003). For MICS-ASIA III, biomass burning emissions were processed by re-gridding the Global Fire Emissions Database version 3 (GFEDv3) (0.5 by 0.5 degree). GFED fire emissions are estimated through combining satellite-detected fire activity and vegetation productivity information. Carbon, dry matter, CO_2, CO, CH_4, hydrogen, nitrous oxide, NO_x, NMHC, OC, BC, $PM_{2.5}$, total particulate matter and SO_2 emissions are estimated in monthly temporal resolution. Figure S2 shows the gridded biomass burning emissions for January 2010. Biomass burning

activity is highest in Cambodia and some areas of Myanmar and north of Thailand (Figure S2), and the peak emission season is spring. Although it has been concluded that biomass burning could significantly contribute to aerosol concentrations in China, the contribution is limited for Topic 3 study since the focused region is North China where biomass burning emissions are negligible during winter (Gao et al., 2016a).

2.2.4 Volcanic SO_2 emissions

Volcanoes are important sources of various sulfur and halogen compounds, which play crucial roles in tropospheric and stratospheric chemistry. It is estimated that SO_2 emitted from volcanoes account for about 9% of the total worldwide annual SO_2 flux (Stoiber et al., 1987). The Asia-Pacific region is one of the most geologically unstable regions in the world where many active volcanoes are located. During MICS-ASIA Phase II, the volcano SO_2 emissions had already been provided for chemical transport models (Carmichael et al. 2008b). Volcano SO_2 emissions were provided, with a daily temporal resolution. In January, some volcanoes in Japan are very active, such as Miyakejima (139.53ºE, 34.08ºN, and 775m above sea level) and Sakurajima (130.65ºE, 31.59ºN, 1117m above sea level).

2.2.5 Air and Ship emissions

Fuel burning in aircraft and ship engines produces greenhouse gases and air pollutants. The shipping and aircraft emissions used are based on HTAPv2 emission inventory (0.1 by 0.1 degree) for year 2010 (Janssens-Maenhout et al., 2015), provided on an annual basis. Aircraft emissions include three parts: landing and takeoff (LTO), climbing and descent (CDS), and cruise (CRS). Aircraft emission hot spots are mostly located in Japan, and Beijing, Yangtze River Delta (YRD) and Pearl River Delta (PRD) in China (Figure S3). East China Sea, sea

around Japan and Singapore exhibit high shipping emissions due to active shipping transportation (Figure S3). It is estimated that international shipping contributed about 10% to the global SO_2 emissions, and together with aviation contribute more than 10% of global NO_x emissions (Janssens-Maenhout et al., 2015).

2.3 Boundary conditions

To predict more realistic spatial and temporal variations of air pollutants, boundary conditions from global chemical transport models are necessary to drive regional chemical transport models (Carmichael et al., 2008b). Simulations of two global chemical transport models (i.e., GEOS-Chem and MOZART) were used as boundary conditions for MICS-ASIA III. GEOS-Chem was developed in the USA to simulate tropospheric chemistry driven by assimilated meteorology (Bey et al., 2001). The National Center for Atmospheric Research (NCAR) also provides global simulations of atmospheric chemistry (MOZART model) and an interface to convert them to WRF-Chem boundary conditions (Emmons et al., 2010), and NASA provides global aerosol distributions using the global GOCART chemistry model (Chin et al., 2002). GEOS-Chem was run with 2.5ºx2º resolution and 47 vertical layers. The MOZART-4 simulations were configured at the horizontal resolution of 2.8ºx2.8º, and with 28 vertical levels. NASA GOCART was configured at the same resolution as GEOS-5 meteorology (1.25ºx1º). As listed in Table 1, M1 used climatological data from the NOAA Aeronomy Lab Regional Oxidant Model (NALROM), while M2 used boundary conditions from the MOZART-4 (provided from the NCAR website). M3 and M4 used MOZART-4 as boundary conditions for gases and used GOCART as boundary conditions for aerosols. M6 also used fixed climatology boundary conditions, and M5 and M7 used GEOS-Chem outputs as boundary conditions. The spatial distribution of near surface concentrations of major gases and aerosols from both GEOS-Chem and MOZART are shown in

Figure S4. Even though the same global model is used as boundary conditions, the treatments of inputs might differ in details, which might lead to dissimilarities. In MICS-ASIA II, Holloway et al. (2008) discussed the impacts of uncertainties in global models on regional air quality simulations.

2.4 Observation data

Historically, the lack of reliable air quality measurements in Asia has been a bottleneck in understanding air quality and constraining air quality modeling in Asia. Beginning with MICS-ASIA II, observational data from Acid Deposition Monitoring Network in East Asia (EANET) has been used to evaluate model performance. 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 (EMEP). As of 2010, there are 54 wet deposition sites and 46 dry deposition sites in 13 participating countries. Quality assurance and quality control measures are implemented at the national levels and in the Inter-laboratory Comparison Project schemes to guarantee high quality dataset. EANET supported current activities of MICS-ASIA III, and provided measurements in 2010 to all modeling groups. More information about EANET dataset can be found in http://www.eanet.asia/.

In addition to EANET data, measurements of air pollutants and aerosol optical depth (AOD) collected at the Campaign on Atmospheric Aerosol Research network of China (CARE-China) (Xin et al., 2015) network were also used. Previous successful networks in Europe and the United States underscored the importance of building comprehensive observational networks of aerosols in China to get better understanding of the physical, chemical and optical properties of

atmospheric aerosols across China. As the first comprehensive attempt in China, CARE-China was launched in 2011 by Chinese Academy of Sciences (CAS) (Xin et al., 2015). Before launching this campaign, CAS had already been measuring air pollutants and AOD at some CARE-China sites. Table 2 summaries the locations and characteristics of the CARE-China measurements for January 2010. Air quality measurements include concentrations of $PM_{2.5}$, PM_{10}, SO_2, NO_2, NO, CO, O_3.

In addition, AOD from Aerosol Robotic Network (AERONET) (https://aeronet.gsfc.nasa.gov/) and operational meteorological measurements (near surface temperature, humidity, wind speed and downward shortwave radiation) in China and atmospheric sounding data in Beijing were used. AERONET provides long-term, continuous, readily accessible and globally distributed database of spectral AOD, inversion products and precipitable water. AOD data are calculated for three quality levels: Level 1.0 (unscreened), Level 1.5 (cloud screened), and Level 2.0 (cloud screened and quality assured) (Holben et al., 1998). The locations and characteristics of the AERONET measurements are also summarized in Table 2. In-situ measurements of meteorological data from standard stations in China are operated by China Meteorological Administration (CMA) and different levels of data, including daily, monthly, and annually, are open to the public (http://data.cma.cn/en). The locations of all used observational sites are marked in Figure S5, Figure S6 and Figure S7.

The meteorology measurements (locations are shown in Figure S5) were averaged and compared with model results that averaged across those locations. The radiation measurements were averaged and compared against model results in North China and South China (locations are shown in Figure S6), separately. The CARE-China, AERONET and EANET measurements

(locations are shown in Figure S6 and S7) were compared against model results site by site, and model ensemble mean values were made by averaging all model results.

2.5 Analysis methodology

All groups participating in Topic 3 were requested to simulate meteorology, air quality, radiative forcing and effects of aerosols over the Beijing-Tianjin-Hebei region of east China during two periods: January 2010 and January 2013. Each group was requested to submit the following fields from their simulations.

(1) hourly mean meteorology:

(a) air temperature and water vapor mixing ratio at 2m above ground (T2, Q2), wind speed at 10m above groud (WS10), and shortwave radiation flux (Wm-2) at the surface;

(b) above variables (except shortwave radiation flux) at 1km and 3km above ground.

(2) hourly mean concentrations:

(a) SO_2, NO_x, CO, O3, $PM_{2.5}$, PM_{10} and sulfate, nitrate, ammonium, BC, OC and dust in $PM_{2.5}$;

(b) above variables at 1km and 3km above ground.

(3) hourly mean AOD, aerosol direct radiative forcings at the surface, top of the atmosphere (TOA) and inside the atmosphere (single scattering albedo is an option for participants).

(4) Hourly mean integrated liquid water, cloud optical depth.

(5) Changes in T2, Q2, WS10 and $PM_{2.5}$ concentrations at the surface due to both direct and indirect aerosol's effects.

We calculated multiple model evaluation metrics, including correlation coefficient (r), root mean square error (RMSE), mean bias error (MBE), normalized mean bias (NMB), mean fractional bias (MFB) and mean fractional error (MFE). The equations are presented in supplemental information.

3 General description of meteorology and haze during the study period

Winter haze events frequently happen in east China, which is partially due to the stagnant weather conditions in winter. Here we present general descriptions of the meteorological conditions during January 2010 using the NCEP/NCAR reanalysis products. Figure S8 (a) displays the monthly mean T2 (temperature at 2m) and W10 (wind speeds at 10m). WS10 were very weak in eastern and central China regions, while lower T2 in Mongolia region was associated with Siberian High. As shown in Figure S8 (b), the Siberian High center was about 1040mb, and there was no significant precipitation in North China and heavy rainfall only occurred in Southeast Asia regions. During winters, northern China burns coal for heating, generating more emissions. Under stagnant weather conditions, haze episodes are easily triggered. High concentrations of aerosols during this month provide great opportunity to study aerosol-radiation-weather interactions.

4 Results and discussions

In this section, we present some major features of model performances in meteorological and chemical variables for the January 2010 period. Detail analysis of feedbacks and radiative

forcing are presented in MICS-ASIA III companion papers. Heavy haze occurred over broad regions of East China in January 2010. The plots of observed meteorological variables and $PM_{2.5}$ in Beijing show the general situation (Figure 3). Elevated $PM_{2.5}$ occurred during three periods separated in time by roughly one week (January 8, 16 and 26). The major event occurred during January 15-21. The events occurred during periods of low wind speeds, and increasing temperature and relative humidity. The high $PM_{2.5}$ concentrations during January 15-21 also greatly reduce the downward shortwave radiation. Below we evaluate how well the models predict these features.

4.1 Evaluation of meteorological variables

Air quality is affected by not only emissions, but also meteorological conditions. Meteorology affects air quality through altering emissions, chemical reactions, transport, turbulent mixing, and deposition processes (Gao et al., 2016c). Thus, it is important to assess how well these participating models reproduced meteorological variables. The predicted temperature at 2m high (T2), water vapor mixing ratio at 2m (Q2), wind speed at 10m high (WS10) and daily maximum downward shortwave radiation (SWDOWN) were evaluated against near surface observations at the CMA sites.

Figure 4 (a-c) shows the comparisons between simulated and observed daily mean T2, Q2 and WS10 averaged over stations in East China (locations are shown in Figure S5) during January 2010, along with multi-model ensemble mean and observation standard deviation. The calculated correlation coefficients between models and observations are also shown in Figure 5 and other calculated model evaluation metrics are summarized in Table 3. In general, the simulated magnitudes and temporal variations of T2 and Q2 show high order of consistencies with

observations, with correlation coefficients ranging from 0.88 to 1. For T2, models tend to have a cool bias; M1 and M2 have the lowest RMSE (0.64 and 0.68), lowest MBE (-0.19 and -0.60) and lowest NMB (-0.07% and -0.22%) values (Table 3). For Q2, most models tend to slightly overestimate; M1 and M2 have the best performance, with the lowest RMSE (0.14 and 0.10), lowest MBE (0.02 and -0.01), and lowest NMB (0.84% and -0.55%) values (Table 3).

Simulated WS10 exhibit larger diversity of results. All models tend to overestimate WS10, with MBE ranging from 0.15m/s to 2.37m/s. Overestimating wind speeds under low wind conditions is a common problem of current weather forecasting models, and many factors, including errors in terrain data and reanalysis data, relatively low horizontal and vertical model resolutions, as well as poorly parameterized urban surface effect, contribute to these overestimations. From the calculated RMSE, MBE, and NMB listed in Table 3, M2, M5 and M7 show better skills in capturing WS10. In addition, the multi-model ensemble mean show the lowest RMSE for Q2, and also better skills than most models for T2 and WS10. The correlation coefficients between multi-model ensemble mean and observations are 0.99, 0.99 and 0.98 for T2, Q2 and WS10, respectively.

The accuracy of radiation predictions is of great significance in evaluating aerosol-radiation-weather interactions. We evaluated simulated daily maximum SWDOWN averaged over sites in northern China and southern China separately in January 2010 against observations. The locations of the radiation sites are shown in Figure S6. As shown in Figure 4 (d), over stations in northern China, all models except M6 and M7 reproduce daily maximum SWDOWN well, with correlation coefficients ranging from 0.72 to 0.94. The poor performance of M6 in North China is caused by largely overpredicted liquid water path (LWP) over North China (Figure S9).

SWDOWN decreases under conditions of high PM, as shown for example on January 9 and 15-21. This is one of the important reasons for coupled air quality and meteorology modeling, as they can account for this effect of aerosols. It is worth noting that most models predict higher daily maximum SWDOWN compared to observations when severe haze happened in the North China Plain (16-19 January 2010), indicating aerosol effects on radiation might be underestimated. Clouds are also important to alter radiation. To exclude its impacts on the radiation shown here, we calculated the reduction ratio of radiation due to clouds using radiation prediction for clear sky and with clouds from M2 (shown in Figure S10). During the severe haze period (16-19 January 2010), the averaged reduction fraction is 5.9% in north China and 4.2% in south China, suggesting the relatively lower radiation during this period shown in Figure 4(d) is mainly caused by aerosols, while the lowest radiation on 20 January was caused by clouds (Figure 4(d) and Figure S10). Over southern China sites (Figure 4e), M6 and M7 show a better consistence with observations than over northern China sites. According to the calculated RMSE listed in Table 3, M3 and multi-model ensemble mean exhibit relatively better performance in capturing the observed time series of daily maximum SWDOWN in both northern China and southern China.

The above comparisons show that T2 and Q2 are reproduced well by the participating models, and WS10 is overestimated by all models. Emery et al. (2001) proposed that excellent model performance would be classified as wind speed RMSE smaller than 2 m/s, and wind speed bias smaller than 0.5 m/s. Based on the calculated RMSE and MBE of WS10 shown in Table 3, RMSE values from all models match the proposed RMSE threshold but MBE values are higher than 0.5 m/s. The vertical distributions of temperature, water vapor mixing ratio and wind speeds were also validated against atmospheric sounding data in Beijing at 1km and 3km (Figure S11,

487 averaged at 00:00 and 12:00 UTC) (http://weather.uwyo.edu/upperair/sounding.html). The
488 magnitudes of temperature, water vapor mixing ratio and wind speeds from different models are
489 generally consistent with each other at 1km and 3km, but variations are larger near the surface.

490 **4.2 Evaluation of air pollutants**

491 Figure 5 displays the daily averaged predicted and observed SO_2, NO_x, CO, O_3, $PM_{2.5}$, and PM_{10}
492 at the Beijing station, along with the observation standard deviation (locations are shown in
493 Figure S7). Comparisons for the Tianjin, Shijiazhuang and Xianghe sites are shown in Figure
494 S12-S14. M6 only provided SO_2, NO_x concentrations, so it is not shown in the plots of CO, O_3,
495 $PM_{2.5}$, and PM_{10}. The observed and predicted primary pollutants and $PM_{2.5}$ and PM_{10} show the
496 same monthly variations with elevated values at roughly weekly intervals, with the largest event
497 occurring during January 15-21. For example, as shown in the comparisons of SO_2
498 concentration, the temporal variations are reproduced well by all the models, but peak values are
499 overestimated or underestimated by some models. Based on the calculated MBE values shown in
500 Table 4, all models except M2 tend to underestimate SO_2 in the CARE-China sites. M1 shows
501 the highest correlation (0.90) with SO_2 observations in the Beijing site, and most other models
502 show similar good correlations. The multi-model ensemble mean shows a better agreement with
503 observations with a higher correlation of 0.92, and it falls within the range shown with standard
504 deviation error bar. In general, the predictions for NO_x capture the main features in the
505 observations, with slightly less skill than for the SO_2 prediction. The calculated correlation
506 coefficients for NO_x from different models are close to each other, ranging from 0.63 to 0.88. M2
507 and M5 predict higher NO_x concentrations than observations and other models (MBE in Table
508 4). All models overestimate NO_x concentration in Shijiazhuang (Figure S14), suggesting NO_x

emissions in Shijiazhuang might be overestimated in the MIX emission inventory. All models produce similar CO predictions.

$PM_{2.5}$ concentrations are well modelled, with high correlation coefficients ranging from 0.87 to 0.90 in Beijing, from 0.83 to 0.93 in Tianjin, and from 0.74 to 0.91 in Xianghe. The correlation coefficient of the multi-model ensemble mean for $PM_{2.5}$ reaches 0.94 (Table 4), better than any individual model. The performances of all participating models in reproducing PM_{10} variations are not as good as reproducing $PM_{2.5}$. M1 and M2 overestimate PM_{10} concentrations, and other models underestimate PM_{10} concentrations (MBE in Table 4). These biases are probably related to different treatments of primary aerosols and anthropogenic dust in the models.

The models showed the poorest skill in predicting ozone. All models exhibit different performances in simulating ozone concentrations, and the correlation coefficients between models and observations can reach negative values (Figure S12). M3 and M4 tend to overestimate ozone concentrations, M2 slightly overestimates it, and M1, M5, and M7 slightly underestimate it (MBE in Table 4). According to the calculated RMSE in Table 4, M1 and M7 shows relatively better performance in modeling ozone variations. Although WRF-Chem and NU-WRF models were applied at three institutions, different gas phase chemistry schemes were used, which leads to these diversities among predicted ozone concentrations. The impacts of gas phase chemical mechanisms on ozone simulations have been investigated in Knote et al. (2015).

Figure 6 shows the comparisons between modeled and observed ground level daily averaged concentrations of SO_2, NO_x, O_3 and PM_{10} during January 2010 at the Rishiri site in Japan from EANET. The locations of EANET sites are marked in Figure S7. Comparisons at other EANET sites are shown in Figure S15-S18. The models are able to predict the major features in the observations. For example, low values of most pollutants are observed (and predicted) during the first half of the month, followed by elevated values, which peak on January 21. For SO_2, most models show similar capability in producing the temporal variations in observations with slight

underestimation (MBE in Table 5). According to the calculated RMSE averaged over all the EANET sites, M2 and the multi-model ensemble mean performed the best. For NO_x, the multi-model ensemble mean shows lower RMSE than any individual model (Table 5). Similar to the comparisons over CARE-China sites, large discrepancies exist in ozone predictions, but the model ensemble mean still shows the lowest RMSE for ozone predictions. PM_{10} concentrations are largely underestimated by M1 (largest negative MBE: -21.03ug/m^3) and overestimated by M5 (highest positive MBE: 3.77ug/m^3) (Table 5), which could also be related to differences in the way sea-salt emissions are treated in the various models. Spatial distributions of monthly near surface concentrations of SO_2, NO_x, O_3 and CO for January 2010 from all participating models are shown in Figure S19. The aerosol spatial distributions are discussed in the following section.

4.3 $PM_{2.5}$ and $PM_{2.5}$ chemical composition distribution

Due to different implementations of chemical reactions in the models, predicted $PM_{2.5}$ chemical compositions from participating models differ largely. Figure 7 and Figure 8 show the predicted monthly mean concentrations of sulfate, nitrate, ammonium, BC and OC in $PM_{2.5}$ from all participating models for January 2010.

M1, M2, M3, M4 and M7 all predict quite low sulfate concentrations in east China, but with considerably enhanced sulfate in southwest areas of China and west areas of India. M5 and M6 show similar spatial patterns of sulfate except that M6 produces higher concentrations. The chemical production of sulfate is mainly from gas-phase oxidation of SO_2 by OH radicals and aqueous-phase pathways in cloud water. In cloud water, dissolved SO_2 can be oxidized by O_3,

556 H_2O_2, Fe(III), Mn(II), and NO_2 (Seinfeld and Pandis, 2016). Most chemical transport models have
557 included the above gas phase oxidation of SO_2 by OH and oxidation of SO_2 by O_3 and H_2O_2 in
558 aqueous phase. Under hazy conditions, radiation is largely reduced due to aerosol dimming effects,
559 and sulfate formation from gas phase and aqueous phase oxidation processes are slowed down,
560 which tend to reduce sulfate concentration. However, field observations exhibit an increase in
561 sulfate concentration during haze episode (Zheng et al., 2015). Cheng et al. (2016) proposed that
562 the reactive nitrogen chemistry in aerosol water could contribute significantly to the sulfate
563 increase due to enhanced sulfate production rates of NO_2 reaction pathway under high aerosol pH
564 and elevated NO_2 concentrations in the North China Plain (NCP). Wang et al. (2016) also pointed
565 out the aqueous oxidation of SO_2 by NO_2 is key to efficient sulfate formation on fine aerosols with
566 high relative humidity and NH_3 neutralization or under cloudy conditions. Besides, Zheng et al.
567 (2015) suggested that heterogeneous chemistry on primary aerosols could play an important role
568 in sulfate production and lead to increasing sulfate simulation during haze episodes. X. Huang et
569 al. (2014) found including natural and anthropogenic mineral aerosols can enhance sulfate
570 production through aqueous-phase oxidation of dissolved SO_2 by O_3, NO_2, H_2O_2 and transition
571 metal. Gao et al. (2016), Wang et al. (2014), and Zhang et al. (2015) also emphasized the
572 importance of multiphase oxidation in winter sulfate production. However, these processes are
573 currently not incorporated in the participating models for this study, which might be responsible
574 for the apparent under-predictions of sulfate concentration (Figure 9). M5 incorporated
575 heterogeneous chemical reactions on aerosol surface (Li and Han, 2010), which enhances total
576 sulfate production.

577 M1 and M5 predict relatively small nitrate and ammonium concentrations; while M2, M6 and
578 M7 produce similar magnitudes and spatial patterns of nitrate. Nitrate formation involves both

579 daytime and nighttime chemistry. During daytime, NO_2 can be oxidized by OH to form nitric
580 acid (HNO_3), and by ozone to form NO_3. HNO_3 is easily removed by dry or wet deposition, but
581 NO_3 is easily photolyzed back to NO_2. During nighttime, NO_3 is the major oxidant, which oxides
582 NO_2 to form dinitrogen pentoxide (N_2O_5). Homogenous reaction of N_2O_5 with water vapor is
583 possible but very slow while heterogeneous uptake of N_2O_5 onto aerosol particles has been
584 identified as a major sink of N_2O_5 and an important contributor to particulate nitrate (Kim et al.,
585 2014). The MOSAIC aerosol module (Zaveri et al., 2008) coupled with CBMZ gas phase
586 chemistry in WRF-Chem already includes heterogeneous uptake of N_2O_5 since version v3.5.1
587 (Archer-Nicholls et al., 2014), which is the version used by M2, leading to the high production of
588 nitrate. An et al. (2013) incorporated photoexcited nitrogen dioxide molecules, heterogeneous
589 reactions on aerosol surfaces, and direct nitrous acid (HONO) emissions into the WRF-Chem
590 model and found these additional HONO sources can improve simulations of HONO and nitrate
591 in north China. M7 also predict high nitrate concentrations (N_2O_5 and NO_2 gases react with
592 liquid water, Zheng et al., 2015), and the predicted lower nitrate concentrations from other
593 models are probably due to missing aqueous phase and heterogeneous chemistry, or the
594 implementations of different gas phase oxidation in these models. Many studies have been
595 conducted regarding sulfate formation issues. Nitrate also account for a large mass fraction in
596 $PM_{2.5}$ during winter haze events in north China, yet less attention was attracted to fully
597 understand its formation. It is worth furtherly digging into the details about how different
598 processes contribute to high nitrate concentrations in future studies. M3 and M4 do not include
599 the explicit nitrate and ammonium treatment but ammonium is implicitly considered in total
600 $PM_{2.5}$ mass estimate.

601 The predicted ammonium concentrations are associated with the amounts of sulfate and nitrate,
602 as shown by its similar spatial distribution to sulfate and nitrate. NH_3 neutralizes H_2SO_4 and
603 HNO_3 to form aerosol, so its amount can affect the formation of sulfate, nitrate and ammonium.
604 Since the same emission inventory was used, the amount of ammonia available for neutralizing
605 will not vary greatly among these models. Thus, the rates of H_2SO_4 and HNO_3 production
606 determines the amounts of ammonium. For example, the produced ammonium concentrations are
607 small in M1, similar to their sulfate and nitrate productions. High ammonium concentrations are
608 predicted from M6, due to high productions of nitrate and sulfate (Figure 7).

609 The spatial distributions and magnitudes of predicted BC from all participating models are
610 similar to each other as BC is a primary pollutant whose mass as BC is not impacted by chemical
611 reactions. The concentrations of BC in the atmosphere are mainly influenced by PBL mixing and
612 diffusion, aging, deposition (dry deposition and wet scavenging) and advection. Predicted BC
613 from M2 and M7 are higher than from other models, which might be caused by the treatment of
614 aging and deposition (dry deposition and wet scavenging) processes. For example, in the
615 GOCART aerosol model (M3 and M4), 80% of BC are assumed to be hydrophobic and then
616 undergo aging to become hydrophilic in an e-folding time of 1.2 days. Hydrophilic aerosols will
617 go through wet deposition. But in other models like M2 and M7, BC is assumed to be
618 hydrophobic, thus the wet removal is less.

619 The disparity among predicted OC concentrations is mainly associated with the different
620 treatments of SOA production, given the POC prediction is generally consistent among models
621 using the same emission inventory. The predicted OC concentrations from M1, M2, and M7 are
622 close to each other. M1 uses SORGAM (Secondary Organic Aerosol Model) to simulate SOA, but
623 M2 and M6 did not include any SOA formation mechanism. The similar magnitudes of OC from

M1 suggests that SORGAM in M1 does not produce appreciable amounts of SOA, which is consistent with the findings in Gao et al. (2016a). Although SOA formation is implemented in M5, the production is relatively weak compared to M3 and M4. In the atmosphere, SOA is mainly formed from the condensation of semi-volatile VOCs from oxidation of primary VOCs. An empirical 2-produt model (Odum et al., 1996) is often used to simulate SOA formation, but this method was reported to significantly underestimate measured SOA mass concentrations (Heald et al., 2008). Later, the volatility basis-set approach (Donahue et al., 2006) was developed to represent more realistically the wide range of volatility of organic compounds and more complex processes, and it was found to increase SOA production and to reduce observation-simulation biases in many regions with high emissions (Tsimpidi et al., 2010) including east China (Han et al., 2016). It was also suggested that primary organic aerosols (POA) are semi-volatile and can evaporate to become SOA precursors, which promotes the understanding and improvements of SOA modeling (Kanakidou et al., 2005). In M5, the SOA production is calculated using a bulk yield method via Lack et al. (2004), in which the amount of SOA able to be produced from a unit of reacted VOC from anthropogenic and biogenic origins are used to represent SOA yields. However, the SOA concentration is highly dependent on the yield data. During haze episodes, photochemistry is reduced due to the aerosol dimming effect, thus aqueous reaction processes on aerosol water and cloud/fog water could become much more important in producing SOA. R. Huang et al. (2014) also suggested that low temperature does not significantly reduce SOA formation rates of biomass burning emissions. Most models over-simplified SOA formation.

In M3 and M4, SOA is treated by assuming that 10% of VOCs from terrestrial source are converted to OC (Chin et al., 2002), and these models produce high OC concentrations, with a major contribution from SOA. The 10% yield rate could be unrealistically high during hazy days because

solar radiation was much reduced. Zhao et al. (2015) comprehensively assessed the effect of organic aerosol aging and intermediate-volatile emissions on OA formation and confirmed their significant roles. All these results suggest more complicated SOA scheme are needed to improve organic aerosol simulations during haze events.

The different predictions of $PM_{2.5}$ chemical components lead to differences in $PM_{2.5}$ and PM_{10} concentrations for January 2010, which are shown in the last row of Figure 8. Although spatial distributions of $PM_{2.5}$ from these models are similar, the underlying causes are different. M2, M3 and M5 simulated higher $PM_{2.5}$ levels in deserts of west China, which are contributed by wind-blown dust. M1 and M7 fail to produce high $PM_{2.5}$ concentrations in the deserts of west China, due to omission of dust emissions. M4 presented results in a smaller domain excluding west China. The enhanced $PM_{2.5}$ concentrations in Central China from M2 and M7 are caused by large nitrate production, as shown in Figure 7.

The differences in the predictions of aerosols composition discussed above can be seen clearly in the comparisons at the Beijing site on 13-23 January when a haze event occurred in the NCP (Figure 9). Also shown are the observed values. Most models fail to produce the observed high sulfate concentrations. Only the sulfate predictions from M5 are close to the observed values. Sulfate is much lower than observed for all other models, except M6 which is too high. M2 and M7 predict reasonable nitrate concentrations. M3 and M4 overpredict OC during the haze period, but other models underpredict OC concentrations.

Figure 10 and 11 show the ensemble mean monthly averaged near-surface $PM_{2.5}$, $PM_{2.5}$ composition, along with the spatial distribution of the coefficient of variation. The coefficient of variation is defined as the standard deviation divided by the average (Carmichael et al., 2008b),

and larger values indicate lower consistency among models. Mean concentrations of $PM_{2.5}$ and $PM_{2.5}$ chemical compositions are high in Sichuan Basin and east China. High coefficient of variation are shown in North China for sulfate, and in most areas for nitrate and OC. The diversity in predictions of these species are caused by complexity of secondary formation and different treatments in models as discussed earlier. Higher consistency is shown for model BC with coefficient of variations less than 0.3 in most areas (Figure 10(h)). Coefficient of variations for $PM_{2.5}$ are also low in North China region, which is consistent with good performance of $PM_{2.5}$ predictions shown in above comparisons. However, the coefficient of variation can reach above 1.6 in northwestern regions, partially due to discrepancies in dust predictions.

4.4 Evaluation of AOD

AOD is an indication of aerosol pollution, which tells us how much sunlight is blocked from reaching the surface by suspended aerosols. We used the measurements of AOD at AERONET and CARE-China sites to evaluate how participating models perform in simulating AOD. The submitted AOD from all models except M6 are at 550nm, and AOD from M6 are at 495nm. We used Angstrom exponent relation (Schuster et al., 2006) to convert AOD from M6 at 495nm to 550nm, and all used AERONET and CARE-China AOD data to 550nm. The locations of AERONET and CARE-China AOD measurement sites are shown in Figure S5. Daytime mean AOD are calculated in pairwise manner and the comparisons and performance statistics are shown in Figure 12, 13, and Table 6. On some days, data are missing because AOD cannot be retrieved under serious pollution and cloudy conditions (Gao et al., 2016a). On days with data, the variations of AOD are captured well by all models. However, large disparities exist among models in the simulated peak AOD values (factor of 2) at monitoring stations during the severe

691 haze episode on 15-20 January 2010 (Figure 12 and Figure 13). The participating models exhibit
692 various skill in simulating AOD temporal variation at different sites.

693 At CARE-China sites, M7 produces the best correlation coefficient R (0.83) among models at
694 Baoding and Beijing forest sites, M2 produces the highest R (0.86) at Cangzhou site, whereas
695 M5 shows the highest R (0.93) at the Beijing city site. At AERONET sites, M7 shows the
696 highest R (0.81) at Beijing, whereas M2 and M5 produce R as high as 0.91 at Xianghe site,
697 which is about 60km southeast of downtown Beijing. In terms of AOD magnitude, it is
698 interesting to note that during the severest haze days around 19 January 2010, M2 consistently
699 simulates the highest AOD among models, followed by M5 and M7. M6 simulates the lowest,
700 and other models in the middle at the sites (Baoding, Beijing City, Beijing Forest, Cangzhou,
701 Beijing, Xianghe) in the north China plain (NCP).

702 In M1, M5, and M7, particle size distribution is described by a lognormal function with a geometric
703 mean radius and a geometric standard deviation basically based on OPAC (Optical properties of
704 aerosols and clouds) database (Hess et al. 1998). In M3 and M4, sulfate, BC and OC are
705 parameterized in bulk mode, and a sectional scheme is used for sea-salt and dust aerosols. M2 uses
706 an 8 bins sectional aerosol scheme with size sections ranging from 39nm to 10μm. The refractive
707 index of various aerosol components in the models are mainly taken from d'Almeida et al. (1991)
708 or OPAC database. All models except M6 use a kappa (κ) parameterization (Petters and
709 Kreidenweis, 2007), in which the aerosol hygroscopicity κ largely varies among different aerosol
710 chemical components, such as $\kappa=0$ for black carbon, and $\kappa>0.6$ for inorganic aerosols, but the
711 prescribed κ values could be different in the above models. M6 uses a different hygroscopic growth
712 scheme following Kiehl and Briegleb (1993). WRF-Chem models assume internally mixing

among aerosols within each mode (or size bin) and externally mixing between modes (or size bins), M5 assumes inorganic and carbonaceous aerosols are internally mixed and externally mixed with soil dust and sea-salt. M6 uses an external mixture assumption among aerosols except for hydrophilic BC, which is internally mixed with other aerosols in a core-shell way.

We first look at the mass concentrations of different aerosol components because of their important roles in determining optical properties. The observed total inorganic aerosol concentration in Beijing on 19 January 2010 is about 130μg/m^3, with sulfate and nitrate being about 50 and 65μg/m^3, respectively (Figure 9). The models generally predict a much lower sulfate concentration except that the prediction from M5, which is close to observations, and M6, which shows an overprediction. Most models predict lower nitrate concentration, in contrast to the overprediction by M2. In terms of inorganic aerosols, which have a similar optical properties, the total concentration (the sum of sulfate, nitrate and ammonium) from M2 (175μg/m^3) is higher than observation and other models, and this can explain the largest simulated AOD by M2. M6 simulates a similar level of inorganic aerosols to M2, but the simulated AOD is lower than other models, which could be due to a weak hygroscopicity or lower simulated RH (see Figure S20). For example, high RH on January 19 are captured by M2 and M6, but underpredicted by M6 (Figure S14a). Although M3 and M4 largely overpredict OC concentrations, their simulated AOD are lower than M1 and M5 because their simulated inorganic aerosol concentrations are much lower and OC has a smaller (mass) extinction coefficient than inorganic aerosols. M1 predicts about three times larger BC concentration than the observations, although the mass extinction coefficient of BC is even larger than inorganic aerosols, the mass concentration and hygroscopicity of BC are much smaller and weaker than that of inorganic aerosols, leading to relatively lower AOD from M1 simulation. M5 and M7 predict a similar level of inorganic aerosol concentrations

(80~90μg/m^3) and use a similar hygroscopic growth scheme, and this can help explain their consistency in the simulated AOD magnitude.

As listed in Table 1, only M6 uses external mixing for aerosols, and internal mixing is assumed by other models for major aerosol compositions. Curci et al. (2015) discussed the impacts of mixing state on simulated AOD and found that external mixing state assumption significantly increase simulated AOD. M6 used external mixing but shows a relative lower AOD because of ignorance of other aerosol species like dust, sea-salt, etc. In general, it appears the magnitude of inorganic aerosol concentrations and the hygroscopic growth efficiency (affected by varied RH) can account for or explain the simulated variations and magnitudes of AOD in Beijing during the severe haze event, given the aerosol size lognormal treatments and mixing state are alike among models.

Table 6 shows the statistics for AOD simulation at NCP sites and at all sites. In the NCP region, R ranges from 0.36~0.74 for all the models. M2, M5 and M7 produce R around 0.7, indicating a better simulation of AOD variations. M2 and M7 exhibit the best R (0.65) for all sites. It is noteworthy that R values at the sites in NCP are larger than that at all sites, indicating the larger reliability of model inputs (emissions and boundary conditions) and meteorological simulations. In terms of magnitudes, all models tend to underpredict AOD in the whole domain, with NMB of -2.7 to -71% in the NCP, and larger biases (NMB of -21~-75%) at all sites. M7 shows the smallest MBE (-0.05) and NMB (-2.7%) and M2 produces the smallest RMSE. It is interesting to note that the simulated AOD from the WRF-Chem models differed largely (-12 to -71%) between M1 and M3 at the NCP sites, and the WRF-Chem model using finer grid size (M4) can produced slightly smaller NMB compared with the same model using larger grid size (M3). However, as grid size becomes finer, R and RMSE from M4 may become worse, although AOD magnitude improved. The effect of grid resolution will be a topic of future paper.

5 Summary

The MICS-Asia Phase III Topic 3 examines how current online coupled air quality models perform in reproducing extreme aerosol pollution episodes in North China, and how high aerosol loadings during these episodes interact with radiation and weather. Predicted meteorological variables and air pollutants from these modeling groups were compared against each other, and measurements as well. A new anthropogenic emission inventory was developed for this phase (Li et al., 2017), and this inventory along with biogenic, biomass burning, air and ship, and volcano emissions were provided to all modeling groups. All modelling groups were required to submit results based on the analysis methodology that documented in this paper.

This paper focused on the evaluation of the predictions of meteorological parameters and the predictions of aerosol mass, composition and optical depth. These factors play important roles in feedbacks impacting weather and climate through radiative and microphysical processes. Comparisons against daily meteorological variables demonstrated that all models can capture the observed near surface temperature and water vapor mixing ratio, but near surface wind speeds were overestimated by all models to varying degrees. The observed daily maximum downward shortwave radiation, particularly low values during haze days, were represented in the participating models. Comparisons with measurements of air pollutants, including SO_2, NO_x, CO, O_3, $PM_{2.5}$, and PM_{10}, from CARE-China and EANET networks showed that the main features of accumulations of air pollutants are represented in current generation of online coupled air quality models. The variations in observed AOD from CARE-China and AERONET networks were also reproduced by the participating models. Differences were found between simulated air pollutants, particularly ozone. While winter time ozone levels are typically low

781 (below 40 ppb) as photochemical pathways are slow, the models captured the synoptic variability
782 but differed in the absolute magnitudes of near surface concentrations. The role of dry deposition
783 and the boundary conditions play important roles.

784 Large differences in the models were found in the predicted $PM_{2.5}$ chemical compositions,
785 especially secondary inorganics and organic carbon. During winter haze events, the production
786 from gas phase chemistry is inhibited, and whether including other aerosol formation pathways
787 (such as aqueous phase chemistry), or how the chemistry is parametrized leads to the large
788 difference between simulated concentrations of secondary inorganic aerosols. In addition,
789 differences in treatments of SOA also lead to large discrepancies between simulated OC
790 concentrations. Differences in the simulated variations and magnitudes of AOD in Beijing during
791 the severe haze event could be explained by the differences in simulated inorganic aerosol
792 concentrations and the hygroscopic growth efficiency (affected by varied RH).

793 Results of this inter-comparison show that there remain important issues with current coupled
794 models in predicting winter haze episodes. Low wind speeds play an important role in haze
795 episodes. Current models can predict the low wind speed - high haze relationship, but
796 overestimate the low wind speeds. This contributed to the underestimation of $PM_{2.5}$. The models
797 also underestimate the production of secondary inorganic aerosols. There is currently a great deal
798 of research focused on inorganic aerosol production under winter haze conditions and new
799 pathways need to be included in the models to improve prediction skills. Furthermore, current
800 models have various treatments of SOA production and these lead to wide differences in the
801 contribution of SOA to winter haze episodes.

However, it was also found that using the ensemble mean of the models produced the best prediction skill. While this has been shown for other conditions (for example prediction of high ozone events in the US (Mckeen et al., 2004), this is to our knowledge the first time it has been shown for heavy haze events. The uncertainties in predictions in aerosols mass and composition will impact estimates of the aerosol direct and indirect effects during haze event (Gao et al., 2017a, 2017b, 2017c). The results of the MICS-Asia Topic 3 experiments looking at the direct and indirect effects during these heavy haze events is the subject of companion papers.

ACKNOWLEDGMENTS

The authors would like to acknowledge support of this project from National Natural Science Foundation of China (NSFC (41620104008)), and ground measurements from Yuesi Wang's research group. The ground observation was supported by the National Natural Science Foundation of China (41222033; 41375036) and the CAS Strategic Priority Research Program Grant (XDA05100102, XDB05020103).

Reference

Akimoto, H. (2003). "Global Air Quality and Pollution." Science, 302 (5651), 1716-1719.

An, J., Li, Y., Chen, Y., Li, J., Qu, Y. and Tang, Y., 2013. Enhancements of major aerosol components due to additional HONO sources in the North China Plain and implications for visibility and haze. Advances in Atmospheric Sciences, 30(1), p.57.

Anthes, R. A., A cumulus parameterization scheme utilizing a one-dimensional cloud model, Mon. Wea. Rev., 105, 270–286, 1977.

Archer-Nicholls, S., et al. (2014). "Gaseous chemistry and aerosol mechanism developments for version 3.5.1 of the online regional model, WRF-Chem." Geoscientific Model Development7(6): 2557-2579.

Baklanov, A., et al. (2014). "Online coupled regional meteorology chemistry models in Europe: current status and prospects." Atmospheric Chemistry and Physics14(1): 317-398.

Baklanov, A., Brunner, D., Carmichael, G., Flemming, J., Freitas, S., Gauss, M., Hov, Ø., Mathur, R., Schlünzen, K.H., Seigneur, C. and Vogel, B., 2017. Key issues for seamless integrated chemistry-meteorology modeling. Bulletin of the American Meteorological Society, (2017).

Bey, I., et al. (2001). "Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation." Journal of Geophysical Research: Atmospheres106(D19): 23073-23095.

Carmichael, G. P., L.R. (1984). "An Eulerian transport/transformation/removel model for SO2 and sulfate-I. model development." Atmospheric Environment18(5): 937-951.

Carmichael, G. P., L.R.; Kitada, T. (1986). "A second generation model for regional-scale transport/chemistry/deposition." Atmospheric Environment20(1): 173-188.

Carmichael, G. R., Peters, L.R.; Saylor, R. D. (1991). "The STEM-II regional scale acid deposition and photochemical oxidant model-I. an overview of model development and applications." Atmospheric Environment25A(10): 2077-2090.

Carmichael, G. R., et al. (1998). "Tropospheric ozone production and transport in the springtime in east Asia." Journal of Geophysical Research: Atmospheres103(D9): 10649-10671.

Carmichael, G. R. C., G.; Hayami, H.; Uno, I.; Cho, S.Y.; Engardt, M.; Kim, S.B.; Ichikawa, Y.; Ikeda, Y.; Woo, J.H.; Ueda, H.; Amann, M. (2002). "The MICS-Asia study: model intercomparison of long-range transport and sulfur deposition in East Asia." Atmospheric Environment36: 175-199.

Carmichael, G. R., et al. (2008a). "Predicting air quality: Improvements through advanced methods to integrate models and measurements." Journal of Computational Physics227(7): 3540-3571.

Carmichael, G., et al. (2008b). "MICS-Asia II: The model intercomparison study for Asia Phase II methodology and overview of findings." Atmospheric Environment42(15): 3468-3490.

Carter, W.P., 2000a. Documentation of the SAPRC-99 chemical mechanism for VOC reactivity assessment. Contract, 92(329), pp.95-308.

Carter, W.P., 2000b. Implementation of the SAPRC-99 chemical mechanism into the models-3 framework. Report to the United States Environmental Protection Agency, January, 29.

Cheng, Y. F., Z., G.; Wei, C.; Mu, Q.; Zheng, B.; Wang, Z.; Gao, M.; Zhang, Q.; He, K.; Carmichael, G.; Poschl, U.; Su, Hang (2016). "Reactive nitrogen chemistry in aerosol water as a source of sulfate during haze events in China." Science Advances 2(e1601530)

Chin, M. G., P.; Kinne, S.; Torres, O.; Holben, B.N.; Duncan, B.N.; Martin, R.V.; Logan, J.A.; Higurashi, A.; Nakajima, T. (2002). "Tropospheric aerosol optical thickness from the GOCART Model and Comparisions with satellite and sun photometer measurements." Journal of Atmospheric Sciences(59).

Curci, G., Hogrefe, C., Bianconi, R., Im, U., Balzarini, A., Baró, R., Brunner, D., Forkel, R., Giordano, L., Hirtl, M. and Honzak, L., 2015. Uncertainties of simulated aerosol optical

properties induced by assumptions on aerosol physical and chemical properties: An AQMEII-2 perspective. Atmospheric Environment, 115, pp.541-552.

Donahue, N.M., Robinson, A.L., Stanier, C.O. and Pandis, S.N., 2006. Coupled partitioning, dilution, and chemical aging of semivolatile organics. Environmental Science & Technology, 40(8), pp.2635-2643.

D'Almeida, G. A., P. Koepke, and E. P. Shettle (1991), Atmospheric Aero-sols: Global Climatology and Radiative Characteristics,A.Deepak, Hampton, Va.

Emery, C. T., E.; Yarwood, G. (2001). Enhanced meteorological modeling and performance evaluation for two Texas ozone episodes.

Emmons, L. K. W., S.; Hess, P.G.; Lamarque, J-F.; Pfister G.G.; Fillmore, D.; Granier, C.; Guenther, A.; Kinnison, D.; Laeple, T.; Orlando, J.; Tie, X.; Tyndall, G.; Wiedinmyer, C.; Baughcum, S.L.; Kloster, S. (2010). "Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4)." Geoscientific Model Development3: 43-67.

Fast, J.D., Gustafson, W.I., Easter, R.C., Zaveri, R.A., Barnard, J.C., Chapman, E.G., Grell, G.A. and Peckham, S.E., 2006. Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of Houston using a fully coupled meteorology‐chemistry‐aerosol model. Journal of Geophysical Research: Atmospheres, 111(D21).

Forkel, R., Balzarini, A., Baró, R., Bianconi, R., Curci, G., Jiménez-Guerrero, P., Hirtl, M., Honzak, L., Lorenz, C., Im, U. and Pérez, J.L., 2015. Analysis of the WRF-Chem contributions to AQMEII phase2 with respect to aerosol radiative feedbacks on meteorology and pollutant distributions. Atmospheric Environment, 115, pp.630-645.

Galmarini, S., Koffi, B., Solazzo, E., Keating, T., Hogrefe, C., Schulz, M., Benedictow, A., Griesfeller, J.J., Janssens-Maenhout, G., Carmichael, G. and Fu, J., 2017. Coordination and harmonization of the multi-scale, multi-model activities HTAP2, AQMEII3, and MICS-

898 Asia3: simulations, emission inventories, boundary conditions, and model output
899 formats. Atmospheric Chemistry and Physics, 17(2), pp.1543-1555.

900 Gao, M., et al. (2015). "Health impacts and economic losses assessment of the 2013 severe haze
901 event in Beijing area." Sci Total Environ511: 553-561.

902 Gao, M., et al. (2016a). "Modeling study of the 2010 regional haze event in the North China
903 Plain." Atmospheric Chemistry and Physics16(3): 1673-1691.

904 Gao, M., et al. (2016b). "Improving simulations of sulfate aerosols during winter haze over
905 Northern China: the impacts of heterogeneous oxidation by NO2." Frontiers of
906 Environmental Science & Engineering10(5).

907 Gao, M., et al. (2016c). "Response of winter fine particulate matter concentrations to emission
908 and meteorology changes in North China." Atmospheric Chemistry and Physics16(18):
909 11837-11851.

910 Gao, M., et al. (2017a). "Chemical and Meteorological Feedbacks in the Formation of Intense
911 Haze Events." Air Pollution in Eastern Asia: An Integrated Perspective. Springer, Cham.
912 437-452.

913 Gao, M., et al. (2017b). Distinguishing the roles of meteorology, emission control measures,
914 regional transport, and co-benefits of reduced aerosol feedbacks in “APEC
915 Blue”. Atmospheric Environment, 167, 476-486.

916 Gao, M., et al. (2017c). "Estimates of Health Impacts and Radiative Forcing in Winter Haze in
917 eastern China through constraints of surface PM2.5 predictions." Environ Sci Technol.

918 Gery, M. W. W., G.Z.; Killus, J.P.; Dodge, M.C. (1989). "A photochemical kinetics mechanism
919 for urban and regional scale computer modeling " Journal of Geophysical Research94(D10):
920 12925-12956.

921 Grell, G. A., et al. (2005). "Fully coupled “online” chemistry within the WRF model."
922 Atmospheric Environment39(37): 6957-6975.

923 Guenther, A. K., T.; Harley, P.; Wiedinmyer, C.; Palmer, P.I.; Geron, C. (2006). "Estimates of
924 global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and
925 Aerosols from Nature)." Atmospheric Chemistry and Physics6: 3181-3210.

926 Han, Zhiwei, Hiromasa Ueda, Kazuhide Matsuda, Renjian Zhang, Kimio Arao, Yutaka Kanai,
927 Hisashi Hasome, 2004. Model study on particle size segregation and deposition during
928 Asian dust events in March 2002, Journal of Geophysical Research, 109, D19205, doi:
929 10.1029/2004jd004920.

930 Han, Zhiwei. (2010). "Direct radiative effect of aerosols over East Asia with a Regional coupled
931 Climate/Chemistry model." Meteorologische Zeitschrift, 19(3): 287-298.

932 Han Zhiwei, Jiawei Li, Xiangao Xia, Renjian Zhang, 2012. Investigation of direct radiative
933 effects of aerosols in dust storm season over East Asia with an online
934 coupled regional climate-chemistry-aerosol model. Atmospheric Environment, 54, 688-699.

935 Han Zhiwei, Jiawei Li, Weidong Guo, Zhe Xiong, Wu Zhang, 2013. A study of dust radiative
936 feedback on dust cycle and meteorology over East Asia by a coupled regional climate-
937 chemistry-aerosol model. Atmospheric Environment, 68, 54-63.

938 Han Zhiwei et al.,2016. Modeling organic aerosols over east China using a volatility basis-set
939 approach with aging mechanism in a regional air quality model. Atmospheric Environment
940 124 , 186-198.

941 Heald, C.L., Henze, D.K., Horowitz, L.W., Feddema, J., Lamarque, J.F., Guenther, A., Hess,
942 P.G., Vitt, F., Seinfeld, J.H., Goldstein, A.H. and Fung, I., 2008. Predicted change in global
943 secondary organic aerosol concentrations in response to future climate, emissions, and land
944 use change. Journal of Geophysical Research: Atmospheres, 113(D5).

945 Hess, M., Koepke, P., Schuit, I., 1998. Optical properties of aerosols and clouds: the software
946 package OPAC. Bull. Am. Meteorol. Soc. 79, 831-844.

947 Holben, B. N., Eck, T.F., Slutsker, I., Tanre, D., Buis, J.P., Setzer, A., Vermote, E., Reagan, J.A.,
948 Kaufman, Y.J., Nakajima, T. and Lavenu, F. (1998). "AERONET—A federated instrument

network and data archive for aerosol characterization." Remote sensing of environment66(1): 1-16.

Holloway, T., et al. (2008). "MICS-Asia II: Impact of global emissions on regional air quality in Asia." Atmospheric Environment42(15): 3543-3561.

Holtslag, A. A. M., and B. A. Boville, Local versus nonlocal boundary-layer diffusion in a global climate model, J. Climate, 6, 1993.

Huang, M., Carmichael, G.R., Pierce, R.B., Jo, D.S., Park, R.J., Flemming, J., Emmons, L.K., Bowman, K.W., Henze, D.K., Davila, Y. and Sudo, K., 2017. Impact of intercontinental pollution transport on North American ozone air pollution: an HTAP phase 2 multi-model study. Atmospheric Chemistry and Physics, 17(9), pp.5721-5750.

Huang, X., Song, Y., Zhao, C., Li, M., Zhu, T., Zhang, Q. and Zhang, X., 2014. Pathways of sulfate enhancement by natural and anthropogenic mineral aerosols in China. Journal of Geophysical Research: Atmospheres, 119(24).

Huang, R.J., Zhang, Y., Bozzetti, C., Ho, K.F., Cao, J.J., Han, Y., Daellenbach, K.R., Slowik, J.G., Platt, S.M., Canonaco, F. and Zotter, P., 2014. High secondary aerosol contribution to particulate pollution during haze events in China. Nature, 514(7521), pp.218-222.

Jacobson, M. Z., 2001: Global direct radiative forcing due to multicomponent anthropogenic and natural aerosols. J. Geophys. Res., 106, 1551–1568.

Janssens-Maenhout, G., et al. (2015). "HTAP_v2.2: a mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution." Atmospheric Chemistry and Physics15(19): 11411-11432.

Kanakidou, M., Seinfeld, J.H., Pandis, S.N., Barnes, I., Dentener, F.J., Facchini, M.C., Dingenen, R.V., Ervens, B., Nenes, A.N.C.J.S.E., Nielsen, C.J. and Swietlicki, E., 2005. Organic aerosol and global climate modelling: a review. Atmospheric Chemistry and Physics, 5(4), pp.1053-1123.

974 Kim, S. W., et al. (2009). "NO2columns in the western United States observed from space and
975 simulated by a regional chemistry model and their implications for NOxemissions." Journal
976 of Geophysical Research114(D11).

977 Kim, Y. J. S., S.N.; Carmichael, G.R.; Riemer, N.; Stanier, C.O. (2014). "Modeled aerosol
978 nitrate formation pathways during wintertime in the Great Lakes region of North America."
979 Journal of Geophysical Research: Atmospheres119: 12420-12445.

980 Kiehl, J.T., Briegleb, B.P., 1993. The relative roles of sulfate aerosols and greenhouse gases in
981 climate forcing. Science 260, 311-314.

982 Knote, C., et al. (2015). "Influence of the choice of gas-phase mechanism on predictions of key
983 gaseous pollutants during the AQMEII phase-2 intercomparison." Atmospheric
984 Environment115: 553-568.

985 Lack, D. A., et al. (2004). "Seasonal variability of secondary organic aerosol: A global modeling
986 study." Journal of Geophysical Research: Atmospheres109(D3): n/a-n/a.

987 Lelieveld, J., et al. (2015). "The contribution of outdoor air pollution sources to premature
988 mortality on a global scale." Nature525(7569): 367-371.

989 Li, M., et al. (2017). "MIX: a mosaic Asian anthropogenic emission inventory under the
990 international collaboration framework of the MICS-Asia and HTAP." Atmospheric
991 Chemistry and Physics17(2): 935-963.

992 Li, J. and Han, Z., 2016. Aerosol vertical distribution over east China from RIEMS-Chem
993 simulation in comparison with CALIPSO measurements. Atmospheric Environment, 143,
994 pp.177-189.

995 McKeen, S., Wilczak, J., Grell, G., Djalalova, I., Peckham, S., Hsie, E.Y., Gong, W., Bouchet,
996 V., Menard, S., Moffet, R. and McHenry, J., 2005. Assessment of an ensemble of seven
997 real‐time ozone forecasts over eastern North America during the summer of 2004. Journal
998 of Geophysical Research: Atmospheres, 110(D21).

Makar, P.A., Gong, W., Milbrandt, J., Hogrefe, C., Zhang, Y., Curci, G., Žabkar, R., Im, U., Balzarini, A., Baró, R. and Bianconi, R., 2015a. Feedbacks between air pollution and weather, Part 1: Effects on weather. Atmospheric Environment, (115), pp.442-469.

Makar, P.A., Gong, W., Hogrefe, C., Zhang, Y., Curci, G., Žabkar, R., Milbrandt, J., Im, U., Balzarini, A., Baró, R. and Bianconi, R., 2015b. Feedbacks between air pollution and weather, part 2: effects on chemistry. Atmospheric environment, 115, pp.499-526.

Menon, S. H., J.; Nazarenko, N.; Luo, Y. (2002). "Climate Effects of Black Carbon Aerosols in China and India." Science.

Nenes, A., Pandis, S.N. and Pilinis, C., 1998. ISORROPIA: A new thermodynamic equilibrium model for multiphase multicomponent inorganic aerosols. Aquatic geochemistry, 4(1), pp.123-152.

Nogherotto, R., Tompkins, A.M., Giuliani, G., Coppola, E. and Giorgi, F., 2016. Numerical framework and performance of the new multiple-phase cloud microphysics scheme in RegCM4. 5: precipitation, cloud microphysics, and cloud radiative effects. Geoscientific Model Development, 9(7), pp.2533-2547.

Odum, J.R., Huffmann, T., Bowman, F., Collins, D., Flagan, R.C., Seinfeld, J.H., 1996.Gas/Particle partitioning and secondary organic aerosol yields. Environ. Sci. Technol. 30, 2580-2585.

Peters-Lidard, C. D., E. M. Kemp, T. Matsui, J.A. Santanello Jr., S.V. Kumar, J.P. Jacob, T. Clune, W.-K. Tao, M. Chin, A. Hou, J.L. Case, D. Kim, K.-M. Kim, W. Lau, Y. Liu, J. Shi, D. Starr, Q. Tan, Z. Tao, B.F. Zaitchik, B. Zavodsky, S.Q. Zhang, and M. Zupanski, Integrated modeling of aerosol, cloud, precipitation and land processes at satellite-resolved scales. Environmental Modeling & Software, 67, 149-159, doi:10.1016/j.envsoft.2015.01.007, 2015.

Petters, M.D., Kreidenweis, S.M., 2007. A single parameter representation of hygroscopic growth and cloud condensation nucleus activity. Atmos. Chem. Phys. 7, 1961-1971.

Ramanathan, V. C., G. (2008). "Global and regional climate changes due to black carbon." Nature Geoscience1(4): 221-227.

San José, R., Pérez, J.L., Balzarini, A., Baró, R., Curci, G., Forkel, R., Galmarini, S., Grell, G., Hirtl, M., Honzak, L. and Im, U., 2015. Sensitivity of feedback effects in CBMZ/MOSAIC chemical mechanism. Atmospheric Environment, 115, pp.646-656.

Schuster, G. L., et al. (2006). "Angstrom exponent and bimodal aerosol size distributions." Journal of Geophysical Research111(D7).

Seinfeld, J.H. and Pandis, S.N., 2016. Atmospheric chemistry and physics: from air pollution to climate change. John Wiley & Sons.

Stockwell, W. R., et al. (1997). "A new mechanism for regional atmospheric chemistry modeling." Journal of Geophysical Research: Atmospheres102(D22): 25847-25879.

Stoiber, R. E. W., S.N.; Huebert, B. (1987). "Annual contribution of sulfur dioxide to the atmosphere by volcanoes." Journal of Volcanology and Geothermal Research33: 1-8.

Streets, D. G., et al. (2003). "Biomass burning in Asia: Annual and seasonal estimates and atmospheric emissions." Global Biogeochemical Cycles17(4): n/a-n/a.

Tao, Z., J. A. Santanello, M. Chin, S. Zhou, Q. Tan, E. M. Kemp, and C. D. Peters-Lidard, Effect of land cover on atmospheric processes and air quality over the continental United States – A NASA Unified WRF (NU-WRF) model study. Atmospheric Chemistry & Physics, 13: 6207-6226, doi: 10.5194/acp-13-6207-2013, 2013.

Tao, Z., H. Yu, and M. Chin, Impact of transpacific aerosol on air quality over the United States: A perspective from aerosol-cloud-radiation interactions. Atmospheric Environment, 125: 48-60, doi:10.1016/j.atmosenv.2015.10.083, 2016.

Tao, Z., H. Yu, and M. Chin, The role of aerosol-cloud-radiation interactions in regional air quality – A NU-WRF study over the United States. Atmosphere, 6, 1045-1068, doi:10.3390/atmos6081045, 2015.

1051 Tsimpidi et al., 2010. Evaluation of the volatility basis-set approach for the simulation of organic
1052 aerosol formation in the Mexico City metropolitan area. Atmos. Chem. Phys., 10, 525–546.

1053 Atmospheres102(D23): 28589-28612.

1054 Yu, S., Mathur, R., Pleim, J., Wong, D., Gilliam, R., Alapaty, K., Zhao, C. and Liu, X., 2013.
1055 Aerosol indirect effect on the grid-scale clouds in the two-way coupled WRF-CMAQ:
1056 model description, development, evaluation and regional analysis. Atmospheric Chemistry
1057 and Physics Discussion, p.25649.

1058 Wang et al., 2016. Persistent sulfate formation from London Fog to Chinese haze. PNAS,
1059 113(48), 13630–13635.

1060 Wang, J., et al. (2014). "Impact of aerosol–meteorology interactions on fine particle pollution
1061 during China's severe haze episode in January 2013." Environmental Research Letters9(9):
1062 094002.

1063 Wang, T., et al. (2010). "Investigations on direct and indirect effect of nitrate on temperature and
1064 precipitation in China using a regional climate chemistry modeling system." Journal of
1065 Geophysical Research115.

1066 Wang, Z. Maeda., T.; Hayashi, M.; Hsiao, L.F.; Liu, K.Y. (2001). "A nested air quality
1067 prediction modeling system for urban and regional scales: application for high-ozone
1068 episode in Taiwan." Water, Air, & Soil Pollution130(1): 391-396

1069 Xiao, H., et al. (1997). "Long-range transport of Sox and dust in East Asia during the PEM B
1070 Experiment." Journal of Geophysical Research:

1071 Xin, J., et al. (2015). "The Campaign on Atmospheric Aerosol Research Network of China:
1072 CARE-China." Bulletin of the American Meteorological Society96(7): 1137-1155.

1073 Zaveri, R. A., et al. (2008). "Model for Simulating Aerosol Interactions and Chemistry
1074 (MOSAIC)." Journal of Geophysical Research113(D13).

Zaveri, R. A. and L. K. Peters (1999). "A new lumped structure photochemical mechanism for large-scale applications." Journal of Geophysical Research: Atmospheres104(D23): 30387-30415.

Zhang, Y., et al. (2010). "Simulating chemistry–aerosol–cloud–radiation–climate feedbacks over the continental U.S. using the online-coupled Weather Research Forecasting Model with chemistry (WRF/Chem)." Atmospheric Environment44(29): 3568-3582.

Zheng et al., 2015. Heterogeneous chemistry: a mechanism missing in current models to explain secondary inorganic aerosol formation during the January 2013 haze episode in North China. Atmos. Chem. Phys., 15, 2031–204

Table 1 Participating models in Topic 3

Models	M1: WRF-Chem	M2: WRF-Chem	M3: NU-WRF	M4: NU-WRF	M5: RIEMS-Chem	M6: RegCCMS	M7: WRF-CMAQ
Modelling Group	Pusan National University	University of Iowa	USRA/NASA	USRA/NASA	Institute of Atmospheric Physics	Nanjing University	University of Tennessee
Grid Resolution	45km	50km	45km	15km	60km	50km	45km
Vertical Layers	40 layers to 50mb	27 layers to 50mb	60 layers to 20mb	60 layers to 20mb	16 layers to 100mb	18 layers to 50mb	
Gas phase chemistry	RACM-ESRL	CBMZ	RADM2	RADM2	CBM4	CBM4	SAPRC99
Aerosols	MADE/SORGAM; modal scheme	MOSAIC-8bin; sectional scheme	GOCART; bulk scheme	GOCART; bulk scheme	Sulfate, nitrate, ammonium, BC, OC, SOA, 5 bins of soil dust, and 5 bins of sea salt modal scheme;	Sulfate, nitrate, ammonium, BC and POC; bulk scheme	AE06 modal scheme
Chemical Boundary Conditions	Climatological data from NALROM	MOZART	MOZART GOCART	MOZART GOCART	GEOS-Chem	Climatological data	GEOS-Chem
Meteorological Boundary Conditions	NCEP FNL	NCEP FNL	NASA MERRA	NASA MERRA	NCEP FNL	NCEP-NCAR	NCEP FNL
BVOC emissions	prescribed	Internal calculation	Internal calculation	Internal calculation	prescribed	NA	Internal calculation
Dust	NA	GOCART AFWA	GOCART dust	GOCART dust	Han et al. (2004)	NA	NA

Microphysics	Lin scheme	Morrison double-moment	GCE (Goddard Cumulus Ensemble)	GCE	Reisner mixed phase	Nogherotto et al. (2016)	
Longwave radiation	RRTMG	RRTMG	Goddard	Goddard	CCM3	CCM3	RRTM
Shortwave radiation	RRTMG	RRTMG	Goddard	Goddard	Revised CCM3	CCM3	Goddard
Boundary Layer	Yonsei University	Yonsei University	YSU	YSU	MRF	Holtslag et al. (1990)	Yonsei University
Cu physics	Grell 3D	Grell 3D	Grell 3D	Grell 3D	Grell 3D	Anthes et al. (1997)	Grell 3D
Surface physics	Thermal diffusion	Unified Noah	Unified Noah	Unified Noah	BATS	BATs	
Aerosol-radiation	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Aerosol-microphysics	Yes	Yes	Yes	Yes	Yes	Yes	No
Mixing state	Internal mixing	Internal mixing	Internal mixing	Internal mixing	Internal mixing among inorganic aerosols and BC and OC, and external mixing between dust, sea-salt and other aerosols	External mixing	Internal mixing

NA represent not available; M1: WRF-Chem v3.7.1; M2: WRF-Chem v3.5.1; M3&M4: NU-WRF v7lis7-3.5.1-p3; M5: RIMES-Chem; M6: RegCCMS; M7: WRFv3.4.1&CMAQv5.0.2

Table 2 CARE-Chine network sites

ID	Site name	Characteristics	Longitude	Latitude
1	Beijing	AOD	116.37	39.97
2	Tianjin	Air quality*	117.21	39.08
3	Shijiazhuang	Air quality	114.53	38.03
4	Xianghe	Air quality	116.96	39.75
5	Xinglong	Air quality	117.58	40.39
6	Beijing Forest	AOD	115.43	39.97
7	Baoding	AOD	115.51	38.87
8	Cangzhou	AOD	116.80	38.28
9	Shenyang	AOD	123.63	41.52
10	Jiaozhou Bay	AOD	120.18	35.90

*Air quality: surface $PM_{2.5}$, PM_{10}, SO_2, NO_x, CO, O_3

Table 3 Performance Statistics of Meteorology Variables (RMSE and MBE units: degree for T2; g/kg for Q2; m/s for WS10; W/m^2 for SWDOWN)

Metrics	Models	T2	Q2	WS10	SWDOWN South	SWDOWN North
RMSE	M1	0.64	0.14	2.04	86.32	69.39
	M2	0.68	0.10	0.95	96.71	72.76
	M3	2.34	0.16	1.16	60.34	59.56
	M4	2.90	0.43	1.44	100.34	74.89
	M5	2.97	0.46	0.91	91.06	65.27
	M6	3.57	0.76	2.48	85.63	222.00
	M7	2.05	0.17	0.22	158.10	218.67
	Ensemble	1.81	0.10	1.28	81.96	62.51
MBE	M1	-0.19	0.02	2.01	66.58	59.94
	M2	-0.60	-0.01	0.91	83.88	62.38
	M3	-2.18	-0.04	1.11	36.44	47.74
	M4	-2.09	0.11	1.40	26.78	33.59
	M5	-2.73	0.43	0.74	49.06	51.00
	M6	-3.06	-0.56	2.37	-0.49	-202.26
	M7	-2.02	-0.12	0.15	145.24	159.02
	Ensemble	-1.71	-0.02	1.25	65.54	36.37
NMB (%)	M1	-0.07%	0.19%	17.58%	14.61%	13.34%
	M2	-0.21%	-0.12%	7.94%	18.41%	13.88%
	M3	-0.79%	-0.34%	9.73%	8.00%	10.63%
	M4	-0.75%	0.95%	12.26%	5.88%	7.48%
	M5	-0.98%	3.65%	6.45%	10.77%	11.35%
	M6	-1.10%	-4.77%	20.73%	-0.11%	-45.02%
	M7	-0.72%	-1.05%	1.31%	31.88%	35.39%
	Ensemble	-0.61%	-0.14%	10.98%	14.38%	8.10%

Table 4 Performance Statistics of Air Pollutants at the CARE-China sites (RMSE and MBE units: ppbv for gases and μg/m^3 for PM)

Metrics	Models	SO_2	NO_x	O_3	$PM_{2.5}$	PM_{10}		SO_2	NO_x	O_3	$PM_{2.5}$	PM_{10}
r	**M1**	0.76	0.60	0.46	0.85	0.76	**MBE**	-17.14	-5.53	-1.54	55.69	30.70
	M2	0.77	0.65	0.48	0.90	0.85		2.10	33.41	2.53	48.44	12.94
	M3	0.69	0.66	0.39	0.85	0.68		-15.89	-8.00	23.93	8.13	-19.92
	M4	0.67	0.61	0.42	0.88	0.73		-9.98	0.28	24.49	23.12	-3.23
	M5	0.72	0.73	0.39	0.91	0.84		-9.69	64.29	-5.30	1.68	-52.49
	M6	0.62	0.48	-	-	-		-27.53	-29.98	-	-	-
	M7	0.57	0.58	0.48	0.82	0.77		-25.56	7.85	-3.09	43.59	-21.00
	Ensemble	0.79	0.71	0.51	0.94	0.87		-14.81	8.90	6.84	30.11	-8.83
RMSE	**M1**	27.63	33.51	6.40	73.37	79.06	**NMB (%)**	-14.05	-5.41	7.37	63.57	18.93
	M2	21.00	66.30	8.15	72.44	80.72		12.13	69.58	39.87	54.07	6.38
	M3	29.50	36.87	24.76	47.20	78.21		-10.44	-6.26	306.33	9.67	-12.41
	M4	26.86	36.10	25.34	49.13	72.25		0.31	4.51	316.99	27.03	-1.78
	M5	32.17	87.48	7.90	45.32	81.00		6.83	127.45	-38.49	0.52	-32.94
	M6	33.95	48.62	-	-	-		-51.28	-48.59	-	-	-
	M7	34.75	35.88	6.89	64.25	70.19		-37.87	18.32	-7.78	48.92	-12.78
	Ensemble	24.10	29.12	8.86	45.25	56.65		-13.48	22.80	104.04	33.96	-5.77
MFB (%)	**M1**	-17.32	5.26	-5.06	64.34	21.98	**MFE (%)**	53.73	43.79	54.54	69.92	41.95
	M2	9.09	32.82	19.88	51.18	3.44		43.18	73.39	60.79	59.87	39.35
	M3	-12.96	4.52	113.60	32.67	-4.62		57.87	46.69	113.60	50.10	36.83
	M4	1.53	15.34	114.35	45.27	6.07		46.30	48.13	114.35	55.03	34.72
	M5	-20.24	67.25	-62.65	16.88	-35.15		63.69	72.07	80.92	48.17	45.09
	M6	-77.13	-56.89	-	-	-		84.21	69.66	-	-	-
	M7	-46.67	21.80	-19.50	57.19	-7.02		72.35	49.18	60.64	66.27	35.83
	Ensemble	-14.17	26.41	62.86	50.61	3.12		43.13	42.94	71.14	55.86	28.05

Table 5 Performance Statistics of Air Pollutants at the EANET sites (RMSE and MBE units: ppbv for gases and μg/m^3 for PM)

Metrics	Models	SO_2	NO_x	O_3	PM_{10}		SO_2	NO_x	O_3	PM_{10}
	M1	0.57	0.64	0.14	0.59		-0.68	0.68	-6.16	-21.03
	M2	0.59	0.45	0.30	0.75		-0.45	-0.39	5.50	3.12
	M3	0.50	0.55	0.26	0.51		-0.37	-0.21	3.67	3.55
	M4	0.45	0.55	0.25	0.49		-0.57	-0.61	4.28	2.96
	M5	0.58	0.54	0.01	0.03		-0.57	1.28	4.67	3.77
r	**M6**	0.33	0.24	-	-	**MBE**	0.32	-1.68	-	-
	M7	0.53	0.49	0.38	0.55		-0.03	0.64	-1.89	-15.75
	Ensemble	0.60	0.66	0.32	0.59		-0.34	-0.07	1.68	-3.89
	M1	-46.45	41.49	-15.03	-82.29		1.18	1.37	8.23	23.39
	M2	-29.64	-29.75	13.47	18.90		1.01	1.35	7.29	10.01
	M3	-25.42	-17.75	9.01	19.46		1.02	1.02	6.44	13.71
NMB	**M4**	-39.63	-35.84	10.47	16.95	**RMSE**	1.14	0.97	6.35	13.78
(%)	**M5**	-34.23	38.50	11.38	31.80		1.27	2.75	12.27	23.10
	M6	12.63	-93.57	-	-		1.38	1.85	-	-

M7	17.42	31.47	-4.71	-56.18		1.04	1.57	6.52	18.76
Ensemble	-20.76	-10.79	4.10	-8.56		0.96	0.79	4.98	11.69

Table 6 Performance Statistics of AOD

Metrics	Models	M1	M2	M3	M4	M5	M6	M7	Ensemble
R	North China	0.63	0.74	0.57	0.51	0.68	0.36	0.71	0.77
	All	0.60	0.65	0.46	0.42	0.53	0.33	0.64	0.75
MBE	North China	-0.25	-0.10	-0.09	-0.07	-0.13	-0.21	-0.05	-0.03
	All	-0.18	-0.02	-0.01	-0.01	-0.01	-0.11	0.00	-0.12
NMB (%)	North China	-71.25	-23.28	-12.63	-9.59	-28.34	-59.19	-2.70	-30.17
	All	-74.94	-30.69	-25.68	-23.64	-28.24	-55.38	-21.12	-28.91
RMSE	North China	0.35	0.20	0.26	0.28	0.24	0.36	0.22	0.22
	All	1.16	1.13	1.15	1.15	1.15	1.17	1.14	0.20

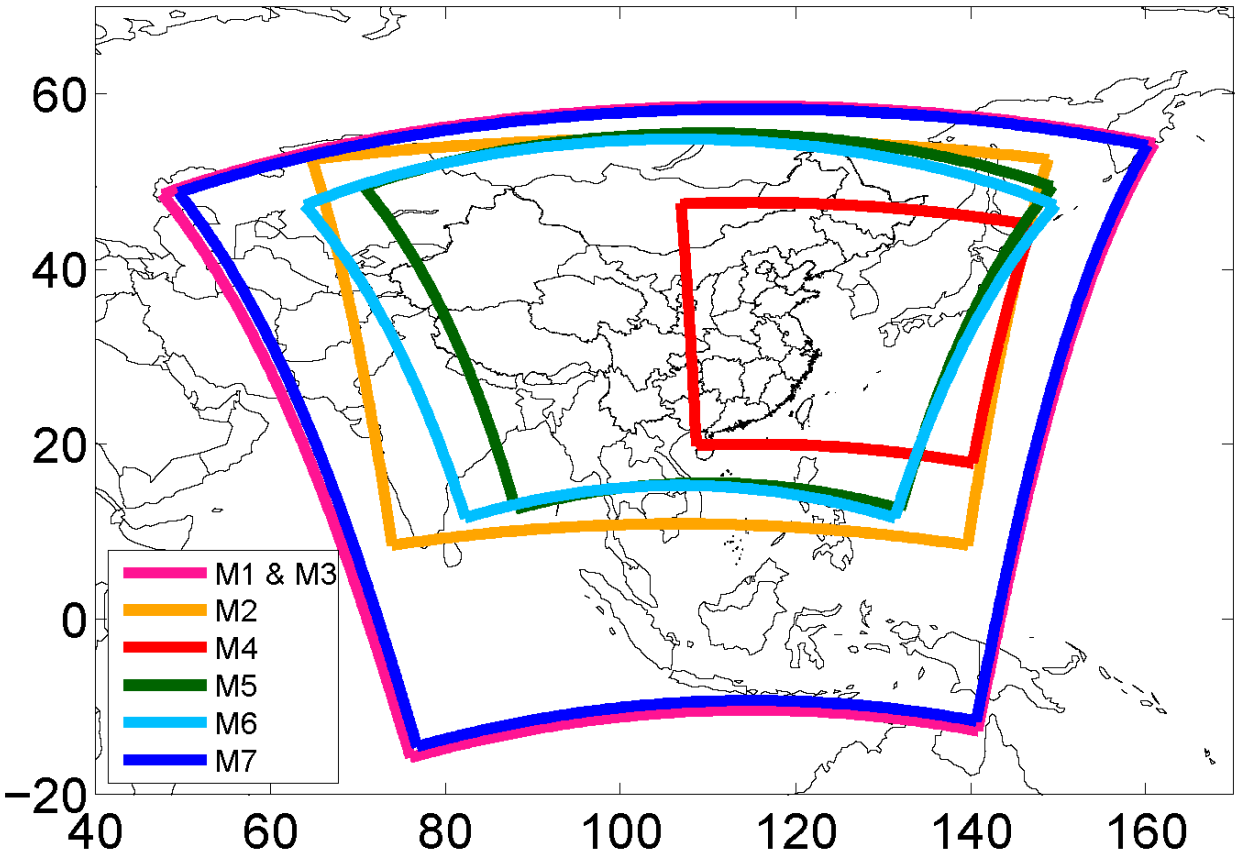

Figure 1. MICS-ASIA III Topic 3 modeling domains (descriptions of each model are documented in Table 1) M1: WRF-Chem 45km; M2: WRF-Chem 50km; M3: NU WRF 45km; M4: NU-WRF 15km; M5: RIEMS-IAP 60km; RegCCMS 50km; WRF-CMAQ 45km

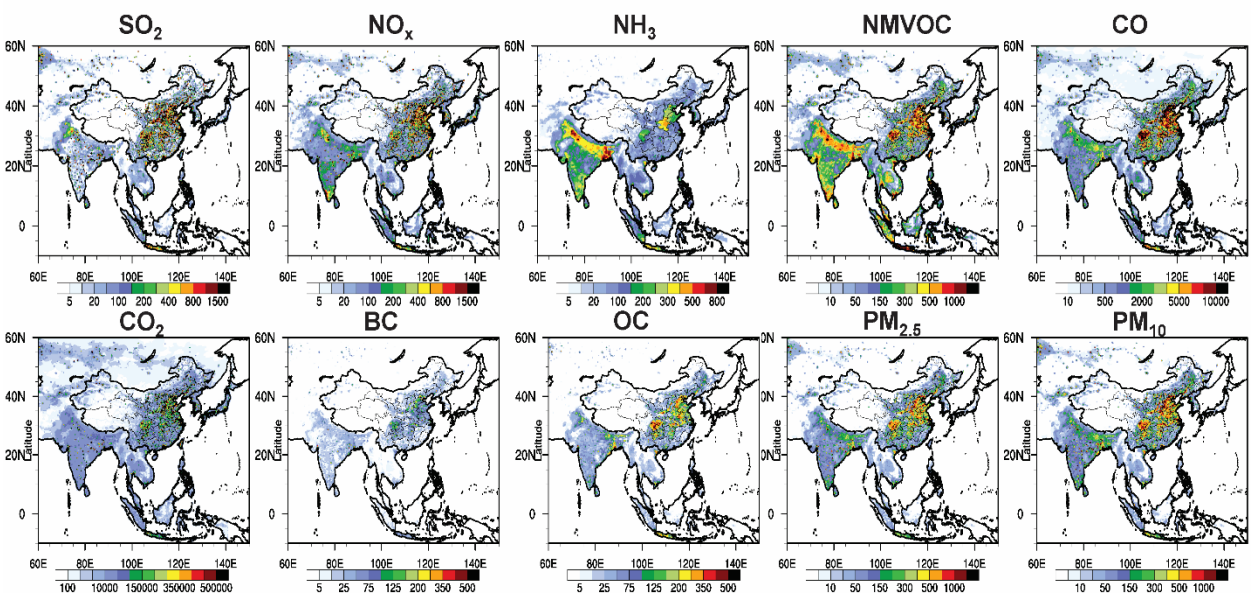

Figure 2. MIX emission inventory for January 2010 (Mg/month/grid)

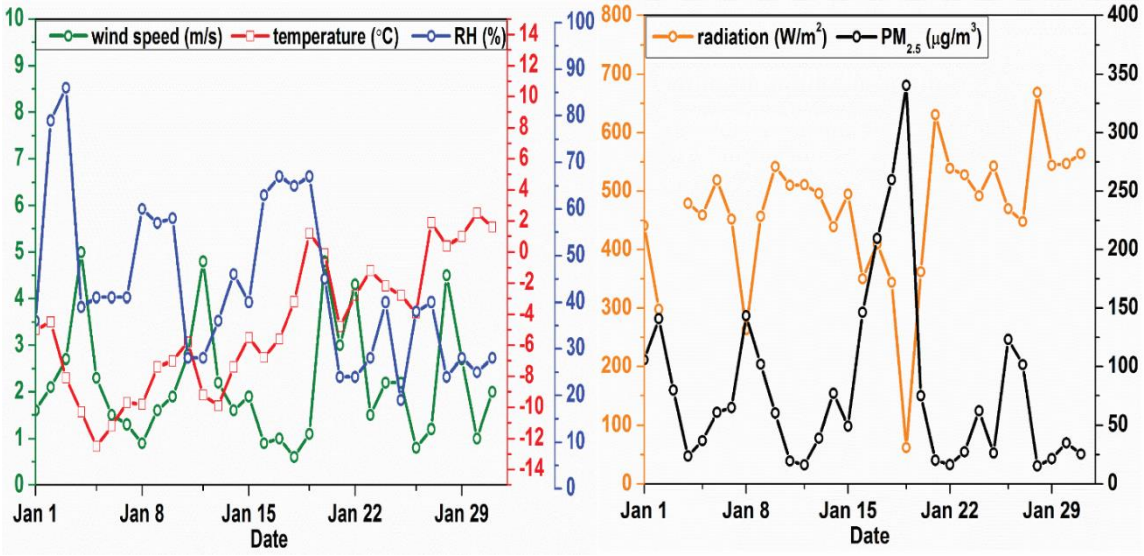

Figure 3. Observed near surface daily meteorological variables and $PM_{2.5}$ concentrations in Beijing for January 2010

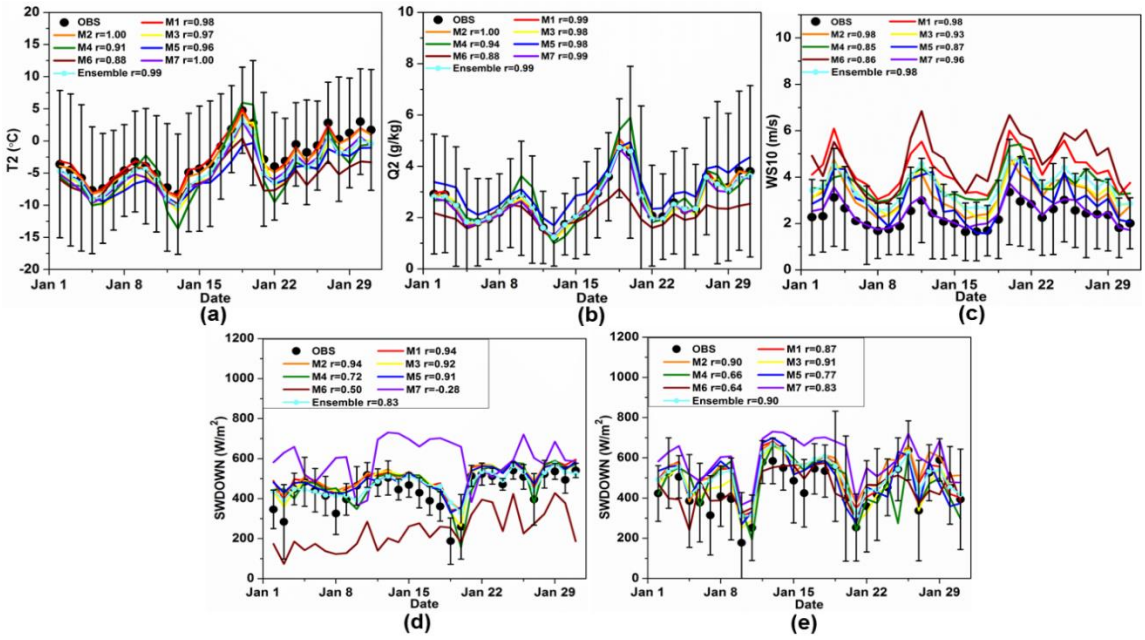

Figure 4. Comparisons between simulated and observed near surface temperature (a), water vapor mixing ratio (b), and wind speeds (c) (T2, Q2, and WS10), downward shortwave radiation in North China (d) and South China (e) (spatial daily values are averaged over measurements shown in S4 and S5; the error bars show the standard deviation of values over the measurement sites)

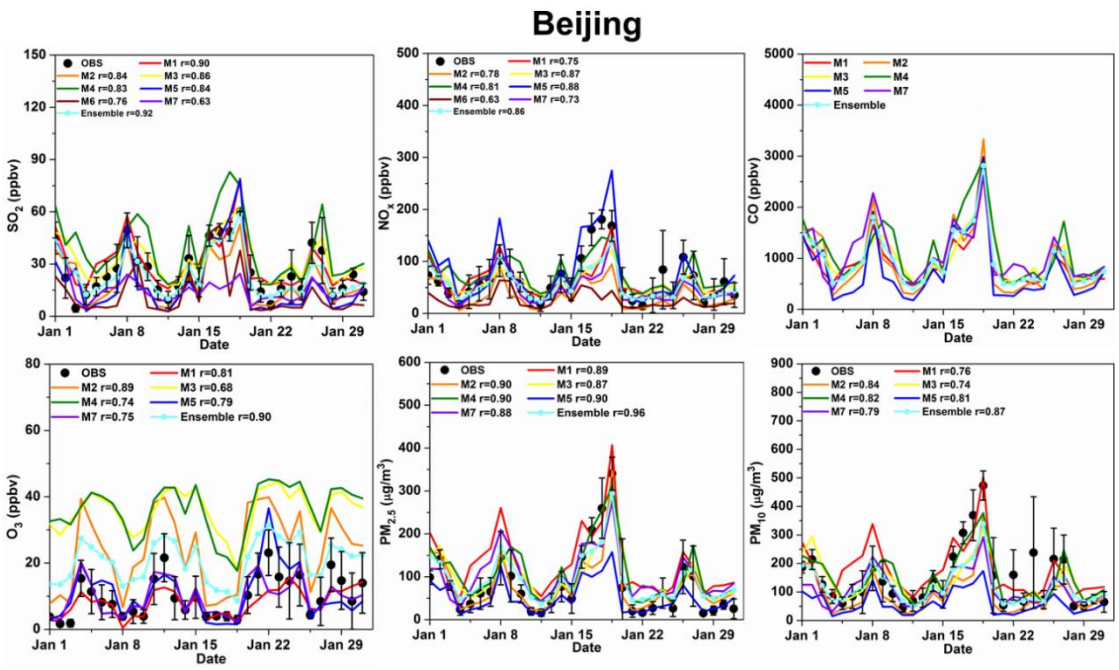

Figure 5. Comparisons between simulated and observed daily air pollutants (SO_2, NO_x, CO, O_3, $PM_{2.5}$ and PM_{10}) at the Beijing CARE-China site

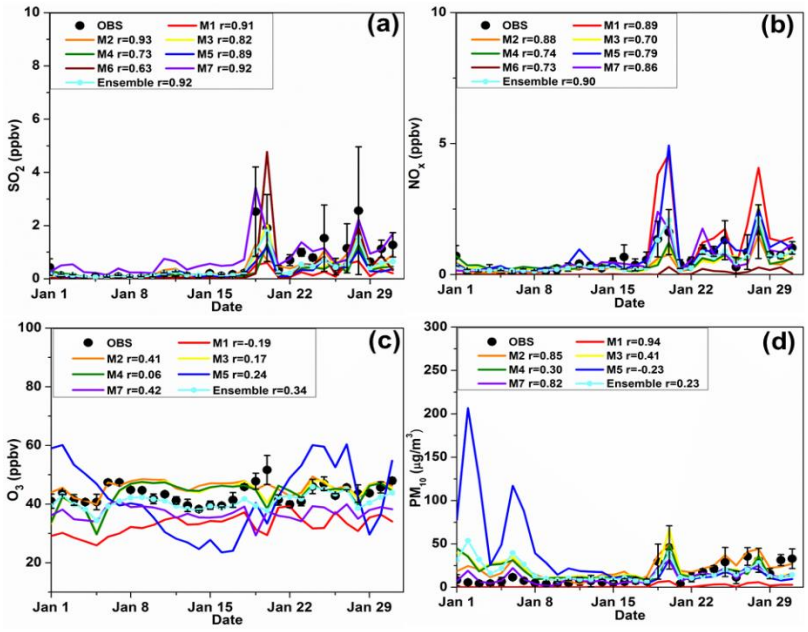

Figure 6. Comparisons between simulated and observed daily air pollutants (SO_2, NO_x, O_3, and PM_{10}) at the Rishiri EANET sites

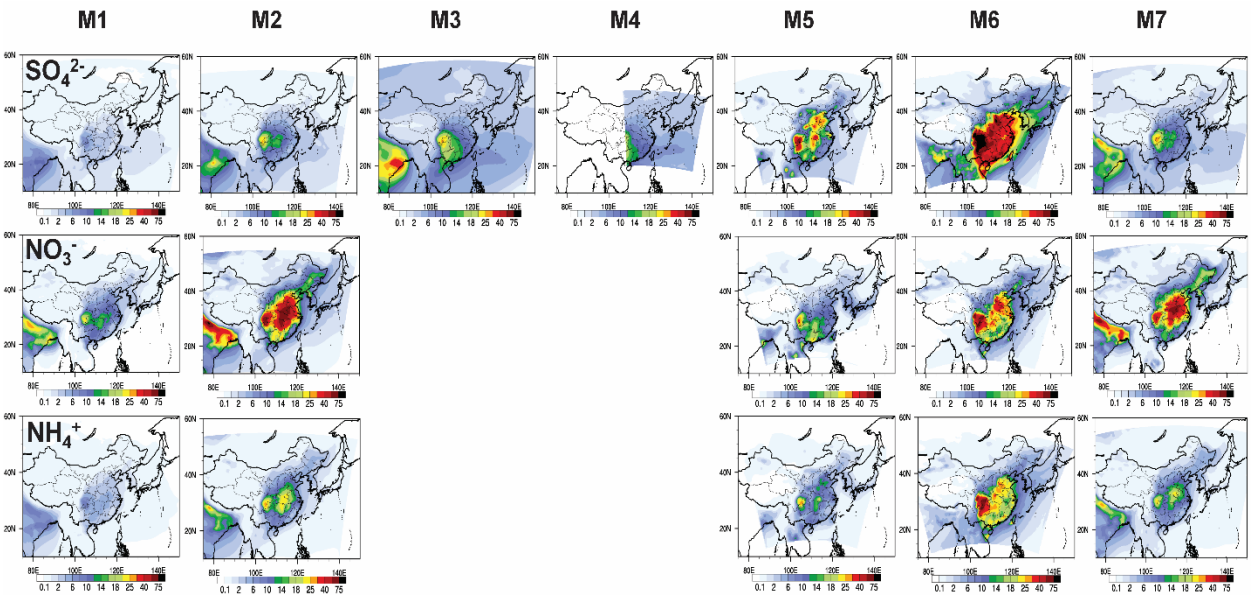

Figure 7. Simulated monthly concentrations of major $PM_{2.5}$ components (μg/m^3) for January 2010 from all participating models

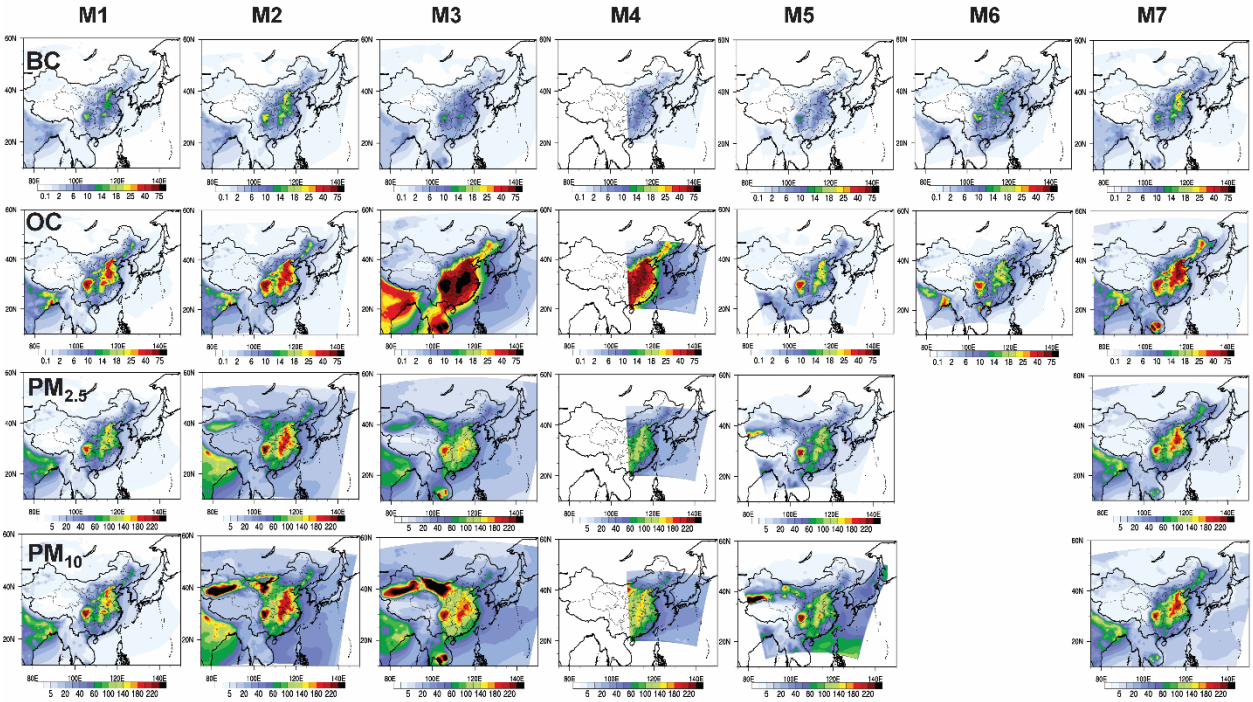

Figure 8. Simulated monthly concentrations of $PM_{2.5}$ and major $PM_{2.5}$ components ($\mu g/m^3$) for January 2010 from all participating models

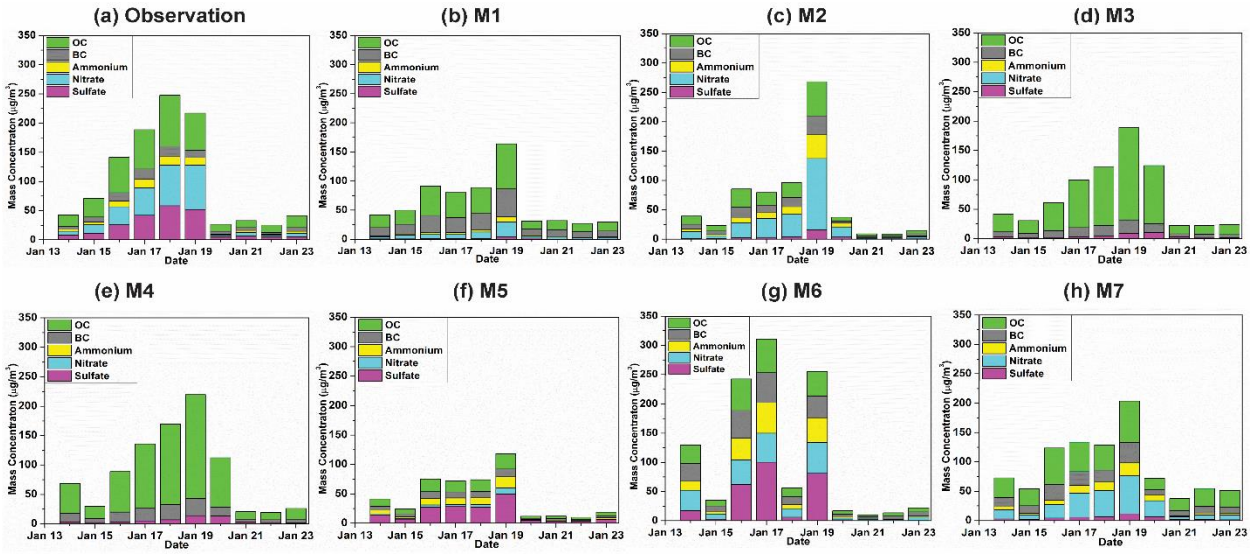

Figure 9. Observed and simulated daily mean concentrations of major $PM_{2.5}$ chemical components in the urban Beijing site

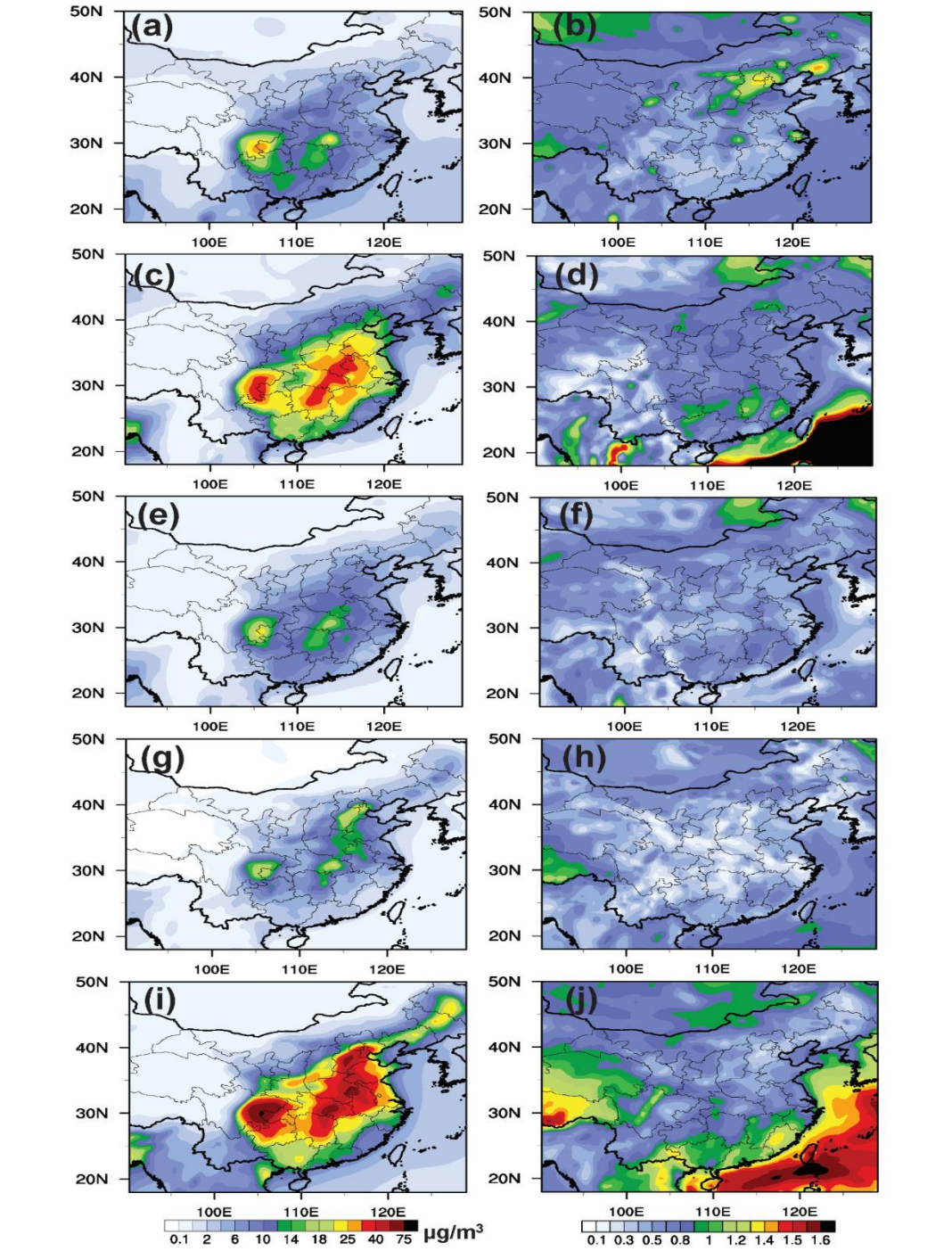

Figure 10. The ensemble mean monthly averaged near-surface distributions of $PM_{2.5}$ compositions for January 2010 (sulfate (a), nitrate (c), ammonium (e), BC (g), and OC (i)), along with the spatial distribution of the coefficient of variation ((b), (d), (f), (h), and (j), standard deviation divided by the average)

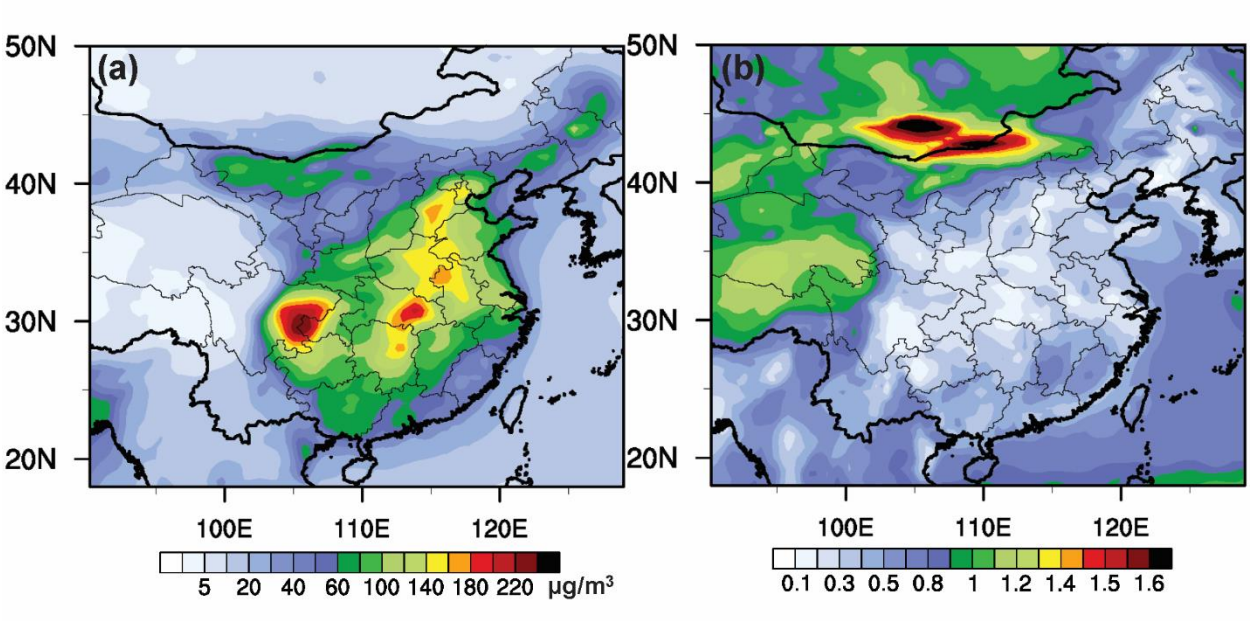

Figure 11. The ensemble mean monthly averaged near-surface distributions of $PM_{2.5}$ for January 2010 (a), along with the spatial distribution of the coefficient of variation (b, standard deviation divided by the average)

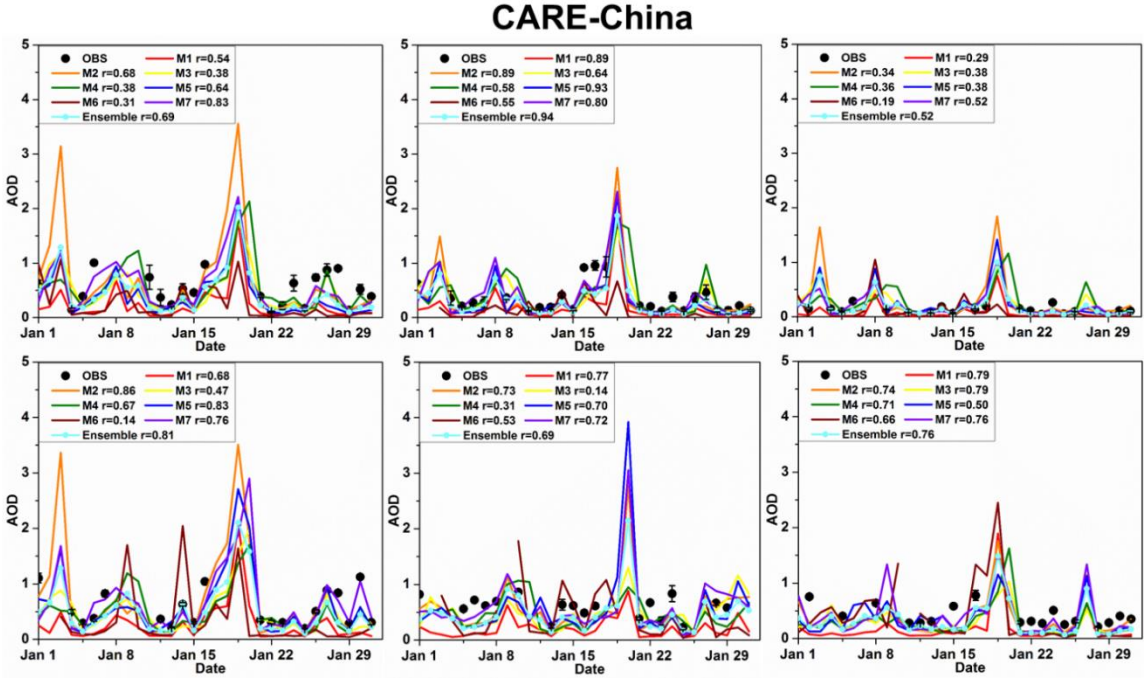

Figure 12. Comparisons between simulated and observed daily (daytime) mean AOD at the CARE-China sites (Baoding, Beijing City, Beijing Forest, Cangzhou, Jiaozhou, Shenyang,)

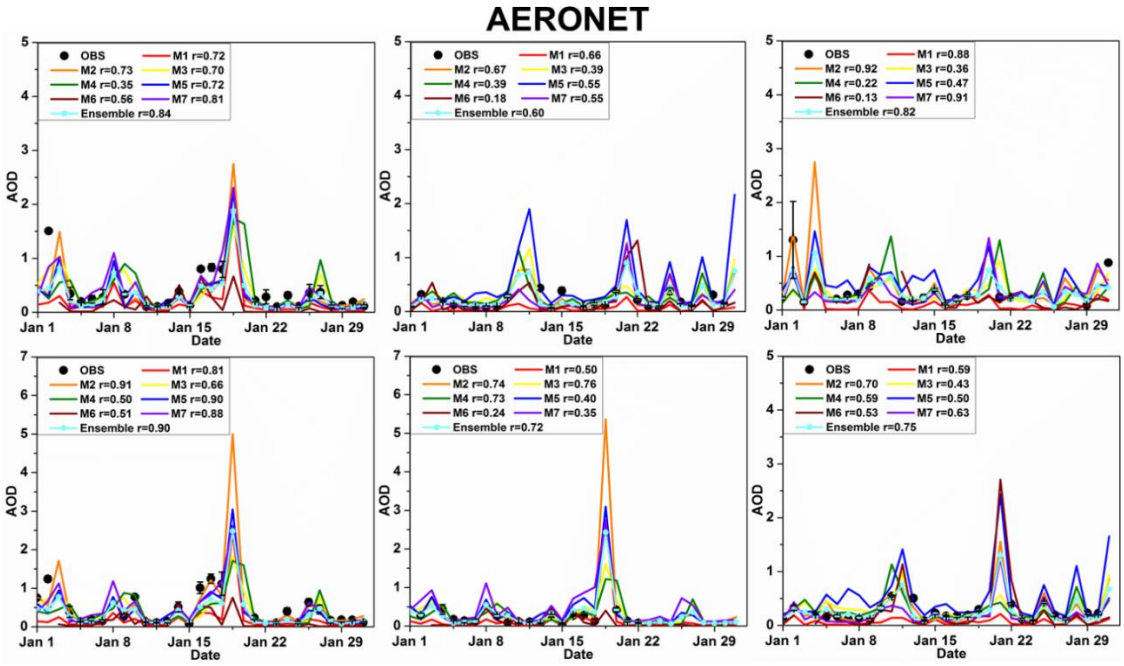

Figure 13. Comparisons between simulated and observed daily (daytime) mean AOD at the AERONET sites (Beijing, Shirahama, GIST, Xianghe, Xinglong, Osaka)