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, 7], Jinyuan Xin[5], Zhining Tao[8,9], Jiawei Li[3,4], Jeong-Eon Kang[10], Kan Huang[11], Xinyi Dong[11], Bingliang Zhuang[12], Shu Li[12], Baozhu Ge[5], Qizhong Wu[13], Yafang Cheng[7], Yuesi Wang[5], Hyo-Jung Lee[10], Cheol-Hee Kim[10], Joshua S. Fu[11], Tijian Wang[12], Mian Chin[9], 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 Multiphase Chemistry Department, Max Planck Institute for Chemistry, Mainz, Germany

8 Universities Space Research Association, Columbia, MD, USA

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

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

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

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

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

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 data from 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. 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 aerosols 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 differences in composition and AOD suggest that future models can be improved by including 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 more detailed evaluation of the dependence of aerosol optical properties on size distribution and mixing state. 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.

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

In 2010, MICS-Asia phase III was launched and three topics for this phase were decided during the first and second Workshop on Atmospheric Modeling in East Asia. Phase III aims to evaluate strengths and weaknesses of current air quality models and provide techniques to reduce uncertainty in Asia (Topic 1), to develop a reliable anthropogenic emission inventory in Asia (Topic 2), and to evaluate aerosol-weather-climate interactions (Topic 3). Various multi-scale models participated in this phase, and the study periods range from year to month depending on study topics. This phase benefits from the Acid Deposition Monitoring Network in East Asia (EANET) measurements, in addition to new observations related to atmospheric chemistry in this region. A detailed overview of the MICS-Asia Phase III, including the descriptions of different research topics and participating models, will be published in a companion paper. An important advance to this phase is the inclusion of multiple online-coupled chemistry-meteorology models to investigate aerosol-weather-climate interactions, which is the target of

topic 3. Online coupled models are playing important roles in air quality, meteorology and climate applications, but many important 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 were studied (Forkel et al., 2015; Makar et al., 2015a, 2015b; San Jose et al., 2015) over Europe and North America.

This paper overviews 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 the participating models, how the experiments are designed, and the 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. 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 comparisons and discussions focused on the results related to the meteorological and air pollution conditions during the January 2010 heavy haze episode.

2 Inter-comparison framework

In North China, severe aerosol pollution frequently happens and attracts enormous interests from both the public and the scientific community (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 on 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 simulate 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) and the observational datasets, and present 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 enable 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 northern region 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, Stockwell et al., 1997, Earth System Research Laboratory, Kim et al., 2009) gas phase chemistry was used. RACM was developed based on the Regional Acid Deposition Model (RADM2) to simulate regional atmospheric chemistry (Stockwell et al., 1990) (including 237 reactions) and the rate coefficients were updated in the 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) used 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 the SAPRC99 mechanism 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, Ackermann et al., 1998) aerosol module coupled by Schell et al. (2001) was 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 used an 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, BC, 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 were 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) was 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 aerosols are 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, Carlton et al., 2010) mechanism was

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 adds 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 used the National Centers for Environmental Prediction (NCEP) final analysis (FNL) data to drive the model; M3 and M4 used NASA MERRA reanalysis data and M6 used NCEP-NCAR reanalysis 1 dataset (Kalnay et al., 1996).

(5) Soil dust: M1, M6 and M7 did not include soil dust calculation. M3 and M4 used GOCART dust module (Ginoux et al., 2001), and M2 used a GOCART version that modified by AFWA (Air Force Weather Agency). M5 used 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 3.

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 transport, volcano emissions, and dust emissions were used. This section provides some basic descriptions of these 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 were provided with CB-05 and SAPRC-99 speciation datasets. Speciation mapping of NMVOC emissions for groups using other gas-phase chemical mechanisms, such as CBMZ, RADM2 and CBM4, used the speciation framework documented in Li et al. (2014). Emissions of these species were prepared for years 2008 and 2010 with a 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 conditions, which include solar transmission, air temperature, humidity, wind speed, and soil moisture. In the preparation of MEGAN biogenic emissions, land cover information was taken from the NASA MODIS products, and weather conditions were calculated using the 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 used prescribed biogenic VOCs emissions, other models except M6 used internal calculation.

2.2.3 Biomass burning emissions

Biomass burning 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, Randerson et al., 2015) (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 with monthly temporal resolution. Figure S2 shows the gridded biomass burning emissions for January 2010. Biomass burning activity was 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 the 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% to global NO_x emissions (Janssens-Maenhout et al., 2015).

2.2.6 Dust emissions

In M2, the Air Force Weather Agency (AFWA) version of the GOCART dust model was used. It calculates the saltation flux as a function of friction velocity (u_*) and threshold friction velocity (u_{*t}):

$$Q = C\frac{\rho_0}{g}u_*^3\,(1+\frac{u_{*t}}{u_*})(1-\frac{u_{*t}^2}{u_*^2}) \text{ when } u_* \geq u_{*t}$$

where C is a tunable empirical constant, ρ_0 is air density, and g is gravitational acceleration. The bulk vertical dust flux is estimated by $F = \alpha QE$ (Marticorena and Bergametti, 1995), in which α is the sandblasting efficiency and E is the dust erodibility factor. The erodibility factor data is included in the model geography dataset.

In M3 and M4, the dust emissions are estimated using the GOCART dust model (Ginoux et al., 2001), which determined by soil texture, moisture and surface wind speed. The drier the soil and the stronger the wind, the higher dust emissions over the regions where the erodibility factor is not zero. In M5, soil dust emissions were estimated by the approach from Han et al. (2004):

$$F = C_0 u_*^4 \left(1 - \frac{u_{*t}}{u_*}\right)(1 - f_i R_i) \text{ when } u_* \geq u_{*t} \text{ , } RH \leq RH_t$$

C_0 is a constant (1.4×10^{-15}), R_i is the reduction factor and f_i is the factional coverage of i type of vegetation in a model grid (considering that vegetation cover can reduce dust emissions). u_* and u_{*t} are the friction and threshold friction velocities. RH and RH_t are the relative humidity and threshold relative humidity near the surface. The total dust emission flux is apportioned to each size bin based on field measurements of vertical dust flux size distribution s in Chinese deserts.

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 if the same global chemistry 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 an obstacle in understanding air quality and constraining air quality modeling in Asia. Beginning with MICS-ASIA II, observational data from the 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 were 54 wet deposition sites and 46 dry deposition sites in the 13 participating countries. Quality assurance and quality control measures were 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 the EANET dataset can be found in http://www.eanet.asia/.

In addition to the 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 the Aerosol Robotic Network (AERONET) (https://aeronet.gsfc.nasa.gov/) and the 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 in pairs. 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 calculated 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 2 meters above ground (T2, Q2), wind speed at 10 meters 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, O_3, $PM_{2.5}$, 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 aerosol optical depth (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 for these metrics are presented in supplemental information.

3 Results and discussions

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 displays the monthly mean T2 (temperature at 2 meters), WS10 (wind speeds at 10 meters) and total precipitation. Near surface wind speeds were very weak in eastern and central China regions, and there was no significant precipitation in North China (Figure S8). During winters, northern China

burns coal for heating, generating more emissions of air pollutants. 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.

In this section, we present some major features of model performances in meteorological and chemical variables for the January 2010 period. Detailed analyses of aerosol 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.

3.1 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 2 meters (T2), water vapor mixing ratio at 2 meters (Q2), wind speed at 10 meters (WS10) and daily maximum downward shortwave radiation (SWDOWN) were evaluated against near surface observations at the CMA sites.

458 Figure 4 (a-c) shows the comparisons between simulated and observed daily mean T2, Q2 and
459 WS10 averaged over stations in East China (locations are shown in Figure S5) during January
460 2010, along with multi-model ensemble mean and observational standard deviation. The
461 calculated correlation coefficients between models and observations are also shown in Figure 4
462 and other calculated model evaluation metrics are summarized in Table 3. In general, the
463 simulated magnitudes and temporal variations of T2 and Q2 show high order of consistencies
464 with observations, with correlation coefficients ranging from 0.88 to 1. For T2, models tend to
465 have a cool bias; M1 and M2 have the lowest RMSE (0.64 and 0.68), lowest MBE (-0.19 and -
466 0.60) and lowest NMB (-0.07% and -0.22%) values (Table 3). For Q2, most models tend to
467 slightly overestimate; M1 and M2 have the best performance, with the lowest RMSE (0.14 and
468 0.10), lowest MBE (0.02 and -0.01), and lowest NMB (0.84% and -0.55%) values (Table 3).

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

479 The accuracy of radiation predictions is of great significance in evaluating aerosol-radiation-
480 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). The slightly higher daily maximum SWDOWN from M7 than other models is due to the deactivation of aerosol-radiation interactions in the presented M7 simulation.

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. 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. Besides, clouds are also important to alter radiation. To exclude clouds' impacts on the radiation shown here, we calculated the radiation reduction ratio due to clouds using radiation prediction for clear sky and for cloudy conditions 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. Thus, the relatively lower radiation during this period (Figure 4(d)) is mainly caused by aerosols, but 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.

503 The above comparisons show that T2 and Q2 were reproduced well by the participating models,
504 but wind speeds were overestimated by all models. Emery et al. (2001) proposed that excellent
505 model performance would be classified as wind speed RMSE smaller than 2 m/s, and wind speed
506 bias smaller than 0.5 m/s. Based on the calculated RMSE and MBE of WS10 shown in Table 3,
507 RMSE values from all models match the proposed RMSE threshold but MBE values are higher
508 than 0.5 m/s. The vertical distributions of temperature, water vapor mixing ratio and wind speeds
509 were also validated against atmospheric sounding data in Beijing at 1km and 3km (Figure S11,
510 averaged at 00:00 and 12:00 UTC) (http://weather.uwyo.edu/upperair/sounding.html). The
511 magnitudes of temperature, water vapor mixing ratio and wind speeds from different models are
512 generally consistent with each other at 1km and 3km, but variations are larger near the surface.

513 **3.2 Evaluation of air pollutants**

514 Figure 5 displays the daily averaged predicted and observed SO_2, NO_x, CO, O_3, $PM_{2.5}$, and PM_{10}
515 concentrations at the Beijing station, along with the observational standard deviation (locations
516 are shown in Figure S7). Comparisons for the Tianjin, Shijiazhuang and Xianghe sites are shown
517 in Figure S12-S14. M6 only provided SO_2, NO_x concentrations, so it is not shown in the plots of
518 CO, O_3, $PM_{2.5}$, and PM_{10}. The observed and predicted primary gaseous pollutants, $PM_{2.5}$ and
519 PM_{10} show the same monthly variations with elevated values at roughly weekly intervals, with
520 the largest event occurring during January 15-21. For example, as shown in the comparisons of
521 SO_2 concentration, the temporal variations are reproduced well by all the models, but peak
522 values are overestimated or underestimated by some models. Based on the calculated MBE
523 values shown in Table 4, all models except M2 tend to underestimate SO_2 in the CARE-China
524 sites. M1 shows the highest correlation (0.90) with SO_2 observations in the Beijing site, and most
525 other models show similar good correlations. The multi-model ensemble mean shows a better

agreement with observations with a higher correlation of 0.92, and it falls within the range shown with standard deviation error bar. In general, the predictions for NO_x capture the main features in the observations, with slightly less skills than for the SO_2 prediction. The calculated correlation coefficients for NO_x from different models are close to each other, ranging from 0.63 to 0.88. M2 and M5 predict higher NO_x concentrations than observations and other models (MBE in Table 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. In winter of the North China Plain, soil dust generally contributes about 10% to $PM_{2.5}$ concentrations (He et al., 2014), but there is also primary PM from anthropogenic activity, such as power plant, traffic, and construction etc. The primary particles are mostly in coarse mode, which might contribute to PM_{10} concentrations, but are highly uncertain compared with other anthropogenic emission sectors.

The models showed the poorest skills 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

549 overestimate ozone concentrations, M2 slightly overestimates it, and M1, M5, and M7 slightly
550 underestimate it (MBE in Table 4). According to the calculated RMSE in Table 4, M1 and M7
551 show relatively better performance in modeling ozone variations. Although WRF-Chem and NU-
552 WRF models were applied at three institutions, different gas phase chemistry schemes were
553 used, which leads to these diversities among predicted ozone concentrations. The impacts of gas
554 phase chemical mechanisms on ozone simulations have been investigated in Knote et al. (2015).
555 The overestimations of ozone concentrations from M3 and M4 primarily occur during nighttime,
556 implying the underestimated titration of ozone by NO_x. Forkel et al. (2015) reported that the
557 RADM2 solver in WRF-Chem has the problem of underestimating ozone titration in areas with
558 high NO emissions, which is the version that applied in M3 and M4.

559 Figure 6 shows the comparisons between modeled and observed ground level daily averaged
560 concentrations of SO_2, NO_x, O_3 and PM_{10} during January 2010 at the Rishiri site in Japan from
561 EANET. The locations of EANET sites are marked in Figure S7. Comparisons at other EANET
562 sites are shown in Figure S15-S18. The models are able to predict the major features in the
563 observations. For example, low values of most pollutants are observed (and predicted) during the
564 first half of the month, followed by elevated values, which peak on January 21. For SO_2, most
565 models show similar capability in producing the temporal variations in observations with slight
566 underestimation (MBE in Table 5). According to the calculated RMSE averaged over all the
567 EANET sites, M2 and the multi-model ensemble mean performed the best. For NO_x, the multi-
568 model ensemble mean shows lower RMSE than any individual model (Table 5). Similar to the
569 comparisons over CARE-China sites, large discrepancies exist in ozone predictions, but the
570 model ensemble mean still shows the lowest RMSE for ozone predictions. PM_{10} concentrations
571 are largely underestimated by M1 (largest negative MBE: -21.03ug/m^3) and overestimated by

572 M5 (highest positive MBE: 3.77ug/m^3) (Table 5), which could be related to the differences in
573 sea-salt treatments. Spatial distributions of the monthly near surface concentrations of SO_2, NO_x,
574 O_3 and CO for January 2010 from all participating models are shown in Figure S19. The aerosol
575 spatial distributions are discussed in the following section.

576

577 **3.3 $PM_{2.5}$ and $PM_{2.5}$ chemical composition distribution**

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

582 M1, M2, M3, M4 and M7 all predict quite low sulfate concentrations in east China, but with
583 considerably enhanced sulfate in southwestern China and western India. M5 and M6 show similar
584 spatial patterns of sulfate except that M6 produces higher concentrations. The chemical production
585 of sulfate is mainly from gas-phase oxidation of SO_2 by OH radicals and aqueous-phase pathways
586 in cloud water. In cloud water, dissolved SO_2 can be oxidized by O_3, H_2O_2, Fe(III), Mn(II), and
587 NO_2 (Seinfeld and Pandis, 2016). Most chemical transport models have included the above gas
588 phase oxidation of SO_2 by OH, and oxidation of dissolved SO_2 by O_3 and H_2O_2 in aqueous phase.
589 Under hazy conditions, radiation is largely reduced due to aerosol dimming effects, and sulfate
590 formation from gas phase and aqueous phase oxidation processes are slowed down, which tend to
591 reduce sulfate concentration. However, field observations exhibit an increase in sulfate
592 concentration during haze episode (Zheng et al., 2015). Cheng et al. (2016) proposed that the
593 reactive nitrogen chemistry in aerosol water could contribute significantly to the sulfate increase

due to enhanced sulfate production rates of NO_2 reaction pathway under high aerosol pH and elevated NO_2 concentrations in the North China Plain (NCP). Wang et al. (2016) also pointed out the aqueous oxidation of SO_2 by NO_2 is key to efficient sulfate formation on fine aerosols with high relative humidity and NH_3 neutralization, or under cloudy conditions. Besides, Zheng et al. (2015) suggested that heterogeneous chemistry on primary aerosols could play an important role in sulfate production and lead to increasing sulfate simulation during haze episodes. X. Huang et al. (2014) found including natural and anthropogenic mineral aerosols can enhance sulfate production through aqueous-phase oxidation of dissolved SO_2 by O_3, NO_2, H_2O_2 and transition metal. Gao et al. (2016b), Wang et al. (2014), and Zhang et al. (2015) also emphasized the importance of multiphase oxidation in winter sulfate production. However, these processes are currently not incorporated in the participating models for this study, which might be responsible for the apparent under-predictions of sulfate concentrations (Figure 9). M5 incorporated heterogeneous chemical reactions on aerosol surface (Li and Han, 2010), which enhances total sulfate production.

M1 and M5 predict relatively small nitrate and ammonium concentrations; while M2, M6 and M7 produce similar magnitudes and spatial patterns of nitrate. Nitrate formation involves both daytime and nighttime chemistry. During daytime, NO_2 can be oxidized by OH to form nitric acid (HNO_3), and by ozone to form NO_3. HNO_3 is easily removed by dry or wet deposition, but NO_3 is easily photolyzed back to NO_2. During nighttime, NO_3 is the major oxidant, which oxides NO_2 to form dinitrogen pentoxide (N_2O_5). Homogenous reaction of N_2O_5 with water vapor is possible but very slow while heterogeneous uptake of N_2O_5 onto aerosol particles has been identified as a major sink of N_2O_5 and an important contributor to particulate nitrate (Kim et al., 2014). The MOSAIC aerosol module (Zaveri et al., 2008) coupled with CBMZ gas phase

chemistry in WRF-Chem already includes heterogeneous uptake of N_2O_5 since version v3.5.1 (Archer-Nicholls et al., 2014), which is the version used by M2, leading to the high production of nitrate. An et al. (2013) incorporated photoexcited nitrogen dioxide molecules, heterogeneous reactions on aerosol surfaces, and direct nitrous acid (HONO) emissions into the WRF-Chem model and found these additional HONO sources could improve simulations of HONO and nitrate in north China. M7 also predicts high nitrate concentrations (N_2O_5 and NO_2 gases react with liquid water, Zheng et al., 2015), and the predicted lower nitrate concentrations from other models are probably due to missing aqueous phase and heterogeneous chemistry, or the implementations of different gas phase oxidation in these models. Many studies have been conducted regarding sulfate formation issues. Nitrate also account for a large mass fraction in $PM_{2.5}$ during winter haze events in north China, yet less attention was attracted to fully understand its formation. It is worth furtherly digging into the details about how different processes contribute to high nitrate concentrations in future studies. M3 and M4 do not include the explicit nitrate and ammonium treatment but ammonium is implicitly considered in total $PM_{2.5}$ mass estimate.

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

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

The disparity among predicted OC concentrations is mainly associated with the different treatments of SOA production, given the POC prediction is generally consistent among models using the same emission inventory. The predicted OC concentrations from M1, M2, and M7 are close to each other. M1 uses SORGAM (Secondary Organic Aerosol Model) to simulate SOA, but M2 and M6 did not include any SOA formation mechanism. The similar magnitudes of OC from M1 suggest 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 was 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, which are the products of the 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 (VBS) approach (Donahue et al., 2006) was developed to represent the wide range of volatility of organic compounds and complex

processes. It was found that the VBS approach was able to increase SOA production, and able 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 (Kanakidou et al., 2005). In M5, the SOA production is calculated using a bulk yield method (Lack et al., 2004), which uses yields that represent the maximum amount of SOA able to be produced from a unit of reacted VOCs. However, the SOA concentration is highly dependent on the yields 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.

However, most models over-simplified SOA formation. In M3 and M4, SOA was treated by assuming that 10% of VOCs from terrestrial source are converted to OC (Chin et al., 2002), and these models produced 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-

686 blown dust. M1 and M7 failed to produce high $PM_{2.5}$ concentrations in the deserts of west China,
687 due to omission of dust emissions. The spatial distributions of predicted wind-blown dust from
688 M5 are slightly different from M2 and M3, with lower concentrations over the Gobi desert (in
689 west Inner Mongolia) (PM_{10} in Figure 8). M2 and M3 used similar GOCART dust emission
690 schemes based on wind speeds and erodible areas, while M5 furtherly considered the dust
691 reduction by vegetation cover, which could partially explain the relatively lower wind-blown dust
692 predictions from M5. The enhanced $PM_{2.5}$ concentrations in Central China from M2 and M7 are
693 caused by large nitrate production, as shown in Figure 7.

694 The differences in the predictions of aerosols composition discussed above can be seen clearly in
695 the comparisons at the Beijing site during the 13-23 January period when a haze event occurred
696 in the NCP (Figure 9). Most models failed to produce the observed high sulfate concentrations.
697 Only the sulfate predictions from M5 are close to the observed high values. M2 and M7 predict
698 reasonable nitrate concentrations. M3 and M4 overpredict OC during the haze period, but other
699 models underpredict OC concentrations.

700 Figure 10 and 11 show the ensemble mean monthly averaged near-surface $PM_{2.5}$, $PM_{2.5}$
701 composition, along with the spatial distribution of the coefficient of variation. The coefficient of
702 variation (CV) is defined as the standard deviation divided by the average (Carmichael et al.,
703 2008b), and larger values indicate lower consistency among models. The mean concentrations of
704 $PM_{2.5}$ and the mean concentrations of $PM_{2.5}$ chemical compositions are high in Sichuan Basin and
705 east China. High CV values are shown in North China for sulfate, and in most areas for nitrate and
706 OC. The diversity in predictions of these species is caused by complexity of secondary formation
707 and different model treatments, which have been discussed above. Higher consistency is shown
708 for model BC with CV values less than 0.3 in most areas (Figure 10(h)). The CV values for $PM_{2.5}$

are also low in the North China region, which is consistent with the good performance of $PM_{2.5}$ predictions shown in above comparisons. However, the CV values can reach above 1.6 in northwestern China regions, partially due to discrepancies in dust predictions.

3.4 Evaluation of AOD

We used the AOD measurements from the AERONET and CARE-China networks to evaluate how participating models perform in simulating AOD. The submitted AOD data from all models except M6 were at 550nm, and AOD predictions from M6 were at 495nm. We used the Angstrom exponent relation (Schuster et al., 2006) to convert AOD at 495nm to 550nm, and all the used AERONET and CARE-China AOD data to 550nm. The locations of the AERONET and CARE-China AOD measurement sites are marked in Figure S6. 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) during the severe haze episode of 15-20 January 2010 (Figure 12 and Figure 13). M2 consistently simulated the highest AOD values among models, followed by M5 and M7, while M6 simulated the lowest.

In M1, M5, and M7, particle size distribution is described by a lognormal function with a geometric mean radius and a geometric standard deviation based on the OPAC (Optical properties of aerosols and clouds) database (Hess et al. 1998). In M3 and M4, sulfate, BC and OC are parameterized in bulk mode, and a sectional scheme is used for sea-salt and dust aerosols. M2 uses an 8 bins sectional aerosol scheme with size sections ranging from 39nm to 10μm. The refractive index of

different aerosol components in the models are mainly taken from the d'Almeida et al. (1991) literature or the OPAC database. All models except M6 use a kappa (κ) parameterization to describe aerosol hygroscopic growth (Petters and Kreidenweis, 2007), in which the hygroscopicity κ values largely vary among different aerosol chemical components. For example, κ=0 for black carbon, and κ>0.6 for inorganic aerosols. M6 uses a different hygroscopic growth 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 that inorganic and carbonaceous aerosols are internally mixed but 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.

As shown in Figure 9, the observed total inorganic aerosol concentration in Beijing on 19 January 2010 was about 130μg/m^{3}with sulfate concentrations higher than 50μg/m^3 and nitrate concentrations over 60μg/m^3. However, all models except M5 largely underestimated sulfate concentrations. Most models except M2 underpredicted nitrate concentrations. . The predicted concentrations of inorganic aerosols (the sum of sulfate, nitrate and ammonium) from M2 (175μg/m^3) is higher than observations and other models (Figure 9), which can partially explain the largest simulated AOD by M2. The largest simulated AOD by M2 could also be related to different vertical distributions of aerosols. M6 simulated a similar level of inorganic aerosols as M2, but the simulated AOD is lower than other models, which could be caused by weaker hygroscopicity from a different scheme (Kiehl and Briegleb, 1993) and/or lower simulated RH (see Figure S20). Although M3 and M4 largely overpredict OC concentrations, but the mass extinction coefficient of OC is smaller than inorganic aerosols. M1 predicts about three times larger BC concentrations than the observations. Although the mass extinction coefficient of BC is

larger than inorganic aerosols, the mass concentrations and hygroscopicity of BC are smaller than those of inorganic aerosols, leading to relatively lower AOD from M1 simulation. M5 and M7 show high consistency in the simulated AOD due to similar levels of predicted inorganic aerosol concentrations (80~90μg/m^3) and similar hygroscopic growth assumptions.

As listed in Table 1, internal mixing is assumed by all the participating models except M6 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 uses external mixing but shows relatively lower AODmainly due to its ignorance of other aerosol species such as dust, sea-salt, etc. In general, the magnitudes of simulated inorganic aerosol concentrations and the hygroscopic growth efficiency (affected by varied RH) can explain the simulated variations and magnitudes of AOD in Beijing during the severe haze event, given that most models use lognormal size distributions and internal mixing assumptions.

Table 6 shows the statistics for AOD simulations at the Norch China sites and at all sites. In the NCP region, R ranges from 0.36~0.74 for all the models. It is noteworthy that R values at the sites in NCP are larger than those at all sites, indicating the larger reliability of model inputs (emissions) and meteorological simulations in North China. In terms of magnitudes, all models tend to underpredict AOD, with NMB of -2.7% to -71% in the NCP, and larger biases (NMB of -21% to-75%) at all sites. It is interesting to note that using finer grid size (M4) can produced slightly smaller NMB compared with the same model using larger grid size (M3). The effect of grid resolution will be a topic of a future paper.

4 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. 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, volcano and dust emissions were used to all the 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 could 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 the CARE-China and EANET networks showed that the main features of accumulations of air pollutants are generally represented in the current generation of online coupled air quality models. The observed variations in AOD from both the CARE-China and AERONET networks were also reproduced well by the participating models. Differences were found between simulated air pollutants, particularly ozone. While winter time ozone levels are typically low (below 40 ppb) as photochemical pathways are slow, the models captured the synoptic variability but differed in the absolute magnitudes of near surface concentrations.

Large differences in the models were found in the predicted $PM_{2.5}$ chemical compositions, especially secondary inorganics and organic carbon. During winter haze events, the production from gas phase chemistry is inhibited, and whether including other aerosol formation pathways (such as aqueous phase chemistry) leads to the large differences between simulated concentrations of secondary inorganic aerosols. In addition, different SOA treatments also lead to large discrepancies between simulated OC concentrations. Differences in the simulated variations and magnitudes of AOD in Beijing during the January 2010 haze episodes could be explained by the differences in simulated inorganic aerosol concentrations and the hygroscopic growth efficiency (affected by varied RH).

Results from this inter-comparison demonstrate that there remain important issues with current coupled models in predicting winter haze episodes. Low wind speeds play an important role in haze episodes. Current models can predict the low wind speed - high haze relationship, but overestimate the low wind speeds. This contributed to the underestimation of $PM_{2.5}$. The models also underestimate the production of secondary inorganic aerosols. There is currently a great deal of research focused on inorganic aerosol production under winter haze conditions and new pathways need to be included in the models to improve prediction skills. Furthermore, current models have various treatments of SOA production, leading to large differences in SOA predictions during 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 of aerosol composition concentrations and optical depth will impact estimates of the aerosol direct and indirect effects

during haze events (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 are the subjects of companion papers.

ACKNOWLEDGMENTS

The authors would like to acknowledge support of this project from the National Natural Science Foundation of China (91644217 and 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

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

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.

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

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

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

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

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

855 Carlton, A.G., Bhave, P.V., Napelenok, S.L., Edney, E.O., Sarwar, G., Pinder, R.W., Pouliot,
856 G.A. and Houyoux, M., 2010. Model representation of secondary organic aerosol in
857 CMAQv4. 7. Environmental science & technology, 44(22), pp.8553-8560.

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

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

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

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

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

871 Carmichael, G. R., et al. (2008a). "Predicting air quality: Improvements through advanced
872 methods to integrate models and measurements." Journal of Computational Physics227(7):
873 3540-3571.

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

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

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

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

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

887 Chou, M.D. and Suarez, M.J., 1994. An efficient thermal infrared radiation parameterization for
888 use in general circulation models (p. 85). National Aeronautics and Space Administration,
889 Goddard Space Flight Center.

890 Curci, G., Hogrefe, C., Bianconi, R., Im, U., Balzarini, A., Baró, R., Brunner, D., Forkel, R.,
891 Giordano, L., Hirtl, M. and Honzak, L., 2015. Uncertainties of simulated aerosol optical
892 properties induced by assumptions on aerosol physical and chemical properties: An
893 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.

Ek, M.B., Mitchell, K.E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G. and Tarpley, J.D., 2003. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. Journal of Geophysical Research: Atmospheres, 108(D22).

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-

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

923 Ginoux, P., Chin, M., Tegen, I., Prospero, J.M., Holben, B., Dubovik, O. and Lin, S.J., 2001.
924 Sources and distributions of dust aerosols simulated with the GOCART model. Journal of
925 Geophysical Research: Atmospheres, 106(D17), pp.20255-20273.

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

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

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

933 Gao, M., et al. (2016c). "Response of winter fine particulate matter concentrations to emission
934 and meteorology changes in North China." Atmospheric Chemistry and Physics16(18):
935 11837-11851.

936 Gao, M., et al. (2017a). "Chemical and Meteorological Feedbacks in the Formation of Intense
937 Haze Events." Air Pollution in Eastern Asia: An Integrated Perspective. Springer, Cham.
938 437-452.

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

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

944 Gery, M. W. W., G.Z.; Killus, J.P.; Dodge, M.C. (1989). "A photochemical kinetics mechanism
945 for urban and regional scale computer modeling " Journal of Geophysical Research94(D10):
946 12925-12956.

947 Grell, G.A., 1993. Prognostic evaluation of assumptions used by cumulus
948 parameterizations. Monthly Weather Review, 121(3), pp.764-787.Grell, G. A., et al. (2005).
949 "Fully coupled “online” chemistry within the WRF model." Atmospheric
950 Environment39(37): 6957-6975.

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

954 Han, Zhiwei, Hiromasa Ueda, Kazuhide Matsuda, Renjian Zhang, Kimio Arao, Yutaka Kanai,
955 Hisashi Hasome, 2004. Model study on particle size segregation and deposition during
956 Asian dust events in March 2002, Journal of Geophysical Research, 109, D19205, doi:
957 10.1029/2004jd004920.

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

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

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

966 Han Zhiwei et al.,2016. Modeling organic aerosols over east China using a volatility basis-set
967 approach with aging mechanism in a regional air quality model. Atmospheric Environment
968 124 , 186-198.

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

973 Henderson-Sellers, A., 1993. A factorial assessment of the sensitivity of the BATS land-surface
974 parameterization scheme. Journal of climate, 6(2), pp.227-247.

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

977 He, H., Wang, Y., Ma, Q., Ma, J., Chu, B., Ji, D., Tang, G., Liu, C., Zhang, H. and Hao, J., 2014.
978 Mineral dust and NOx promote the conversion of SO 2 to sulfate in heavy pollution
979 days. Scientific reports, 4, p.4172.

980 Holben, B. N., Eck, T.F., Slutsker, I., Tanre, D., Buis, J.P., Setzer, A., Vermote, E., Reagan, J.A.,
981 Kaufman, Y.J., Nakajima, T. and Lavenu, F. (1998). "AERONET—A federated instrument
982 network and data archive for aerosol characterization." Remote sensing of
983 environment66(1): 1-16.

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

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

988 Hong, S.Y. and Pan, H.L., 1996. Nonlocal boundary layer vertical diffusion in a medium-range
989 forecast model. Monthly weather review, 124(10), pp.2322-2339.

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

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

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

1000 Iacono, M.J., Delamere, J.S., Mlawer, E.J., Shephard, M.W., Clough, S.A. and Collins, W.D.,
1001 2008. Radiative forcing by long-lived greenhouse gases: Calculations with the AER
1002 radiative transfer models. Journal of Geophysical Research: Atmospheres, 113(D13).

1003 Jacobson, M. Z., 2001: Global direct radiative forcing due to multicomponent anthropogenic and
1004 natural aerosols. J. Geophys. Res., 106, 1551–1568.

1005 Janssens-Maenhout, G., et al. (2015). "HTAP_v2.2: a mosaic of regional and global emission
1006 grid maps for 2008 and 2010 to study hemispheric transport of air pollution." Atmospheric
1007 Chemistry and Physics15(19): 11411-11432.

1008 Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha,
1009 S., White, G., Woollen, J. and Zhu, Y., 1996. The NCEP/NCAR 40-year reanalysis
1010 project. Bulletin of the American meteorological Society, 77(3), pp.437-471.

1011 Kanakidou, M., Seinfeld, J.H., Pandis, S.N., Barnes, I., Dentener, F.J., Facchini, M.C.,
1012 Dingenen, R.V., Ervens, B., Nenes, A.N.C.J.S.E., Nielsen, C.J. and Swietlicki, E., 2005.
1013 Organic aerosol and global climate modelling: a review. Atmospheric Chemistry and
1014 Physics, 5(4), pp.1053-1123.

1015 Kim, S. W., et al. (2009). "NO2columns in the western United States observed from space and
1016 simulated by a regional chemistry model and their implications for NOxemissions." Journal
1017 of Geophysical Research114(D11).

1018 Kim, Y. J. S., S.N.; Carmichael, G.R.; Riemer, N.; Stanier, C.O. (2014). "Modeled aerosol
1019 nitrate formation pathways during wintertime in the Great Lakes region of North America."
1020 Journal of Geophysical Research: Atmospheres119: 12420-12445.

1021 Kiehl, J.T., Briegleb, B.P., 1993. The relative roles of sulfate aerosols and greenhouse gases in
1022 climate forcing. Science 260, 311-314.

1023 Knote, C., et al. (2015). "Influence of the choice of gas-phase mechanism on predictions of key
1024 gaseous pollutants during the AQMEII phase-2 intercomparison." Atmospheric
1025 Environment115: 553-568.

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

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

Li, M., Zhang, Q., Streets, D.G., He, K.B., Cheng, Y.F., Emmons, L.K., Huo, H., Kang, S.C., Lu, Z., Shao, M. and Su, H., 2014. Mapping Asian anthropogenic emissions of non-methane volatile organic compounds to multiple chemical mechanisms. Atmospheric Chemistry and Physics, 14(11), p.5617.

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

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

Lin, Y.L., Farley, R.D. and Orville, H.D., 1983. Bulk parameterization of the snow field in a cloud model. Journal of Climate and Applied Meteorology, 22(6), pp.1065-1092.

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

Maloney, E.D. and Hartmann, D.L., 2001. The sensitivity of intraseasonal variability in the NCAR CCM3 to changes in convective parameterization. Journal of Climate, 14(9), pp.2015-2034.

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.

Mlawer, E.J., Taubman, S.J., Brown, P.D., Iacono, M.J. and Clough, S.A., 1997. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated‐k model for the longwave. Journal of Geophysical Research: Atmospheres, 102(D14), pp.16663-16682.

Morrison, H., Curry, J.A. and Khvorostyanov, V.I., 2005. A new double-moment microphysics parameterization for application in cloud and climate models. Part I: Description. Journal of the Atmospheric Sciences, 62(6), pp.1665-1677.

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.

1079 Petters, M.D., Kreidenweis, S.M., 2007. A single parameter representation of hygroscopic
1080 growth and cloud condensation nucleus activity. Atmos. Chem. Phys. 7, 1961-1971.

1081 Ramanathan, V. C., G. (2008). "Global and regional climate changes due to black carbon."
1082 Nature Geoscience1(4): 221-227.

1083 Randerson, J.T., G.R. van der Werf, L. Giglio, G.J. Collatz, and P.S. Kasibhatla. 2015. Global
1084 Fire Emissions Database, Version 4 (GFEDv4). ORNL DAAC, Oak Ridge, Tennessee,
1085 USA. http://dx.doi.org/10.3334/ORNLDAAC/1293

1086 Reisner, J., RaSmuSSen, R.M. and Bruintjes, R.T., 1998. Explicit forecasting of supercooled
1087 liquid water in winter storms using the MM5 mesoscale model. Quarterly Journal of the
1088 Royal Meteorological Society, 124(548), pp.1071-1107.

1089 San José, R., Pérez, J.L., Balzarini, A., Baró, R., Curci, G., Forkel, R., Galmarini, S., Grell, G.,
1090 Hirtl, M., Honzak, L. and Im, U., 2015. Sensitivity of feedback effects in CBMZ/MOSAIC
1091 chemical mechanism. Atmospheric Environment, 115, pp.646-656.

1092 Schell, B., Ackermann, I.J., Hass, H., Binkowski, F.S. and Ebel, A., 2001. Modeling the
1093 formation of secondary organic aerosol within a comprehensive air quality model
1094 system. Journal of Geophysical Research: Atmospheres, 106(D22), pp.28275-28293.

1095 Schuster, G. L., et al. (2006). "Angstrom exponent and bimodal aerosol size distributions."
1096 Journal of Geophysical Research111(D7).

1097

1098 Seinfeld, J.H. and Pandis, S.N., 2016. Atmospheric chemistry and physics: from air pollution to
1099 climate change. John Wiley & Sons.

1100 Stockwell, W.R., Middleton, P., Chang, J.S. and Tang, X., 1990. The second generation regional
1101 acid deposition model chemical mechanism for regional air quality modeling. Journal of
1102 Geophysical Research: Atmospheres, 95(D10), pp.16343-16367.

1103 Stockwell, W. R., et al. (1997). "A new mechanism for regional atmospheric chemistry
1104 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, W.K., Simpson, J., Baker, D., Braun, S., Chou, M.D., Ferrier, B., Johnson, D., Khain, A., Lang, S., Lynn, B. and Shie, C.L., 2003. Microphysics, radiation and surface processes in the Goddard Cumulus Ensemble (GCE) model. Meteorology and Atmospheric Physics, 82(1), pp.97-137.

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.

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

Atmospheres102(D23): 28589-28612.

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

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

1132 Wang, J., et al. (2014). "Impact of aerosol–meteorology interactions on fine particle pollution
1133 during China's severe haze episode in January 2013." Environmental Research Letters9(9):
1134 094002.

1135 Wang, T., et al. (2010). "Investigations on direct and indirect effect of nitrate on temperature and
1136 precipitation in China using a regional climate chemistry modeling system." Journal of
1137 Geophysical Research115.

1138 Wang, Z. Maeda., T.; Hayashi, M.; Hsiao, L.F.; Liu, K.Y. (2001). "A nested air quality
1139 prediction modeling system for urban and regional scales: application for high-ozone
1140 episode in Taiwan." Water, Air, & Soil Pollution130(1): 391-396

1141 Xiao, H., et al. (1997). "Long-range transport of Sox and dust in East Asia during the PEM B
1142 Experiment." Journal of Geophysical Research:

1143 Xin, J., et al. (2015). "The Campaign on Atmospheric Aerosol Research Network of China:
1144 CARE-China." Bulletin of the American Meteorological Society96(7): 1137-1155.

1145 Zaveri, R. A., et al. (2008). "Model for Simulating Aerosol Interactions and Chemistry
1146 (MOSAIC)." Journal of Geophysical Research113(D13).

1147 Zaveri, R. A. and L. K. Peters (1999). "A new lumped structure photochemical mechanism for
1148 large-scale applications." Journal of Geophysical Research: Atmospheres104(D23): 30387-
1149 30415.

1150 Zhao, B., Wang, S., Donahue, N.M., Jathar, S.H., Huang, X., Wu, W., Hao, J. and Robinson,
1151 A.L., 2016. Quantifying the effect of organic aerosol aging and intermediate-volatility
1152 emissions on regional-scale aerosol pollution in China. Scientific reports, 6, p.28815.

1153 Zhang, Y., et al. (2010). "Simulating chemistry–aerosol–cloud–radiation–climate feedbacks over
1154 the continental U.S. using the online-coupled Weather Research Forecasting Model with
1155 chemistry (WRF/Chem)." Atmospheric Environment44(29): 3568-3582.

1156 Zheng et al., 2015. Heterogeneous chemistry: a mechanism missing in current models to explain
1157 secondary inorganic aerosol formation during the January 2013 haze episode in North
1158 China. Atmos. Chem. Phys., 15, 2031–204

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	34 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 (Lin et al., 1983)	Morrison double-moment (Morrison et al., 2005)	GCE (Goddard Cumulus Ensemble)	GCE (Tao et al., 2003)	Reisner et al., 1998	Nogherotto et al., 2016	Lin scheme
Longwave radiation	RRTMG (Iacono et al., 2008)	RRTMG	Goddard (Chou et al., 1994)	Goddard	CCM3 (Maloney et al., 2001)	CCM3	RRTM (Mlawer et al. 1997)
Shortwave radiation	RRTMG	RRTMG	Goddard	Goddard	Revised CCM3	CCM3	Goddard
Boundary Layer	Yonsei University	Yonsei University	YSU	YSU	MRF (Hong and Pan, 1996)	Holtslag et al., 1993	Yonsei University

Table 1

	M1	M2	M3	M4	M5	M6	M7
Cu physics	Grell 3D (Grell, 1993)	Grell 3D	Grell 3D	Grell 3D	Grell 3D	Anthes et al. (1977)	Grell 3D
Surface physics	Thermal diffusion	Unified Noah (Ek et al., 2003)	Unified Noah	Unified Noah	BATS (Henderson-Sellers, 1993)	BATS	Unified Noah
Aerosol-radiation	Yes	Yes	Yes	Yes	Yes	Yes	Yes [1]
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

represents not considered in the simulaiton; 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

1 Online coupled WRF-CMAQ only considers aerosol-radiation interactions but no aerosol indirect effects. The WRF-CMAQ results shown in this paper are from an offline simulation (aerosol-radiation interaction was turned off).

Table 2 CARE-Chine network sites

ID	Site name	Characteristics	Longitude	Latitude
1	Beijing	Air quality*, 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	Beijing Forest	AOD	115.43	39.97
6	Baoding	AOD	115.51	38.87
7	Cangzhou	AOD	116.80	38.28
8	Shenyang	AOD	123.63	41.52
9	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

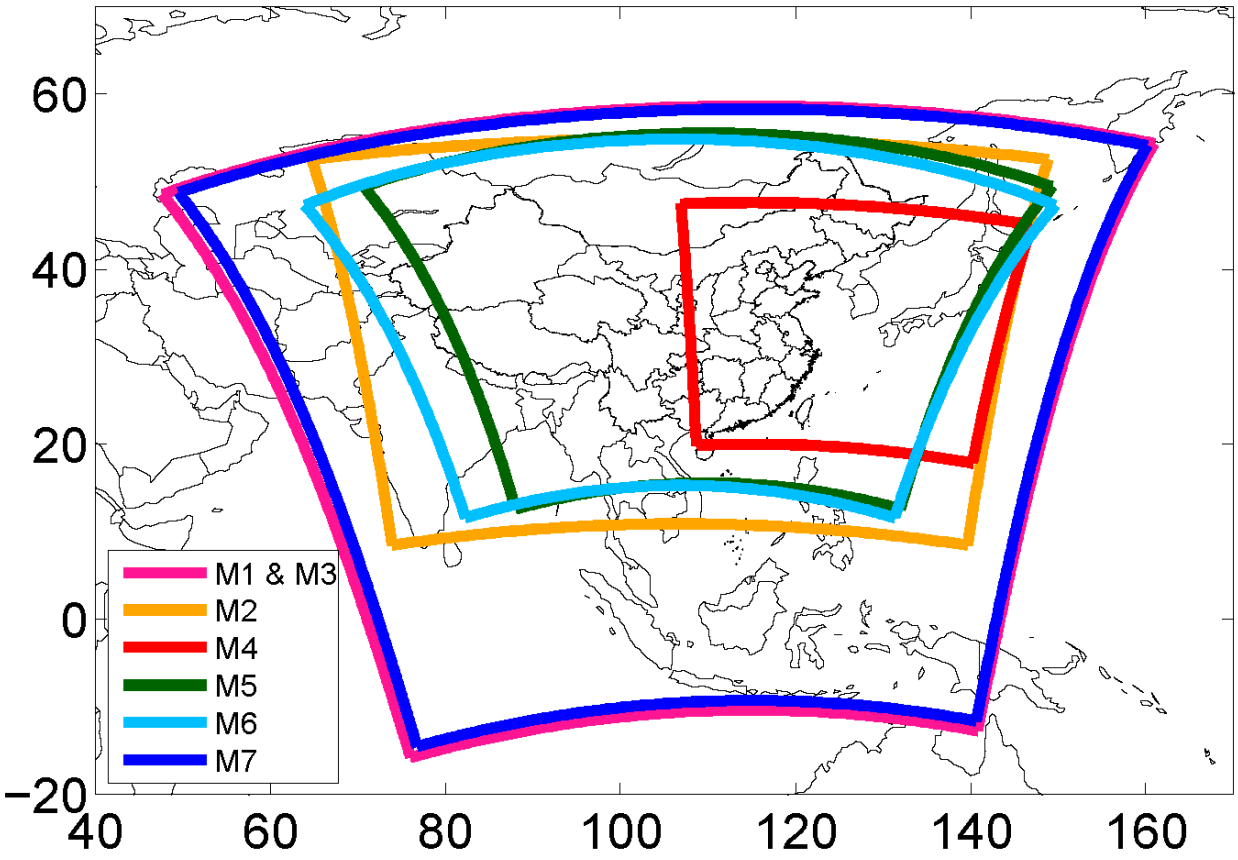

Figure 1. MICS-ASIA III Topic 3 modeling domains (descriptions of each model are documented in Table 1) M1: WRF-Chem 45km; M2: WRF-Chem 50km; M3: NU WRF 45km; M4: NU-WRF 15km; M5: RIEMS-IAP 60km; M6: RegCCMS 50km; M7: WRF-CMAQ 45km

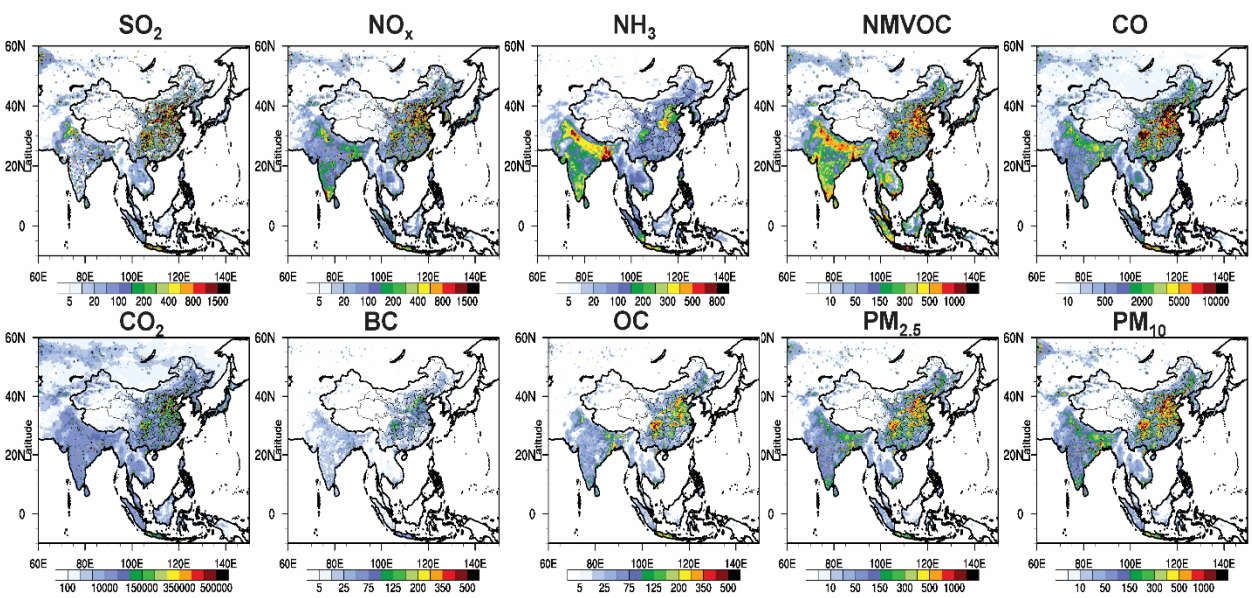

Figure 2. MIX emission inventory for January 2010 (Mg/month/grid)

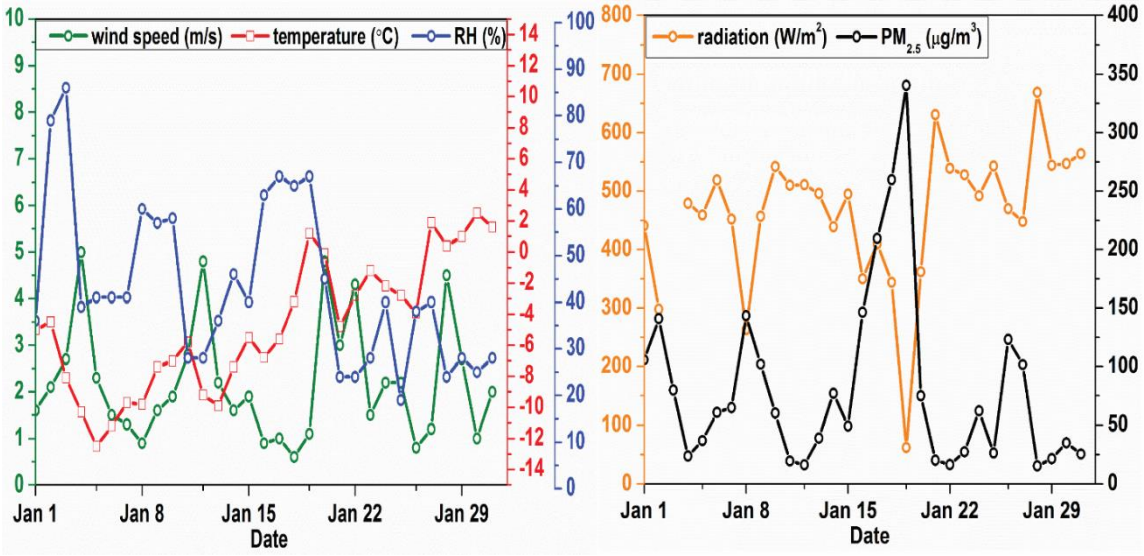

Figure 3. Observed near surface daily meteorological variables and $PM_{2.5}$ concentrations in Beijing for January 2010

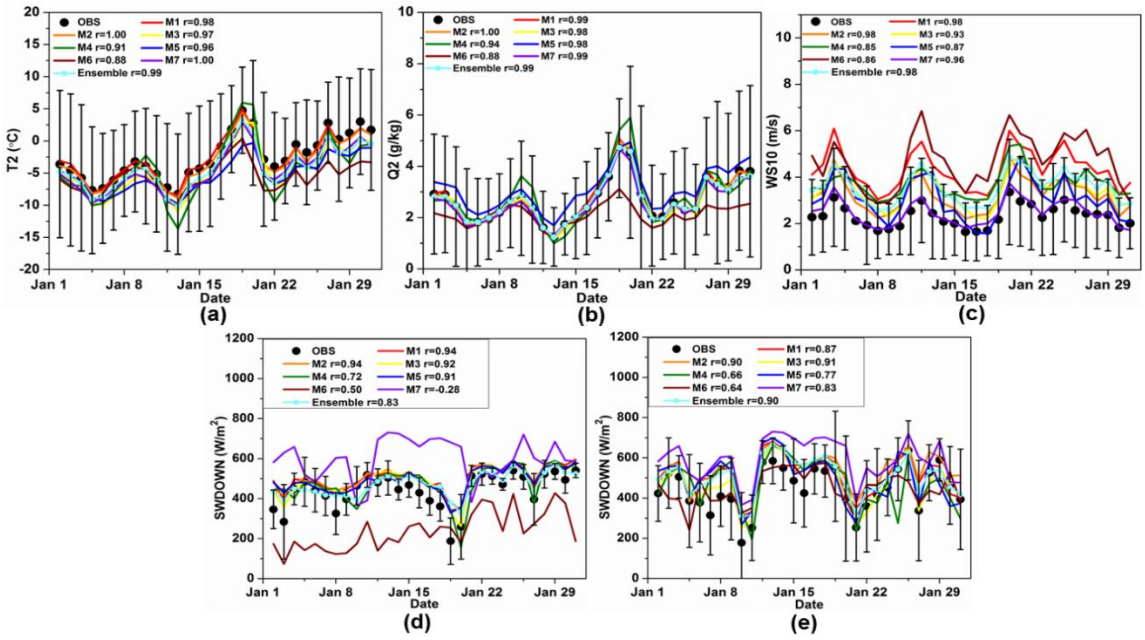

Figure 4. Comparisons between simulated and observed near surface temperature (a), water vapor mixing ratio (b), and wind speeds (c) (T2, Q2, and WS10), downward shortwave radiation in North China (d) and South China (e) (spatial daily values are averaged over measurements shown in S4 and S5; the error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

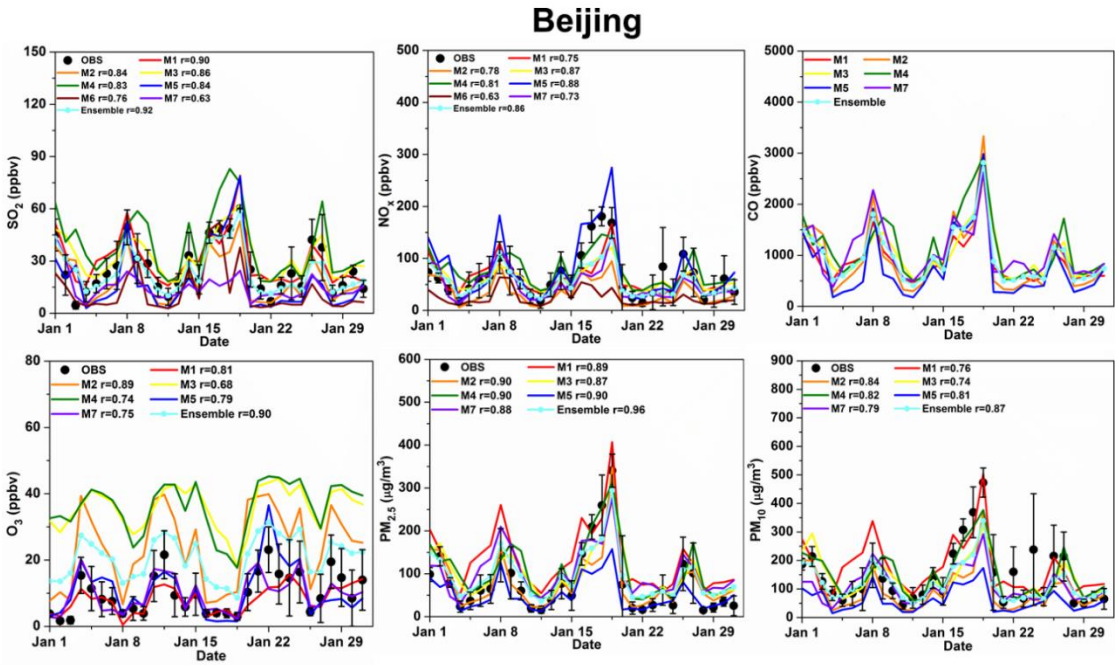

Figure 5. Comparisons between simulated and observed daily air pollutants (SO_2, NO_x, CO, O_3, $PM_{2.5}$ and PM_{10}) at the Beijing CARE-China site

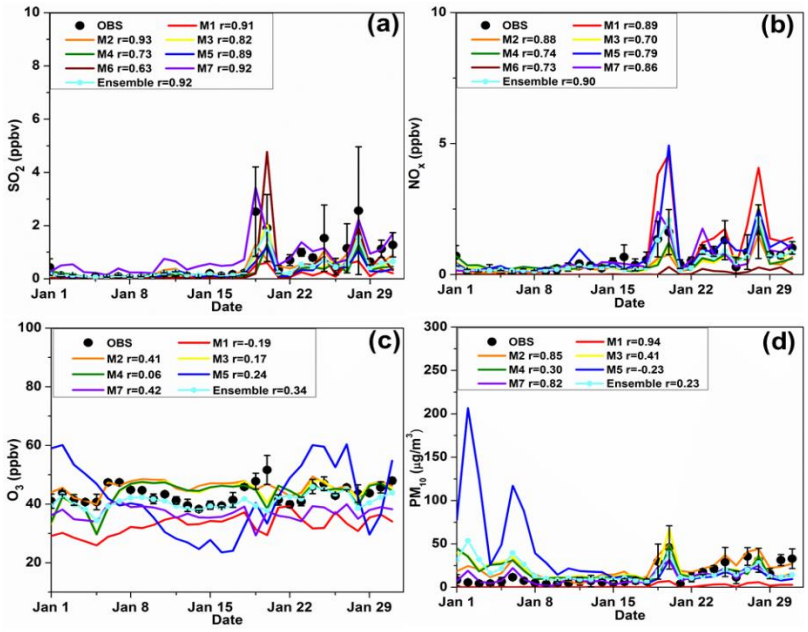

Figure 6. Comparisons between simulated and observed daily air pollutants (SO_2, NO_x, O_3, and PM_{10}) at the Rishiri EANET sites

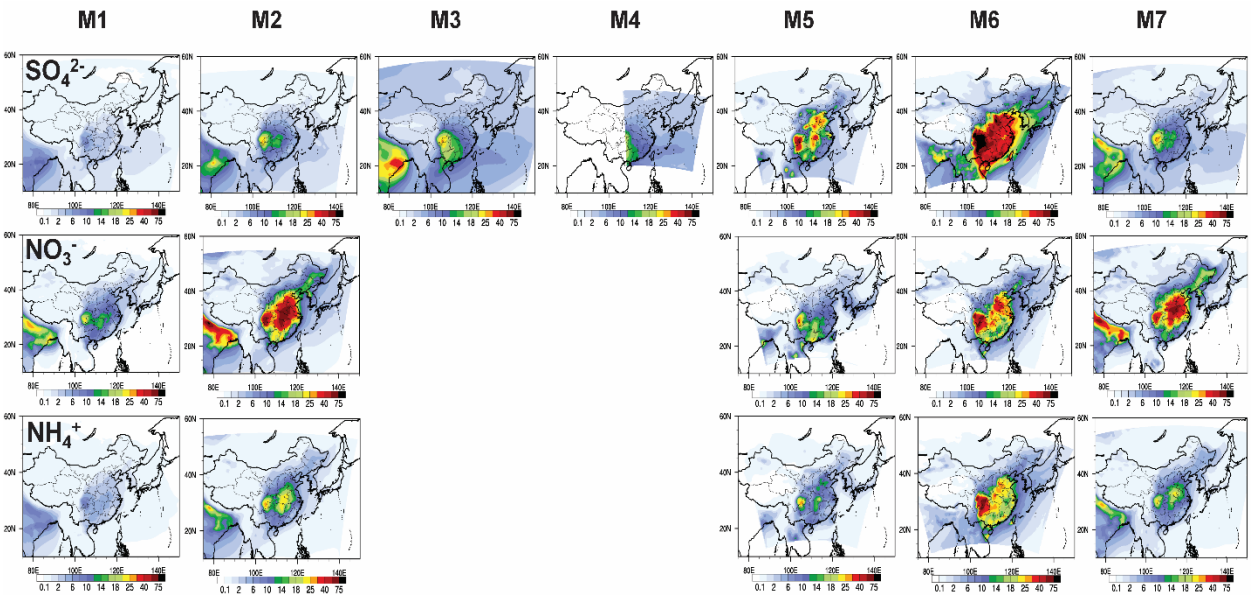

Figure 7. Simulated monthly concentrations of major $PM_{2.5}$ components (μg/m^3) for January 2010 from all participating models

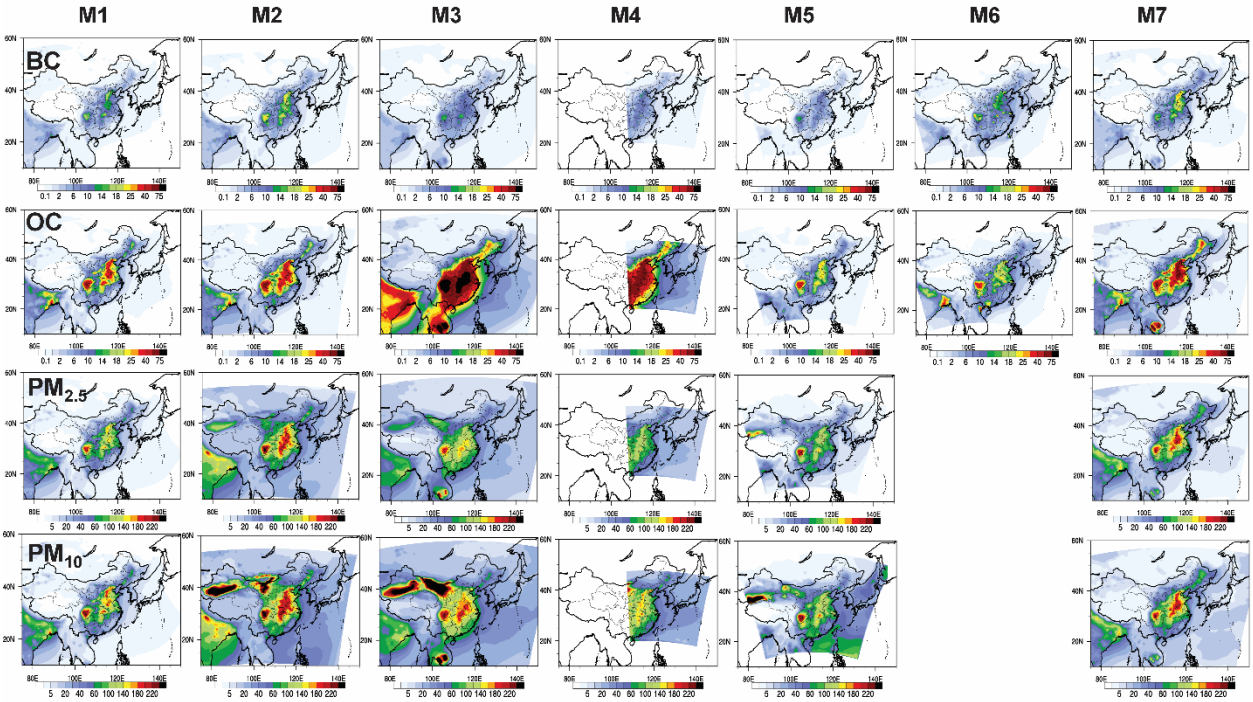

Figure 8. Simulated monthly concentrations of $PM_{2.5}$ and major $PM_{2.5}$ components ($\mu g/m^3$) for January 2010 from all participating models

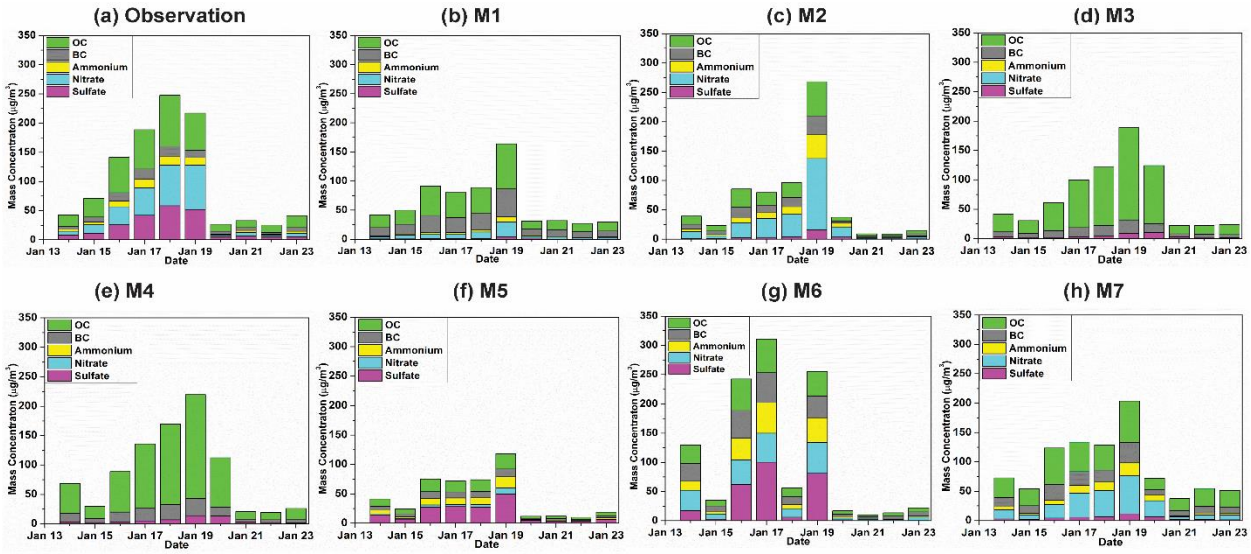

Figure 9. Observed and simulated daily mean concentrations of major $PM_{2.5}$ chemical components in the urban Beijing site

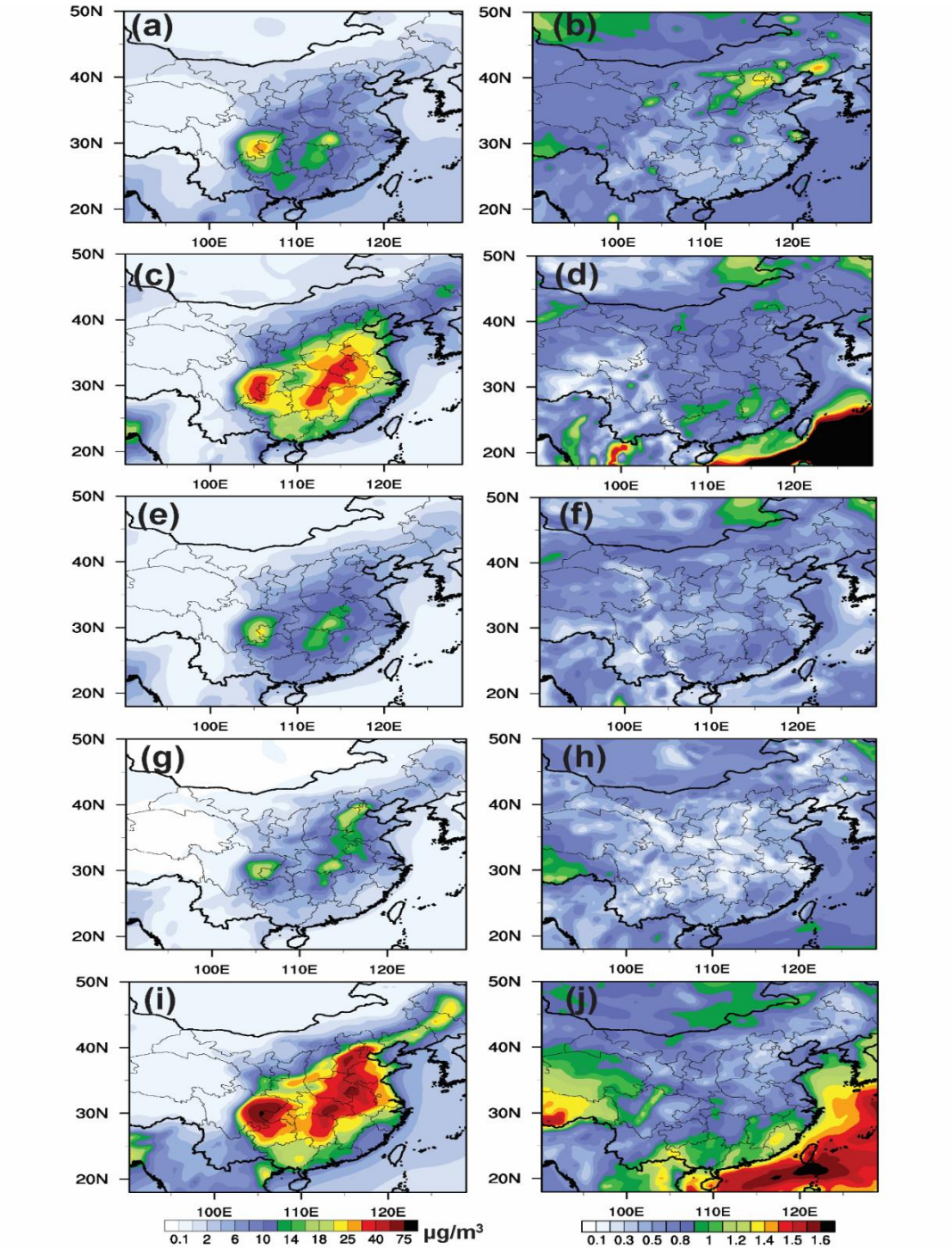

Figure 10. The ensemble mean monthly averaged near-surface distributions of $PM_{2.5}$ compositions for January 2010 (sulfate (a), nitrate (c), ammonium (e), BC (g), and OC (i)), along with the spatial distribution of the coefficient of variation ((b), (d), (f), (h), and (j), standard deviation divided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

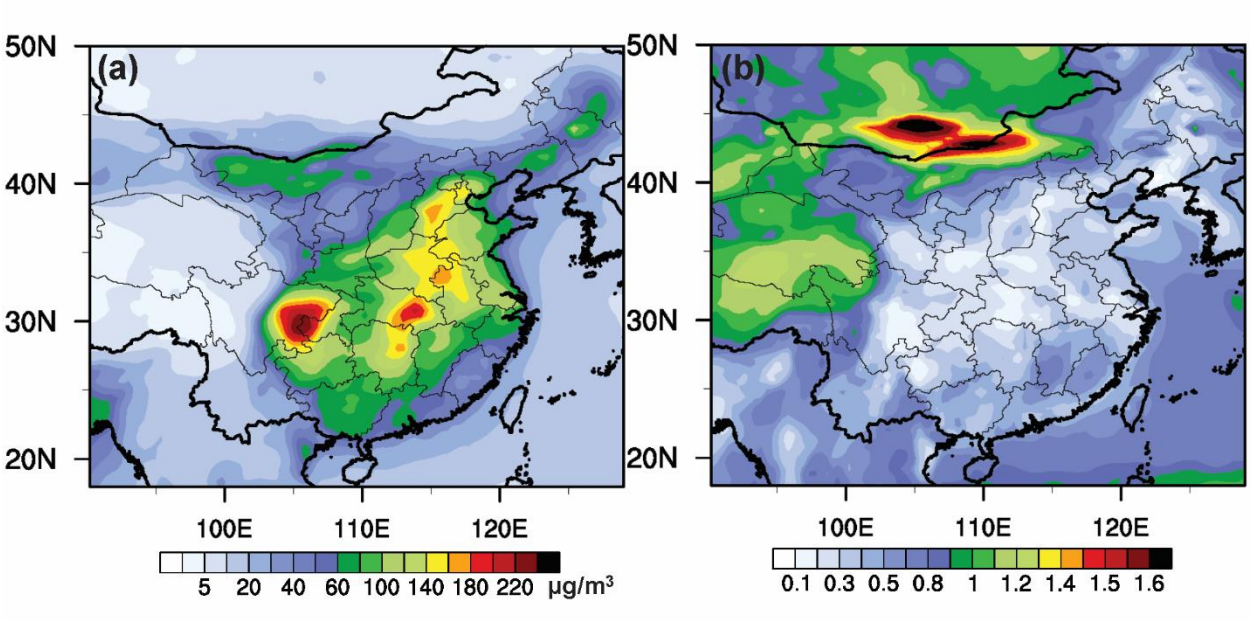

Figure 11. The ensemble mean monthly averaged near-surface distributions of $PM_{2.5}$ for January 2010 (a), along with the spatial distribution of the coefficient of variation (b, standard deviation divided by the average)

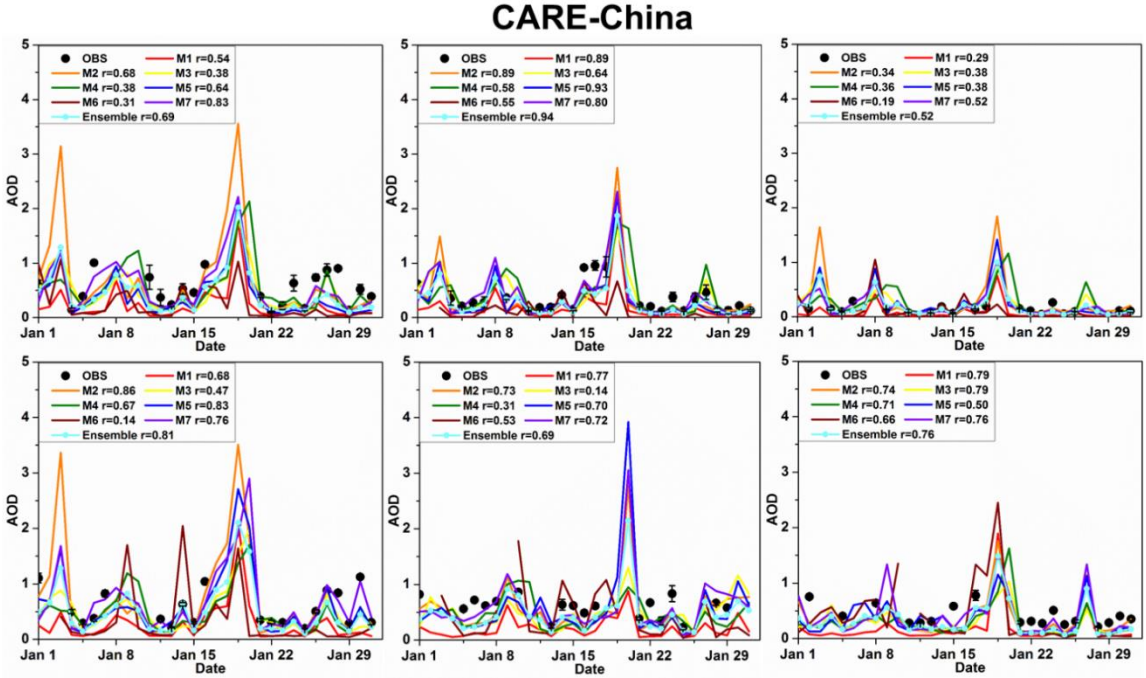

Figure 12. Comparisons between simulated and observed daily (daytime) mean AOD at the CARE-China sites (Baoding, Beijing City, Beijing Forest, Cangzhou, Jiaozhou, Shenyang,)

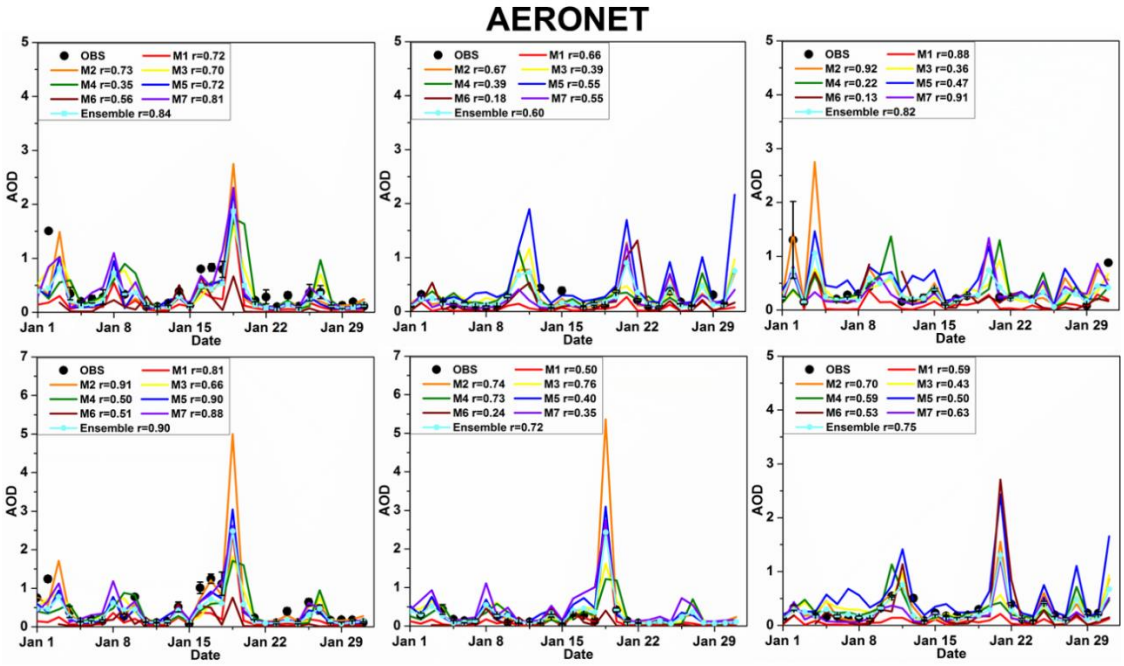

Figure 13. Comparisons between simulated and observed daily (daytime) mean AOD at the AERONET sites (Beijing, Shirahama, GIST, Xianghe, Xinglong, Osaka)