Inter-comparison of multiple two-way coupled meteorology and air quality models (WRF v4.1.1-CMAQ v5.3.1, WRF-Chem v4.1.1, and WRF v3.7.1-CHIMERE v2020r1) in eastern China

Chao Gao1,2, Xuelei Zhang1,2, Aijun Xiu1,2, *, Qingqing Tong1,2, Hongmei Zhao1,2, Shichun Zhang1,2, Guangyi Yang1,2,3, Mengduo Zhang1,2,3, and Shengjin Xie1,2,4

1Key Laboratory of Wetland Ecology and Environment, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, 130102, China
2Key Laboratory of Wetland Ecology and Environment, State Key Laboratory of Black Soils Conservation and Utilization, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, 130102, China
3University of Chinese Academy of Sciences, Beijing, 100049, China
4School of Environment, Harbin Institute of Technology, 150000, Harbin, China
Correspondence to: X.L. Zhang (zhangxuelei@iga.ac.cn) & A.J. Xiu (xiuaijun@iga.ac.cn)

Abstract

In the eastern China region, two-way coupled meteorology and air quality models have been applied aiming to more realistically simulate meteorology and air quality by accounting for the aerosol–radiation–cloud interactions. There have been numerous related studies being conducted, but the performances of multiple two-way coupled models simulating meteorology and air quality have not been compared in this region. In this study, we systematically evaluated annual and seasonal meteorological and air quality variables simulated by three open-source and widely used two-way coupled models (i.e., WRF-CMAQ, WRF-Chem, and WRF-CHIMERE) by validating the model results with surface and satellite observations for eastern China during 2017. Note that although we have done our best to keep the same configurations, this study is not aiming to screen which model is better or worse since different setups are still presented in simulations. Our evaluation results showed that all three two-way coupled models reasonably well simulated the annual spatiotemporal distributions of meteorological and air quality variables. The impacts of aerosol-cloud interaction (ACI) on model performances’ improvements were limited compared to aerosol-radiation interaction (ARI), and several possible improvements on ACI representations in two-way coupled models are further discussed and proposed. When sufficient computational resources become available, two-way coupled models should be applied for more accurate air quality forecast and timely warning of heavy air pollution events in atmospheric environmental management. The potential improvements of two-way coupled models are proposed in future research perspectives.
Aerosols in the atmosphere due to anthropogenic and nature emissions not only cause air pollution but also induce climate and meteorological impacts through aerosol-radiation interaction (ARI) and aerosol-cloud interaction (ACI) (Carslaw et al., 2010; Rosenfeld et al., 2014; Fan et al., 2016; IPCC, 2021). The feedbacks of aerosols to meteorology have been widely investigated by two-way coupled meteorology and air quality models in the past two decades (Jacobson, 1994, 1997, 1998, 2001, 2002; Grell et al., 2005; Wong et al., 2012; Wang et al., 2014; Zhou et al., 2016; Briant et al., 2017; Feng et al., 2021). In these models, two-way interactions between meteorology and aerosols are enabled by including all the processes involving ARI or/and ACI (Grell and Baklanov, 2011; Wang et al., 2014; Briant et al., 2017; Wang et al., 2021). The fundamental theories, modeling technics, developments, and applications of two-way coupled meteorology and air quality models in North America, Europe and Asia have been systemically reviewed (Zhang, 2008; Baklanov et al., 2014; Gao et al., 2022).

As pointed out by these review papers, the treatments and parameterization schemes of all the physiochemical processes involving ARI and ACI can be very different in two-way coupled models, so that the simulation results from these models could vary in many aspects. At the same time, the configurations of coupled models, such as meteorological and chemical initial and boundary conditions (ICs and BCs), horizontal and vertical resolutions, and emission inventories and processing tools, etc., play important roles in models’ simulations. In the past, model inter-comparison projects have been carried out targeting various two-way coupled meteorology and air quality models. For example, the Air Quality Model Evaluation International Initiative Phase II focused on the performance of multiple two-way coupled models and the effects of aerosol feedbacks in Europe and the United States (Brunner et al., 2015; Im et al., 2015a, b; Makar et al., 2015a, b). In Asia, the Model Inter-Comparison Study for Asia Phase III was conducted to evaluate ozone (O$_3$) and other gaseous pollutants, fine particular matter (PM$_{2.5}$), and acid and reactive nitrogen deposition with various models with/out ARI or/and ACI (Li et al., 2019; Chen et al., 2019; Itahashi et al., 2020; Ge et al. al., 2020; Kong et al., 2020). With respect to this project, Gao et al. (2018, 2020) have reviewed in detail the model performance of seven two-way coupled models from different research groups in simulating a heavy air pollution episode during January 2010 in North China Plain and how aerosol feedbacks affected simulations of meteorological variables and PM$_{2.5}$ concentrations. Targeting the heavy polluted India region, Govardhan et al. (2016) compared aerosol optical depth (AOD) and various aerosol species (black carbon, mineral dust, and sea salt) modeled by WRF-Chem (with ARI) and Spectral Radiation-Transport Model for Aerosol Species (with both ARI and ACI), but under different model configurations.

So far, there is no comprehensive comparisons of multiple coupled models under the same model configuration with respect to the high aerosol loading region over eastern China, where has experienced rapid growth of economy, urbanization, population, as well as severe air quality problems in the past decades (He et al., 2002; Wang and Hao, 2012; Gao et al., 2017; Geng et al., 2021). In the eastern China region (ECR), several open-source and proprietary two-way coupled models have been applied...
to investigate the ARI and/or ACI effects, yet most studies have focused on certain short-term episodes of heavy air pollution without any year-long simulations (Xing et al., 2017; Ding et al., 2019; Ma et al., 2021). The commonly used open-source models in ECR are WRF-Chem and WRF-CMAQ (Grell et al., 2005; Wong et al., 2012), but there is no any application of the two-way coupled WRF-CHIMERE model that has been applied to examine aerosol-radiation-cloud interactions in Europe and Africa (Briant et al., 2017; Tuccella et al., 2019). At the same time, model simulations should be compared not only against surface measurement data but also satellite data (Zhao et al., 2017; Hong et al., 2017; Campbell et al., 2017; Wang et al., 2018). Even though the running time of an individual modeling system (e.g., WRF-CMAQ and WRF-CHIMERE) was evaluated by considering its online and offline versions and under various computing configurations (Wong et al., 2012; Briant et al., 2017), the computational efficiencies of multiple two-way coupled models need to be accessed under the same computing conditions as well.

In this paper, a comparative evaluation of three open-sourced two-way coupled meteorology and air quality models (WRF-CMAQ, WRF-Chem and WRF-CHIMERE) in ECR is conducted. The remainder of the paper is organized as follows: Section 2 describes the study methods including model configurations and evaluation protocols. Sections 3 and 4 presents the analyses and intercomparisons of simulations from these three two-way coupled models with regard to meteorology and air quality, respectively. The major findings of this work are summarized in Section 5.

2 Data and methods

2.1 Model configurations and data sources

One-year long-term simulations in eastern China were examined using the two-way coupled WRF-CMAQ, WRF-Chem, and WRF-CHIMERE models, with and without enabling ARI and/or ACI, and with 27-km horizontal grid spacing (there were 110, 120, and 120 grid cells in the east–west direction, and 150, 160, and 170 in the north–south direction for WRF-CMAQ, WRF-Chem, and WRF-CHIMERE, respectively). *All the three coupled models used in this study have 30 levels (i.e., 29 layers) from the surface to 100 hPa with 11 layers in the bottom 1 km and the bottom-layer thickness being 23.2 m.* The anthropogenic emissions of Multi-resolution Emission Inventory for China (MEIC) (Li et al., 2017) and the Fire INventory from NCAR version 1.5 (FINN v1.5) biomass burning emissions (Wiedinmyer et al., 2011) were applied in our simulations, and their spatial, temporal, and species allocations were performed using Python language (Wang et al., 2023). Biogenic emissions were calculated using the Model of Emissions of Gases and Aerosols from Nature version 3.0 (MEGAN v3.0) (Gao et al., 2019). Dust and sea-salt emissions were both used with calculations of inline modules, as shown in Table 1. *The meteorological ICs and lateral BCs were derived from the National Center for Environmental Prediction Final Analysis (NCEP-FNL) datasets (http://rda.ucar.edu/datasets/ds083.2), with a horizontal resolution of 1° × 1° at 6-hour intervals for each of the three coupled models, and the flux in model-top boundary is set zero.* To improve the long-term accuracy of
meteorological variables when using the WRF model, options of observational and grid
four-dimensional data assimilation (FDDA) were turned on, and pressure, station height, 
relative humidity, wind speed, and wind direction were observed four times per day at 
00:00, 06:00, 12:00, and 18:00 UTC from 2168 stations 
coupled models could dampen the simulated aerosol feedbacks (Wong et al., 2012; 
Forkel et al., 2012; Hogrefe et al., 2015; Zhang et al., 2016). To reduce the effects of 
enabling FDDA on aerosol feedbacks in long-term simulations, here the nudging 
coefficients for u/v wind, temperature, and water vapor mixing ratio above the 
planetary boundary layer were set to 0.0001 s$^{-1}$, 0.0001 s$^{-1}$, and 0.000001 s$^{-1}$, 
respectively. The chemical ICs/lateral BCs were downscaled from the Whole 
Atmosphere Community Climate Model (WACCM) for WRF-CMAQ and WRF-Chem 
via the mozart2camx and mozbc tools, respectively. WRF-CHIMERE used the 
climatology from a general circulation model developed at the Laboratoire de 
Météorologie Dynamique (LMDz) coupling a global chemistry and aerosol model 
INteractions between Chemistry and Aerosols (INCA) (Mailler et al., 2017). For 
chemical model-top BCs, WRF-CMAQ and WRF-Chem models both take into account 
the impacts of stratosphere-troposphere O$_3$ exchange using the parameterization of O$_3$-
potential vorticity (Safieddine et al., 2014; Xing et al., 2016), the related options for the 
two models were used in this study. In WRF-CHIMERE, the climatology from LMDz-
INCA data was utilized (Mailler et al., 2017).

The options of parameterization schemes of aerosol–radiation–cloud interactions 
are listed in Table 1. To keep the consistency of physical schemes, the same RRTMG 
shortwave and longwave radiation schemes and Morrison microphysics schemes are 
adopted in both WRF-Chem and WRF-CMAQ. WRF-CHIMERE applied the same 
radiation schemes and Thompson microphysics scheme. The different other schemes 
(cumulus, surface, and land surface) in WRF-CMAQ and WRF-Chem were chosen 
according to widely used options outlined in Table S1 of Gao et al. (2022). The other 
schemes used in WRF-CHIMERE are the same as with WRF-Chem. To consider the 
effects of clouds on radiative transfer calculations, the fractional cloud cover and cloud 
optical properties were included in the RRTMG shortwave/longwave radiation schemes 
used by all three coupled models (Xu and Randall, 1996; Iacono et al., 2008). The 
coupled WRF-CMAQ model with the Kain-Fritsch cumulus scheme included the 
cumulus cloud fraction impacts on RRTMG radiation (Alapaty et al., 2012), but not the 
WRF-Chem and WRF-CHIMERE models with the Grell-Freitas cumulus scheme. In the 
Fast-JX photolysis scheme used by the three coupled models, the impacts of clouds are 
included by considering cloud cover and cloud optical properties. However, the 
calculations of cloud cover and cloud optical properties are different in these models 
and all the relevant information is listed in Table S1. As illustrated in Tables 1 and S2 
for aerosol size distribution, we used modal approach with Aitken, accumulation and 
coarse modes in WRF-CMAQ, and the 4-bin and 10-bin sectional approaches in WRF-
Chem and WRF-CHIMERE models, respectively (Binkowski and Roselle, 2003; Zaveri 
et al., 2008; Nicholls et al., 2014; Menut et al., 2013, 2016).

To demonstrate the capabilities of the three two-way coupled models with/without
feedbacks in simulating meteorology and air quality, we undertook comprehensive evaluations of the strengths and weaknesses each coupled model, validated against extensive ground-based and satellite measurements. Ground-based data included 572 hourly ground-based meteorological observations (air temperature (T2) and relative humidity (RH2) air temperature at 2m above the surface, wind speed at 10m above the surface (WS10), and precipitation (PREC)) (http://data.cma.cn), 327 hourly national environmental observations (fine particulate matter (PM2.5), ozone (O3), nitrogen dioxide (NO2), sulfur dioxide (SO2), and carbon monoxide (CO)) (http://106.37.208.233:20035), 109 hourly surface shortwave radiation (SSR) measurements (Tang et al., 2019) and 74 radiosonde sites retrieved twice per day (Guo et al., 2019); the locations of these data are depicted in Fig. 1. Because there were no observed water vapor mixing ratio (w) data, this parameter was calculated via the formula $w = \frac{r_h}{w_s}$, where $r_h$ is the relative humidity and $w_s$ is the saturation mixing ratio (Wallace and Hobbs, 2006).

Satellite data included the following: monthly average downwelling short-/long-wave flux at the surface and short-/long-wave flux at the top of the atmosphere (TOA) from the Clouds and the Earth’s Radiant Energy System (CERES) (https://ceres.larc.nasa.gov); precipitation from the Tropical Rainfall Measuring Mission (TRMM); cloud fraction, liquid water path (LWP), and aerosol optical depth (AOD) from the Moderate Resolution Imaging Spectroradiometer (MODIS); tropospheric NO2 column and SO2 column in the planetary boundary layer (PBL) from the Ozone Monitoring Instrument (OMI); total CO column from the Measurements of Pollution in the Troposphere (MOPITT) (https://giovanni.gsfc.nasa.gov/giovanni); total column ozone (TCO) from the Infrared Atmospheric Sounding Interferometer-Meteorological Operational Satellite-A (IASI-METOP-A) (https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-ozone?tab=form); and total ammonia (NH3) column from IASI-METOP-B (https://cds-espri.ipsl.fr/iasibl3/iasi_nh3/V3.1.0). These data were downloaded and interpolated to the same horizontal resolution as the model results using Rasterio library (Gillies et al., 2013), then the model and observed values at each grid point were extracted.
Figure 1. Modeling domains (WRF-CMAQ, WRF-Chem, and WRF-CHIMERE), and solar radiation, meteorology, air quality, and radiosonde stations.

Table 1. Model setups and inputs for the two-way coupled models (WRF-CMAQ, WRF-Chem and WRF-CHIMERE).

<table>
<thead>
<tr>
<th></th>
<th>WRF-CMAQ</th>
<th>WRF-Chem</th>
<th>WRF-CHIMERE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Domain</strong></td>
<td>Horizontal grid spacing</td>
<td>27 km (110 × 150)</td>
<td>27 km (120 × 160)</td>
</tr>
<tr>
<td><strong>configuration</strong></td>
<td>Vertical resolution</td>
<td>30 levels</td>
<td>30 levels</td>
</tr>
<tr>
<td><strong>Physics</strong></td>
<td>Shortwave radiation</td>
<td>RRTMG</td>
<td>RRTMG</td>
</tr>
<tr>
<td><strong>parameterization</strong></td>
<td>Longwave radiation</td>
<td>RRTMG</td>
<td>RRTMG</td>
</tr>
<tr>
<td><strong>Cloud microphysics</strong></td>
<td>Morrison</td>
<td>Morrison</td>
<td>Thompson</td>
</tr>
<tr>
<td><strong>PBL</strong></td>
<td>ACM2</td>
<td>YSU</td>
<td>YSU</td>
</tr>
<tr>
<td><strong>Cumulus</strong></td>
<td>Kain-Fritsch</td>
<td>Grell-Freitas</td>
<td>Grell-Freitas</td>
</tr>
<tr>
<td><strong>Surface</strong></td>
<td>Pleim-Xiu</td>
<td>Monin-Obukhov</td>
<td>Monin-Obukhov</td>
</tr>
<tr>
<td><strong>Land surface</strong></td>
<td>Pleim-Xiu LSM</td>
<td>Noah LSM</td>
<td>Noah LSM</td>
</tr>
<tr>
<td><strong>Icloud</strong></td>
<td>Xu-Randall method</td>
<td>Xu-Randall method</td>
<td>Xu-Randall method</td>
</tr>
<tr>
<td><strong>Chemistry</strong></td>
<td>Aerosol mechanism</td>
<td>AERO6</td>
<td>MOSAIC</td>
</tr>
<tr>
<td><strong>scheme</strong></td>
<td>Aerosol size distribution</td>
<td>Modal (3 modes)</td>
<td>Sectional (4 bins)</td>
</tr>
<tr>
<td><strong>Aerosol mixing state</strong></td>
<td>Core-Shell</td>
<td>Core-Shell</td>
<td>Core-Shell</td>
</tr>
<tr>
<td><strong>Gas-phase chemistry</strong></td>
<td>CB6</td>
<td>CBMZ</td>
<td>MELCHIOR2</td>
</tr>
<tr>
<td><strong>Photolysis</strong></td>
<td>Fast-JX with cloud effects</td>
<td>Fast-JX with cloud effects</td>
<td>Fast-JX with cloud effects</td>
</tr>
<tr>
<td><strong>Emission</strong></td>
<td>Anthropogenic emission</td>
<td>MEIC 2017</td>
<td>MEIC 2017</td>
</tr>
<tr>
<td><strong>Biogenic emission</strong></td>
<td>MEGAN v3.0</td>
<td>MEGAN v3.0</td>
<td>MEGAN v3.0</td>
</tr>
<tr>
<td><strong>Biomass burning emission</strong></td>
<td>FINN v1.5</td>
<td>FINN v1.5</td>
<td>FINN v1.5</td>
</tr>
<tr>
<td><strong>Dust emission</strong></td>
<td>Foroutan</td>
<td>GOCART</td>
<td>Menut</td>
</tr>
<tr>
<td><strong>Sea-salt emission</strong></td>
<td>Gong</td>
<td>Gong</td>
<td>Monahan</td>
</tr>
<tr>
<td><strong>Input data</strong></td>
<td>Meteorological ICs and BCs</td>
<td>FNL</td>
<td>FNL</td>
</tr>
<tr>
<td><strong>Chemical ICs and BCs</strong></td>
<td>MOZART</td>
<td>MOZART</td>
<td>LMDZ-INCA</td>
</tr>
</tbody>
</table>

2.2 Scenario setup

To thoroughly assess the performance of WRF v4.1.1-CMAQ v5.3.1, WRF-Chem v4.1.1, and WRF v3.7.1-CHIMERE v2020r1 and its affected by aerosol feedbacks over eastern during 2017, eight sets of annual hindcast simulations with/without ARI and/or ACI were conducted, as presented in Table 2. Compared to WRF v4.1.1-CMAQ v5.3.1 and WRF-Chem v4.1.1, this version of WRF v3.7.1-CHIMERE v2020r1 can be officially obtained and the higher version of WRF-CHIMERE has not been developed. It should be noted that the officially released WRF-Chem and WRF-CHIMERE are capable of simulating ARI and ACI, but WRF-CMAQ is not. In all of the simulations performed in this study, a month of spin-up time was set up to reduce the influence of the initial conditions. Multiple statistical metrics between each scenario simulation and ground-based/satellite-borne observations were used including the correlation coefficient (R), mean bias (MB), normalized mean bias (NMB), normalized gross error (NGE), and root mean square error (RMSE). The mathematical definitions of these...
metrics are provided in Supplement S1. *To compare simulations by three coupled models, the respective model configurations of physics and chemistry routines are set as consistent as possible.* We systemically analyzed the annual and seasonal statistical metrics of meteorological and air quality variables including simulations by all three two-way coupled models with/without enabling ARI and/or ACI effects. We then quantified the respective contributions of the ARI and ACI effects to model performance.

### Table 2. Summary of scenarios setting in three coupled models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Scenario</th>
<th>Configuration option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>WRF-CMAQ</td>
<td>(1) WRF-CMAQ_NO</td>
<td>DO_SW_CAL=F</td>
<td>Without aerosol feedbacks</td>
</tr>
<tr>
<td></td>
<td>(2) WRF-CMAQ_ARI</td>
<td>DO_SW_CAL=T</td>
<td>ARI</td>
</tr>
<tr>
<td>WRF-Chem</td>
<td>(3) WRF-Chem_NO</td>
<td>aer_ra_feedback=0</td>
<td>Without aerosol feedbacks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>wetscav_onoff=0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>cldchem_onoff=0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4) WRF-Chem_ARI</td>
<td>aer_ra_feedback=1</td>
<td>ARI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>wetscav_onoff=0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>cldchem_onoff=0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5) WRF-Chem_BOTH</td>
<td>aer_ra_feedback=1</td>
<td>ARI and ACI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>wetscav_onoff=1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>cldchem_onoff=1</td>
<td></td>
</tr>
<tr>
<td>WRF-CHIMERE</td>
<td>(6) WRF-CHIMERE_NO</td>
<td>direct_feed_chimere=0</td>
<td>Without aerosol feedbacks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>indirect_feed_chimere=0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7) WRF-CHIMERE_ARI</td>
<td>direct_feed_chimere=1</td>
<td>ARI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>indirect_feed_chimere=0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(8) WRF-CHIMERE_BOTH</td>
<td>direct_feed_chimere=1</td>
<td>ARI and ACI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>indirect_feed_chimere=1</td>
<td></td>
</tr>
</tbody>
</table>

### 3 Multi-model meteorological evaluations

This section presents annual and seasonal (March–April–May, Spring; June–July–August, Summer; September–October–November, Autumn; and December–January–February, Winter) statistical metrics of simulated meteorological variables and air quality when compared with ground-based and satellite observations, as well as a discussion of the running times of the eight scenario simulations.

#### 3.1 Ground-based observations

Figures 2 and S1–S7 illustrate the spatial distributions of R, MB, and RMSE for hourly SSR, T2, Q2, RH2, WS10, PREC, PBLH00, and PBLH12 from WRF-CMAQ, WRF-Chem, and WRF-CHIMERE with/without turning on aerosol feedbacks against ground-based observations from each site across the whole of 2017. The calculated annual model evaluation metrics for all sites in eastern China are summarized in Table S1, and the related seasonal R and MB values are presented in Fig. 3. *Here, we mainly focused on the comparisons of SSR, T2, RH2, and WS10, and the analysis of PREC, PBLH00, and PBLH12 are presented in Section 1.1 of Supplement.*
The accuracy of radiation prediction is of great significance in evaluating ARI. Yearly and seasonal average simulated SSR data were compared with ground-based observations (Figs. 3–4 and Table S3), and SSR over eastern China was simulated reasonably well by all models with R values in the range of 0.61–0.78. All simulated results were overestimated at both annual and seasonal scales (MBs in spring and summer were larger than those in autumn and winter). The overestimations of annual SSR were 19.98, 14.48, and 9.24 W m\(^{-2}\) for WRF-CMAQ, WRF-Chem, and WRF-CHIMERE, respectively. Overestimations of SSR by most two-way coupled models were also reported for Europe and North America in the comparative study conducted by Brunner et al. (2015). Such overestimations could be explained by multiple factors, namely, the uncertainties in cloud development owing to PBL and convection parameterizations (Alapaty et al., 2012), and the diversity in treatment of land surface processes (Brunner et al., 2015), which appear to play more important roles than does the enabling of two-way aerosol feedbacks on SSR through ARI and ACI effects in the models. When the three models considered ARI effects, the simulation accuracy of SSR, over both the whole year and in the four seasons were improved, but the enabling of ACI effects resulted in relatively limited improvement. In addition, the MB variations of WRF-CMAQ and WRF-Chem simulations were higher in spring and winter than those in summer and autumn, while the MB of WRF-CHIMERE simulations showed a maximum in summer (−10.33 W m\(^{-2}\)) and minimum in autumn (−7.64 W m\(^{-2}\)). Both the annual and seasonal reductions in SSR simulated by WRF-Chem and WRF-CHIMERE with ACI effects enabled were much smaller than those with ARI effects enabled.

In general, the simulated magnitudes and temporal variations of air temperature at 2 m above the ground showed a high order of consistency with observations (R = 0.88–0.97). Looking at annual and seasonal T2, models tended to have a negative bias, and T2 underestimations in spring and winter were greater than those in summer and autumn (Figs. 3 and 4). As pointed out by Makar et al. (2015a), WRF-CHEM and GEM-MACH gave negative MBs in summer and positive MBs in winter when both ACI and ARI effects were enabled, and WRF-CMAQ with only ARI effects enabled also produced negative MBs in summer over North America during 2010; note that the Makar et al (2015a) study lacked evaluations of meteorology in winter using WRF-CMAQ. The comparison results of MBs indicated that WRF-CHIMERE > WRF-CMAQ > WRF-Chem. The annual and seasonal MBs of WRF-CMAQ and WRF-Chem were approximately −1 °C, while those of WRF-CHIMERE ranged from −2 to −1 °C. The RMSEs were approximately equal for WRF-CMAQ (2.71–3.05 °C) and WRF-Chem (2.82–3.27 °C), and larger for WRF-CHIMERE (3.39–4.53 °C) at both annual and seasonal scales. It is noteworthy that underestimations of annual and seasonal T2 were mitigated in eastern China in the three coupled models when ARI effects were enabled. When ACI effects were enabled, the MBs for T2 simulated by WRF-Chem_BOTH showed no significant changes compared with those of WRF-Chem_NO; WRF-CHIMERE_BOTH further enhanced the underestimations of T2 in the full year (−1.30 °C), spring (−0.12 °C), and winter (−0.40 °C) compared with WRF-CHIMERE_NO.
Looking at RH2, annual and seasonal simulations using WRF-CMAQ had the highest correlation with the observed values, followed by WRF-Chem, and WRF-CHIMERE, and the smallest correlation coefficients for all three models occurred in autumn (~0.5). The spatial MBs between simulations by the three models and observations showed a general converse trend compared with T2 (i.e., RH2 was overestimated where T2 was underestimated, and vice versa). This can be explained by the calculation of RH2 being based on T2 in the models (Wang et al., 2021). The annual and seasonal MBs were approximately 0.65%–71.03% and −21.30% to 60.00%, respectively (Fig. 4 and Table S3), and only WRF-Chem produced negative MBs in summer. The magnitude of RMSE showed an inverse pattern compared with R for all three models, with maximum (28.48%–29.52%) and minimum (12.57%–16.07%) values shown in autumn and summer, respectively. As shown in Figs. 3–4 and Table S3, WRF-CMAQ_ARI further reduced the overestimations of annual and seasonal RH2 in eastern China, while WRF-Chem_ARI (except for summer) and WRF-CHIMERE_ARI showed the opposite trend. Moreover, variations in annual and seasonal RH2 MBs simulated by WRF-Chem_BOTH and WRF-CHIMERE_BOTH were further reduced compared with WRF-Chem_ARI (except for summer) and WRF-CHIMERE_ARI, respectively.

Similar analyses were also performed for WS10, and revealed that WRF-CMAQ performed better in capturing WS10 patterns compared with WRF-Chem and WRF-CHIMERE. The R values for all three models ranged from 0.47 to 0.60; WRF-CMAQ and WRF-Chem overestimated wind speed by approximately 0.5 m s\(^{-1}\), while WRF-CHIMERE overestimated it by approximately 1.0 m s\(^{-1}\) (Table S3 and Figs. 3–4). The overestimation of WS10 under real-world low wind conditions is a common phenomenon of current weather models, which is mainly caused by outdated geographic data, coarse model resolution, and a lack of a good physical representation of the urban canopy (Gao et al., 2015, 2018). All three models presented lower correlations (0.31–0.54) and MBs (0.20–0.86 m s\(^{-1}\)) in summer compared with other seasons, and the RMSEs were approximately 2.0 m s\(^{-1}\). When ARI effects were enabled, the overestimations of the three models were alleviated, especially for WRF-CMAQ_ARI.
Figure 2. Statistical metrics (R, MB, and RMSE) between annual simulations and observations of surface shortwave radiation in eastern China.

Figure 3. Time series of observed and simulated hourly SSR, T2, RH2 and WS10 by coupled WRF-CMAQ, WRF-Chem and WRF-CHIMERE with/without aerosol feedbacks over Eastern China during the year of 2017.
To identify and quantify how well our results compare with previous studies using two-way coupled models, we here discuss comparisons between our work and earlier research in terms of the evaluation results of meteorology and air quality; meteorology is discussed in this section and air quality is discussed in Section 4.1. Box-and-whisker plots were used and the 5th, 25th, 75th, and 95th percentiles were used as statistical indicators. In the plots, the dashed lines in the boxes are the mean values, and the circles represent outliers. Previous studies mainly used WRF-Chem and WRF-CMAQ to evaluate meteorology and air quality, while applications of WRF-NAQPMS and GRAPES-CUACE were scarce. As mentioned in Section 1, investigations of meteorology and air quality using WRF-CHIMERE with/without aerosol feedbacks have not previously been conducted in eastern China. Therefore, only evaluation results involving WRF-Chem and WRF-CMAQ to study aerosol feedbacks are analyzed herein.

The statistical metrics of T2, RH2, Q2, and WS10 in this study compared with the evaluation results of previous studies are presented in Fig. S8. According to the number of samples (NS) in the statistical metrics of each meteorological variable, most previous studies mainly involved the simulation and evaluation of T2, WS10, and RH2, with relatively few studies focusing on Q2. Compared with the evaluation results of previous studies, the ranges of statistical metrics in our study were roughly similar, but there were some important differences. The R values of the WRF-CMAQ and WRF-Chem models in our study were higher than those of previous studies; the MBs of T2 simulated via WRF-CMAQ were smaller, but those of T2 simulated via WRF-Chem were larger; and the RMSEs of the WRF-CMAQ simulation were larger, but those of the WRF-Chem simulation were smaller. For RH2, the R values for WRF-CMAQ and WRF-Chem in this study were all larger than the average level of previous studies, while the MBs and RMSEs for WRF-CMAQ were larger, and those for WRF-Chem were smaller.
than the average of previous studies. For Q2, the model performance of WRF-CMAQ in this study was generally better than the average level of previous studies, but the R between WRF-Chem simulation results and observed values was higher (and MB and RMSE were lower) than the average level of previous studies. We also conclude that the simulation results of WRF-CMAQ and WRF-Chem in our study better reproduced variations in WS10 compared with previous studies.

3.2 Satellite-borne observations

To further evaluate the performance of WRF-CMAQ, WRF-Chem, and WRF-CHIMERE against satellite observations, we analyzed the annual and seasonal statistical metrics of short- and long-wave radiation at the surface, precipitation, cloud cover, and liquid water path simulated by the three coupled models with and without aerosol feedbacks, via comparisons between simulations and satellite-borne observations (Table 3; Figures 5, S9, S12–S14). In addition, the evaluations of short- and long-wave radiation at top of the atmosphere (TOA) are presented in Section 1.2 of Supplement.

As shown in Table 3 and Fig. 5, the three coupled models showed relative poor performance for the shortwave radiation variables at the surface (SSR) annual MBs of 8.21–30.74 W m\(^{-2}\), and correlations ranging from 0.61 to 0.78. A similar poor performance for shortwave radiation was also reported in the USA using the coupled WRF-CMAQ and offline WRF models (Wang et al., 2021). The overall seasonal characteristics of SSR were successfully reproduced by the three coupled models (Fig. S10). Meanwhile, no matter whether aerosol feedbacks were enabled or not, all three models overestimated seasonal SSR (except for WRF-Chem_ARI in winter), and showed higher MBs in spring and summer than in autumn and winter. The seasonal SSR overestimations may be a direct result of the underestimation of calculated AOD when considering ARI effects (Wang et al., 2021). Compared to SSR, the three coupled models predicted the longwave radiation variables at the surface (SLR) well (R values up to 0.99), with annual domain-average MBs of −9.97 to −6.05 W m\(^{-2}\). Significant seasonal differences in simulated longwave radiation were also present among the three coupled models; all WRF-CMAQ and WRF-CHIMERE scenarios gave underestimations, with maximum and minimum values of SLR in winter and summer, respectively, while the maximum underestimations of WRF-Chem occurred in autumn, especially for WRF-Chem_BOTH (Fig. S9).

As all three coupled models adopted the same grid resolution (27 × 27 km) and short- and long-wave radiation schemes (RRTMG), the above analysis demonstrated that the representation differences for aerosol components, size distributions and mechanisms contributed to the diversity of seasonal MBs (Tables 1 and S2). Moreover, the three two-way coupled models with ARI feedbacks enabled effectively improved the performances of annual and seasonal SSR; however, for SLR, performance improvements were much more variable across the three coupled models and across different scenarios with and without ARI and/or ACI feedbacks enabled (Table S4).

When ARI effects are enabled, the diversities of refractive indices of aerosol species groups lead to the discrepancies of online calculated aerosol optical properties in
different shortwave and longwave (SW and LW) bands in the RRTMG SW/LW radiation schemes of WRF-CMAQ, WRF-Chem, and WRF-CHIMERE (Tables S5–S6). The online calculated cloud optical properties induced by aerosol absorption in the RRTMG radiation schemes are different in treatments of aerosol species groups in the three coupled models. With enabling ACI effects, the activation of cloud droplets from aerosols based on the Köhler theory is taken into account in WRF-Chem and WRF-CHIMERE, in comparison to simulations without aerosol feedbacks (Table S7). The treatments of prognostic ice nucleating particles (INP) formed via heterogeneous nucleation of dust particles (diameters > 0.5 µm) and homogeneous freezing of hygroscopic aerosols (diameters > 0.1 µm) are only considered in WRF-CHIMERE, but the prognostic ice nucleating particles are not included in WRF-CMAQ and WRF-Chem. These discrepancies eventually contribute to the differences of simulated radiation changes caused by aerosols.

From IPCC 2007 to IPCC 2021, the effects of aerosol feedbacks (especially for ACI effects) on precipitation and cloud processes remain under debate. Here, we further assessed annual and season al simulated precipitation, cloud cover, and liquid water pathways in eastern China with high aerosol loadings against satellite observations (Table 3 and Figs. S12–S14), and attempted to provide new insights from a yearly perspective into enabling online feedbacks in two-way coupled modeling simulations. The results illustrated those correlations of precipitation via WRF-CMAQ (0.51–0.89) were larger than those of WRF-Chem (0.61–0.73) and WRF-CHIMERE (0.54–0.70). WRF-CMAQ had the best correlation in winter, while WRF-Chem and WRF-CHIMERE had the best correlation in spring; all three models showed their worst correlation in summer. The reason for this is that numerical models struggle to effectively capture enhanced convective activity in summer. Huang and Gao (2018) also pointed out that accurate representations of lateral boundaries are crucial in improving precipitation simulations during summer over China. WRF-CMAQ underestimated annual precipitation, with MBs of −76.49 to −51.93 mm, while WRF-Chem and WRF-CHIMERE produced large precipitation overestimations ranging from +108.04 to +207.05 mm (Table 3), especially in regions of southern China (Fig. S11). WRF-CMAQ also produced negative biases (−27.89 to +42.08 mm) of seasonal precipitation, excluding WRF-CMAQ_ARI in winter. WRF-Chem and WRF-CHIMERE only underestimated seasonal precipitation in autumn (−31.39 to −26.89 mm) and winter (−7.12 to −4.43 mm), respectively (Fig. S12). The variations in annual and seasonal MBs of precipitation were consistent with changes in cloud fraction and LWP (Zhang et al., 2016), which will be discussed in more detail below.

When aerosol feedbacks were considered, the ARI-induced reductions in the annual MBs of precipitation for WRF-CMAQ, WRF-Chem, and WRF-CHIMERE were 24.56, 12.11, and 4.70 mm, respectively. WRF-Chem_BOTH (24.9 mm) and WRF-CHIMERE_BOTH (3.41 mm) enhanced the overestimation of annual precipitation compared with WRF-Chem_ARI and WRF-CHIMERE_ARI, respectively. Significant increases (+53.15 mm) and decreases (−6.3 to −3.41 mm) in MBs in winter and summer, respectively, were produced by WRF-CMAQ and the other two models with ARI effects enabled compared with no feedbacks. WRF-Chem and WRF-CHIMERE with both ARI
and ACI effects enabled led to larger enhancements of MBs (+3.54 to +7.46) at the seasonal scale (Fig. S12). It must be noted that the discrepancies in simulated precipitation could mainly be attributed to the selection of different microphysics and cumulus schemes in WRF-CMAQ (Morrison and Kain-Fritsch), WRF-Chem (Morrison and Grell-Freitas), and WRF-CHIMERE (Thompson and Grell-Freitas).

Cloud fraction (CF) and LWP can significantly influence the spatiotemporal distributions of precipitation; our simulated results of annual and seasonal CF over eastern China are presented in Table 3 and Fig. S13. Overall, WRF-CMAQ performed best in simulating CF. The R values for WRF-Chem during summer (0.69) and winter (0.70) were larger than those of WRF-CMAQ (0.59 and 0.64) and WRF-CHIMERE (0.56 and 0.66), while WRF-CMAQ and WRF-CHIMERE showed better simulation results in winter and autumn with correlations of up to 0.89 and 0.67, respectively. All three coupled models underestimated annual and seasonal CF with MBs that ranged from −16.83% to −6.18% and −21.13% to −4.13%, respectively; these were consistent with previous two-way coupled modeling studies using WRF-CMAQ (−19.7%) and WRF-Chem (−32% to −9%) in China (Hong et al., 2017; Zhao et al., 2017). All models reasonably simulated annual LWP in eastern China, with R values above 0.55 and negative biases varying from −57.36 to −31.29 g m⁻². The underestimations were closely related to missing cloud homogeneity (Wang et al., 2015; Dionne et al., 2020) and excessive conversion of liquid to ice in all selected cloud microphysics schemes (Klein et al., 2009). As shown in Fig. S14, all models showed their best performance in simulating LWP in spring (R = 0.51–0.79) and exhibited the largest underestimations in winter (MBs of −54.82 to −40.89 g m⁻²), except for WRF-Chem, which had its maximum bias in autumn.

In terms of quantitatively determining the functions of aerosol feedbacks on CF and LWP, all simulated scenarios revealed that WRF-CMAQ_ARI overwhelmingly decreased annual and seasonal underestimations of CF (0.48%–1.05%) and LWP (3.03–4.29 g m⁻²), while there were slightly increased underestimations (CF: 0.02%–0.39%; LWP: 0.03–0.58 g m⁻²) in WRF-Chem_ARI and WRF-CHIMERE_ARI. Larger variations in annual and seasonal MBs of CF (0.23%–0.93%) and LWP (−2.96 g m⁻² to 7.38 g m⁻²) were produced by WRF-CHIMERE_BOTH compared with WRF-CHIMERE_ARI. WRF-Chem_BOTH showed equivalent variations (CF: 0.03%–0.71%; LWP: 0.02–2.89 g m⁻²) to those of WRF-Chem_ARI. This may be explained as the different parameterization treatments of cloud droplet number concentration (CDNC) simulated by the three coupled models with/without enabling ACI effects. The cloud condensation nuclei activated from aerosol particles can increase CDNC and impact on LWP and CF. Without enabling any aerosol feedbacks or only enabling ARI, the CDNC is default prescribed as a constant value of 250 cm⁻³ in the Morrison scheme of WRF-CMAQ and WRF-Chem and 300 cm⁻³ in the Thompson scheme of WRF-CHIMERE. When only ACI or both ARI and ACI are enabled, the online calculating of prognostic CDNC is performed in WRF-Chem and WRF-CHIMERE by using the method of maximum supersaturation (Abdul-Razzak and Ghan, 2002; Chapman et al., 2009; Tuccella et al., 2019). Although we have obtained preliminary quantitative results of the ACI effects on regional precipitation, CF, and LWP, it should be kept in mind that
several limitations in representing ACI effects still exist in state-of-the-art two-way coupled models; these include a lack of consideration of the responses of convective clouds to ACI (Tuccella et al., 2019), and a lack of numerical descriptions of giant cloud condensation nuclei (Wang et al., 2021) and heterogeneous ice nuclei (Keita et al., 2020).
Table 4 and Fig. 7 present the statistical metrics of annual and seasonal air pollutant concentrations (PM$_{2.5}$, O$_3$, NO$_2$, SO$_2$, and CO) simulated by each of the three coupled

4 Multi-model air quality evaluations

In a similar way to meteorology, to further determine the quantitative effects of enabling aerosol feedbacks on the simulation accuracy of air quality variables in eastern China, ground-based and satellite-borne observations were adopted as comparisons in the following evaluation analysis. The usage status of computing resources during each simulation process is also assessed in Section 4.3.

4.1 Ground-based observations

Figure 5. Spatial distributions of seasonal SSR between CERES observations and simulations from WRF-CMAQ, WRF-Chem, and WRF-CHIMERE with and without aerosol feedbacks in eastern China.
models. The evaluations between surface measurements and simulations of PM$_{2.5}$ and O$_3$ are presented below, and the performance assessments of other gaseous pollutants are in Section 2 of Supplement.

The R values of annual PM$_{2.5}$ concentrations for WRF-CMAQ (0.68) were the highest, followed by WRF-Chem (0.65–0.68), and WRF-CHIMERE (0.52–0.53). All three models showed higher correlations in winter compared with those in other seasons (Fig. 7). As shown in Table 4 and Figs. 6–7, WRF-CMAQ underestimated annual and seasonal (except for autumn) PM$_{2.5}$ concentrations with NMBs ranging from $-9.78\%$ to $-6.39\%$ and $-17.68\%$ to $+5.17\%$, respectively. WRF-Chem generated both overestimations and underestimations of PM$_{2.5}$ at the annual and seasonal scales, with related NMBs varying from $-39.11\%$ to $+24.72\%$, respectively. Meanwhile, WRF-CHIMERE excessively overestimated annual and seasonal PM$_{2.5}$ concentrations (NMB: $+19.51\%$ to $+75.47\%$). These biases could be related to different aerosol and gas phase mechanisms, dust and sea salt emission schemes, chemical ICs and BCs, and aerosol size distribution treatments applied in the three two-way coupled models. Based on the differences in NMBs between simulations with ARI and those with no aerosol feedbacks, ARI-induced annual and seasonal NMB variations of WRF-CMAQ_ARI and WRF-Chem_ARI ranged from $+3.01\%$ to $+4.21\%$ and $+3.07\%$ to $+5.02\%$, respectively, indicating that the enabling of ARI feedbacks slightly reduced annual and seasonal (except for autumn) underestimations of PM$_{2.5}$ concentrations. Note that WRF-CHIMERE_ARI further overestimated the annual and seasonal PM$_{2.5}$, with an increase in NMB of up to 10.04%. The increases in PM$_{2.5}$ concentrations caused by ARI effects can be attributed to synergetic decreases in SSR, T2, WS10, and PBLH, and increases in RH2. With ACI feedbacks further enabled, WRF-Chem_BOTH largely underestimated the annual and seasonal PM$_{2.5}$, with NMBs ranging from $-24.15\%$ to $-14.44\%$ compared with WRF-Chem_ARI. WRF-CHIMERE_BOTH tended to decrease ($-2.1\%$ to $-0.51\%$) annual and autumn–winter NMBs, and increase ($+0.35\%$ to $+3.04\%$) spring–summer NMBs. Further comparison between ARI- and ACI-induced NMB variations demonstrates the key point that ARI-induced variations in PM$_{2.5}$ concentrations were smaller than those induced by ACI in WRF-Chem, but this pattern was reversed in WRF-CHIMERE. This may be explained by WRF-CHIMERE incorporating the process of dust aerosols serving as IN, which was not included in WRF-Chem in this study.

For O$_3$, WRF-CHIMERE ($R = 0.62$) exhibited the highest correlation, followed by WRF-CMAQ ($R = 0.55$), and WRF-Chem ($R = 0.45$) (Table 4 and Fig. S16). WRF-CMAQ slightly underestimated annual O$_3$, with NMBs and NGEs of $-12.57\%$ to $-11.52\%$, but WRF-Chem and WRF-CHIMERE both significantly overestimated it, with NMBs of 47.82%–48.10% and 29.46%–29.75%, respectively. The seasonal results of statistical metrics showed patterns that were consistent with annual simulations, and summer O$_3$ pollution levels were better simulated than those in other seasons (Fig. 6). All models with ARI feedbacks enabled resulted in slight decreases in annual and seasonal O$_3$ NMBs and NGEs, ranging from $-3.02\%$ to $+0.85\%$ (the only positive value of $+0.85\%$ was produced by WRF-CMAQ in summer) and from $-1.42\%$ to $-0.75\%$, respectively. Meanwhile, for ACI effects, WRF-Chem and WRF-CHIMERE had
increased annual \( O_3 \) NMBs and NGEs of 0.12\%–0.65\% and 0.40\%–0.55\%, respectively. ACI-induced seasonal NMB variations were different for WRF-Chem compared with WRF-ChIMERE; WRF-Chem increased in spring–summer and decreased in autumn–winter, while WRF-ChIMERE increased in all seasons except for winter (Fig. 7). Such diversity in NMB and NGE variations can be explained by two aspect differences. For model-top boundary conditions, the WRF-CMAQ and WRF-Chem models employed the parameterization scheme of \( O_3 \)-potential vorticity and WRF-ChIMERE used the climatological data from LMDz-INCA. For gas-phase chemistry mechanisms, three coupled models incorporate a variety of photolytic reactions, with a more comprehensive explanation provided in Section 4.2.

A comprehensive assessment of the effects of seven gas-phase chemical mechanisms (RADM2, RADMKA, RACM-ESRL, CB05Clx, CB05-TUCL, CBMZ, and MOZART-4) on \( O_3 \) simulations via three two-way coupled models (WRF-Chem, WRF-CMAQ, and COSMO-ART) was conducted by Knote et al. (2015); they concluded that the \( O_3 \) concentrations simulated via WRF-Chem with the CBMZ mechanism were closest to the mean values of multiple models over North America and Europe in spring and summer. However, in contrast to North America and Europe, the two-way coupled WRF-Chem with CBMZ had the poorest performance during spring in eastern China. In addition, ARI and/or ACI effects contribute to atmospheric dynamics and stability (as mentioned in the PBLH evaluation part of Section 1.1 in Supplment), as well as photochemistry and heterogeneous reactions, and, in turn, they will eventually influence \( O_3 \) formation (Xing et al., 2017; Qu et al., 2021; Zhu et al., 2021).

Table 4. Statistical metrics (R, MB, NMB, NGE, and RMSE) between annual simulations and observations of surface PM\(_{2.5}\), \( O_3 \), NO\(_2\), SO\(_2\), and CO in eastern China. The best results are in bold, while mean simulations and observations are in italics.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Statistics</th>
<th>WRF-CMAQ_NO</th>
<th>WRF-CMAQ_ARI</th>
<th>WRF-Chem_NO</th>
<th>WRF-Chem_ARI</th>
<th>WRF-Chem_BOTH</th>
<th>WRF-ChIMERE_NO</th>
<th>WRF-ChIMERE_ARI</th>
<th>WRF-ChIMERE_BOTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM(_{2.5}) (44.99 ( \mu g/m^3))</td>
<td>Mean_sim</td>
<td>40.59</td>
<td>42.12</td>
<td>44.45</td>
<td>46.65</td>
<td>38.33</td>
<td>62.17</td>
<td>65.36</td>
<td>65.13</td>
</tr>
<tr>
<td>R</td>
<td>0.68</td>
<td>0.68</td>
<td>0.65</td>
<td>0.65</td>
<td>0.69</td>
<td>0.52</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
</tr>
<tr>
<td>MB</td>
<td>-4.40</td>
<td>-2.87</td>
<td>-0.54</td>
<td>1.66</td>
<td>-6.66</td>
<td>17.18</td>
<td>20.37</td>
<td>20.14</td>
<td>44.76</td>
</tr>
<tr>
<td>NGE (%)</td>
<td>-9.78</td>
<td>-6.39</td>
<td>-1.21</td>
<td>3.69</td>
<td>-14.81</td>
<td>38.19</td>
<td>45.27</td>
<td>44.76</td>
<td>44.76</td>
</tr>
<tr>
<td>RMSE</td>
<td>27.62</td>
<td>27.69</td>
<td>32.58</td>
<td>34.64</td>
<td>32.48</td>
<td>55.13</td>
<td>60.25</td>
<td>59.41</td>
<td>59.41</td>
</tr>
<tr>
<td>O(_3) (62.23 ( \mu g/m^3))</td>
<td>Mean_sim</td>
<td>55.06</td>
<td>54.41</td>
<td>88.53</td>
<td>87.81</td>
<td>87.89</td>
<td>76.92</td>
<td>76.48</td>
<td>76.89</td>
</tr>
<tr>
<td>R</td>
<td>0.54</td>
<td>0.55</td>
<td>0.46</td>
<td>0.45</td>
<td>0.45</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
</tr>
<tr>
<td>NGE (%)</td>
<td>-11.52</td>
<td>-12.57</td>
<td>42.26</td>
<td>41.10</td>
<td>41.22</td>
<td>23.60</td>
<td>22.90</td>
<td>23.55</td>
<td>23.55</td>
</tr>
<tr>
<td>RMSE</td>
<td>28.32</td>
<td>28.68</td>
<td>48.10</td>
<td>47.99</td>
<td>47.82</td>
<td>29.65</td>
<td>29.46</td>
<td>29.75</td>
<td>29.75</td>
</tr>
<tr>
<td>NO(_2) (31.2 ( \mu g/m^3))</td>
<td>Mean_sim</td>
<td>33.94</td>
<td>34.46</td>
<td>21.17</td>
<td>21.98</td>
<td>21.40</td>
<td>21.85</td>
<td>22.20</td>
<td>22.24</td>
</tr>
<tr>
<td>R</td>
<td>0.59</td>
<td>0.60</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.55</td>
<td>0.56</td>
<td>0.56</td>
<td>0.56</td>
</tr>
<tr>
<td>MB</td>
<td>2.74</td>
<td>3.26</td>
<td>-10.03</td>
<td>-9.22</td>
<td>-9.80</td>
<td>-9.35</td>
<td>-9.00</td>
<td>-8.96</td>
<td>-8.96</td>
</tr>
</tbody>
</table>
Figure 6. Time series of observed and simulated hourly PM$_{2.5}$ and O$_3$ concentrations by WRF-CMAQ, WRF-Chem and WRF-CHIMERE with/without aerosol feedbacks over Eastern China during the year of 2017.
Figure 7. Taylor diagrams (R, normalized standard deviation, and NMB) of seasonal PM$_{2.5}$, O$_3$, NO$_2$, SO$_2$, and CO via three two-way coupled models (WRF-CMAQ, WRF-Chem, and WRF-CHIMERE) with/without ARI and/or ACI effects in eastern China compared with surface observations.

In a similar manner to the meteorological variables presented above, we aimed to conduct quality assurance for the statistical metrics by making further comparisons with PM$_{2.5}$ and O$_3$ results from previous model evaluations (summarized in Fig. S20). The performances of WRF-CMAQ and WRF-Chem in simulating PM$_{2.5}$ in this study were better than the average levels of previous studies from eastern China. For O$_3$, WRF-Chem simulations performed worse than the average level of previous studies. Although the R values of O$_3$ simulated by WRF-CMAQ in this study were lower than the average level of previous studies, the RMSEs in this study were smaller.

4.2 Satellite-borne observations

In this section, we further investigate the discrepancies among different models in terms of the calculated AOD and column concentrations of gases (O$_3$, NO$_2$, SO$_2$, CO, and NH$_3$), and compare them with various satellite observations. For NH$_3$, owing to not setting the output of simulated NH$_3$ concentrations in WRF-CHIMERE, the discussion here only includes the results from WRF-CMAQ and WRF-Chem.

As shown in Table 5, annual AOD at 550 nm, TCO, NO$_2$, and CO simulated by all
three models agreed most closely with satellite observations, with correlation coefficients of 0.80–0.98; these were followed by NH$_3$ (0.75–0.76), and SO$_2$ (0.50–0.53). WRF-CMAQ presented negative biases for annual AOD (~0.01), TCO (~5.92 Dobson Units (DU)), SO$_2$ (~0.03 to ~0.02 DU), CO (~1.25 × 10$^{17}$ molecules cm$^{-2}$), and NH$_3$ (~2.95 × 10$^{15}$ molecules cm$^{-2}$), but a positive bias for NO$_2$ (1.09–1.21 petamolecules cm$^{-2}$). For AOD, WRF-Chem and WRF-CHIMERE produced positive and negative MBs of +0.09 and ~0.06, respectively. Both WRF-Chem and WRF-CHIMERE overestimated NO$_2$ (0.28–0.63 petamolecules cm$^{-2}$) and CO (0.93–1.21 × 10$^{17}$ molecules cm$^{-2}$), and underestimated O$_3$ (-10.99 to -3.63 DU) and SO$_2$ (-0.03 to -0.02 DU). Similar to WRF-CMAQ, WRF-Chem also underestimated NH$_3$ by approximately −3.14 × 10$^{15}$ molecules cm$^{-2}$.

For seasonal variations, relatively high correlation relationships (0.71–0.88) of AOD were present in autumn, with lower values (0.53–0.84) in other seasons (Fig. 8). WRF-CMAQ and WRF-Chem tended to underestimate AOD in summer (MBs of ~0.1 to ~0.4) and overestimate it in other seasons (MBs of 0.01–0.05). WRF-CHIMERE had positive biases (0.03–0.04) in winter and negative biases (~0.10 to ~0.01) in other seasons. For TCO (Fig. S24), the model performances of WRF-CMAQ and WRF-Chem in spring and winter were slightly better than those in summer and autumn, but all seasonal R values were greater than 0.89. Both WRF-CMAQ (~9.53 to ~0.72 DU) and WRF-Chem (~24.62 to +10.57 DU) had negative biases in all seasons (note: WRF-Chem except for autumn). WRF-CHIMERE was better at capturing TCO in spring and summer (overestimations of +9.19 to +29.20 DU) than in autumn and winter (underestimations of ~33.75 to ~19.40 DU). The R values of NO$_2$ columns for all three models were slightly higher in autumn and winter (0.82–0.91) than in spring and summer (0.76–0.84). The seasonal NO$_2$ columns were generally underestimated in WRF-CMAQ (~0.68 to -0.16 DU), WRF-Chem (~1.40 to -0.44 DU), WRF-CHIMERE (~1.31 to -0.19 DU) (Fig. S22). All models overestimated SO$_2$ column concentrations in winter (by approximately 0.01–0.03 DU) but underestimated them in other seasons (~0.05 to ~0.001 DU) (Fig. S23). For NH$_3$, the only primary alkaline gas in the atmosphere, better model performances of WRF-CMAQ and WRF-Chem occurred in summer (R: 0.81–0.87; MB: ~3.42 to 2.07 × 10$^{15}$ molecules cm$^{-2}$) (Fig. S25). Ammonia emissions from fertilizer and livestock have been substantially underestimated in China (Zhang et al., 2017), and peak values occur in spring and summer (Huang et al., 2012).

In addition, bidirectional exchanges of fertilizer-induced NH$_3$ were not considered in our simulations. Compared to above column variables, WRF-CMAQ, WRF-Chem, and WRF-CHIMERE showed relatively poor performances (R: 0.68–0.79) in simulating CO columns during spring, summer, and autumn, respectively, compared with other seasons (Fig. S24). WRF-CMAQ and WRF-CHIMERE respectively underestimated and overestimated CO columns in other seasons except for summer and spring, with MBs of −3.29 to 0.31 × 10$^{17}$ and −0.62 to 2.09 × 10$^{17}$ molecules cm$^{-2}$, respectively. WRF-Chem had positive MBs in summer and autumn (4.03–5.12 × 10$^{17}$ molecules cm$^{-2}$) and negative MBs in spring and winter (~3.15 to ~2.10 × 10$^{17}$ molecules cm$^{-2}$).

Moreover, after comparing the performance results for each pollutant between sections 4.1 and 4.2, the only disparity found between evaluations with ground-based
observations compared with those with satellite-borne observations was for CO. The formation of CO via the oxidation of methane, an important source of CO emissions (Stein et al., 2014), is not considered in the three coupled models, and methane emissions are not included in the MEIC inventory. In addition, the contribution of CO to atmospheric oxidation capacity (OH radicals) was non-negligible (e.g., values were approximately 20.54%–38.97% in Beijing (Liu et al., 2021), and 26%–31% in Shanghai (Zhu et al., 2020). Also, these discrepancies in the model performances for simulating AOD and column concentrations of gases can be explained by differences in the representations of aerosol species groups, Fast-JX photolysis scheme, and gas-phase mechanisms in the three coupled models. More detailed interpretations were grouped into four aspects: (1) AODs are calculated via Mie theory using refractive indices of different numbers (5, 6 and 10) of aerosol species group in different coupled models (WRF-CMAQ, WRF-Chem and WRF-CHIMERE) (Tables S5–S6); (2) 7 (294.6, 303.2, 310.0, 316.4, 333.1, 382.0 and 607.7 nm), 4 (300, 400, 600 and 999 nm), and 5 (200, 300, 400, 600, and 999 nm) effective wavelengths are used in calculating actinic fluxes and photolysis rates in Fast-JX photolysis modules of WRF-CMAQ, WRF-Chem and WRF-CHIMERE, respectively; (3) Different calculating methods of aerosol and cloud optical properties exist in the Fast-JX schemes of three coupled models (Tables S1 and S5–S6); (4) 77, 52 and 40 gas-phase species involve 218, 132, 120 gas-phase reactions in CB6, CBMZ and MELCHIOR2 mechanisms, respectively.

When all three models enabled just ARI effects, improvements in annual AOD and NO2 columns simulated by these models were relatively limited. The AOD simulations improved in spring and summer, but worsened in autumn and winter (Table 4 and Fig. 9). Larger variations in seasonal MBs of NO2 columns induced by ARI effects occurred in WRF-CMAQ (−0.18 to 0.13 petamolecules cm⁻²) compared with WRF-Chem and WRF-CHIMERE (0–0.01 petamolecules cm⁻²). When both ARI and ACI effects were enabled in WRF-Chem, the model performance for seasonal AOD simulations worsened considerably. The annual and seasonal NO2 simulations via WRF-Chem became slightly worse, while those using WRF-CHIMERE became slightly better. In contrast to AOD and NO2 column concentrations, improvements in annual and seasonal column simulations of total ozone, PBL SO2, and NH3 via all two-way coupled models were limited when one or both of ARI and ACI were enabled.

Table 5. Statistical metrics (R, MB, NMB, NGE, and RMSE) of simulated and satellite-retrieved AOD, total column ozone, tropospheric column NO2, PBL column SO2, total column CO, and total column density of NH3 in eastern China. The best results are in bold, while annual mean simulations and observations are in italics.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Statistics</th>
<th>WRF-CMAQ_NO</th>
<th>WRF-CMAQ_ARI</th>
<th>WRF-Chem_NO</th>
<th>WRF-Chem_ARI</th>
<th>WRF-Chem_BOTH</th>
<th>WRF-CHIMERE_NO</th>
<th>WRF-CHIMERE_ARI</th>
<th>WRF-CHIMERE_BOTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOD (0.27)</td>
<td>Mean_sim</td>
<td>0.26</td>
<td>0.27</td>
<td>0.35</td>
<td>0.36</td>
<td>0.25</td>
<td>0.21</td>
<td>0.22</td>
<td>0.22</td>
</tr>
<tr>
<td>R</td>
<td></td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.75</td>
<td><strong>0.87</strong></td>
<td><strong>0.87</strong></td>
<td>0.86</td>
</tr>
<tr>
<td>MB</td>
<td></td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.09</td>
<td>0.09</td>
<td>-0.01</td>
<td>-0.05</td>
<td>-0.05</td>
<td>-0.04</td>
</tr>
<tr>
<td>NMB (%)</td>
<td></td>
<td>-3.99</td>
<td><strong>-2.93</strong></td>
<td>34.14</td>
<td>35.03</td>
<td>-4.92</td>
<td>-18.72</td>
<td>-17.37</td>
<td>-16.22</td>
</tr>
<tr>
<td>NGE (%)</td>
<td></td>
<td>34.90</td>
<td>34.82</td>
<td>58.21</td>
<td>58.89</td>
<td><strong>41.46</strong></td>
<td><strong>32.11</strong></td>
<td><strong>32.06</strong></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td></td>
<td><strong>0.09</strong></td>
<td><strong>0.09</strong></td>
<td>0.15</td>
<td>0.15</td>
<td><strong>0.10</strong></td>
<td><strong>0.09</strong></td>
<td><strong>0.09</strong></td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>Mean_sim</td>
<td>306.15</td>
<td>306.15</td>
<td>300.77</td>
<td>300.73</td>
<td>300.46</td>
<td>307.69</td>
<td>307.47</td>
<td>307.75</td>
</tr>
<tr>
<td>----</td>
<td>----------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>O$_3$ VCDs (312.07 DU)</td>
<td>R</td>
<td>0.98</td>
<td>0.98</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
<td>0.65</td>
<td>0.65</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>MB</td>
<td>-5.92</td>
<td>-5.92</td>
<td>-10.68</td>
<td>-10.72</td>
<td>-10.99</td>
<td>-3.69</td>
<td>-3.91</td>
<td>-3.63</td>
</tr>
<tr>
<td></td>
<td>NMB (%)</td>
<td>-1.90</td>
<td>-1.90</td>
<td>-3.43</td>
<td>-3.44</td>
<td>-3.53</td>
<td>-1.19</td>
<td>-1.26</td>
<td>-1.17</td>
</tr>
<tr>
<td></td>
<td>NGE (%)</td>
<td>2.46</td>
<td>2.46</td>
<td>25.02</td>
<td>25.02</td>
<td>25.08</td>
<td>10.95</td>
<td>10.89</td>
<td>10.93</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>8.91</td>
<td>8.91</td>
<td>83.72</td>
<td>83.73</td>
<td>83.94</td>
<td>39.88</td>
<td>39.71</td>
<td>39.73</td>
</tr>
<tr>
<td>Tropospheric NO$_2$ VCDs (2.71×10$^{15}$ molecules cm$^{-2}$)</td>
<td>Mean_sim</td>
<td>3.80</td>
<td>3.91</td>
<td>3.07</td>
<td>3.08</td>
<td>3.06</td>
<td>2.62</td>
<td>2.63</td>
<td>2.63</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>0.85</td>
<td>0.85</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>MB</td>
<td>1.09</td>
<td>1.21</td>
<td>0.62</td>
<td>0.63</td>
<td>0.61</td>
<td>0.28</td>
<td>0.29</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>NMB (%)</td>
<td>40.35</td>
<td>44.64</td>
<td>25.27</td>
<td>25.52</td>
<td>24.89</td>
<td>12.03</td>
<td>12.47</td>
<td>12.42</td>
</tr>
<tr>
<td></td>
<td>NGE (%)</td>
<td>52.80</td>
<td>55.08</td>
<td>46.01</td>
<td>46.05</td>
<td>45.17</td>
<td>46.06</td>
<td>46.31</td>
<td>46.24</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>3.18</td>
<td>3.33</td>
<td>2.27</td>
<td>2.27</td>
<td>2.27</td>
<td>1.65</td>
<td>1.67</td>
<td>1.68</td>
</tr>
<tr>
<td>PBL SO$_2$ VCDs (0.09 DU)</td>
<td>Mean_sim</td>
<td>0.07</td>
<td>0.07</td>
<td>0.09</td>
<td>0.09</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>0.53</td>
<td>0.53</td>
<td>0.56</td>
<td>0.56</td>
<td>0.54</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>MB</td>
<td>-0.03</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.02</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>NMB (%)</td>
<td>27.32</td>
<td>25.48</td>
<td>32.50</td>
<td>21.50</td>
<td>35.08</td>
<td>28.64</td>
<td>27.31</td>
<td>27.51</td>
</tr>
<tr>
<td></td>
<td>NGE (%)</td>
<td>57.45</td>
<td>58.26</td>
<td>67.55</td>
<td>64.83</td>
<td>68.31</td>
<td>68.61</td>
<td>68.80</td>
<td>68.80</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.07</td>
<td>0.07</td>
<td>0.08</td>
<td>0.08</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>Total CO VCDs (21.60×10$^{17}$ molecules cm$^{-2}$)</td>
<td>Mean_sim</td>
<td>20.34</td>
<td>20.35</td>
<td>22.20</td>
<td>22.20</td>
<td>22.21</td>
<td>22.34</td>
<td>22.36</td>
<td>22.35</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>0.83</td>
<td>0.83</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>MB</td>
<td>-1.26</td>
<td>-1.24</td>
<td>0.93</td>
<td>0.93</td>
<td>0.94</td>
<td>1.19</td>
<td>1.21</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td>NMB (%)</td>
<td>-5.83</td>
<td>-5.75</td>
<td>4.35</td>
<td>4.37</td>
<td>4.44</td>
<td>5.64</td>
<td>5.70</td>
<td>5.65</td>
</tr>
<tr>
<td></td>
<td>NGE (%)</td>
<td>9.33</td>
<td>9.31</td>
<td>10.30</td>
<td>10.28</td>
<td>10.32</td>
<td>11.02</td>
<td>11.06</td>
<td>11.10</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>2.54</td>
<td>2.54</td>
<td>2.69</td>
<td>2.68</td>
<td>2.69</td>
<td>2.57</td>
<td>2.58</td>
<td>2.58</td>
</tr>
<tr>
<td>Total NH$_3$ VCDs (16.05×10$^{15}$ molecules cm$^{-2}$)</td>
<td>Mean_sim</td>
<td>13.06</td>
<td>13.15</td>
<td>12.31</td>
<td>12.27</td>
<td>8.63</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>0.76</td>
<td>0.76</td>
<td>0.73</td>
<td>0.73</td>
<td>0.76</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>MB</td>
<td>-3.00</td>
<td>-2.90</td>
<td>-3.27</td>
<td>-3.32</td>
<td>-3.34</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>NMB (%)</td>
<td>-18.66</td>
<td>-18.08</td>
<td>-21.01</td>
<td>-21.28</td>
<td>-21.41</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>NGE (%)</td>
<td>47.69</td>
<td>48.09</td>
<td>50.84</td>
<td>50.80</td>
<td>50.99</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

NA indicates that outputs of NH$_3$ column concentrations were not extracted from WRF-CHIMERE with/without aerosol feedback simulations.
Figure 8. Spatial distributions of seasonal AOD between MODIS observations and simulations from WRF-CMAQ, WRF-Chem, and WRF-CHIMERE with and without aerosol feedbacks in eastern China.

4.3 Computational performance

Table 5 summarizes the comparative results of central processing unit (CPU) time consumption for one day simulations via WRF-CMAQ, WRF-Chem, and WRF-CHIMERE with and without aerosol feedbacks in 2017. The results show that regardless of whether aerosol feedbacks were enabled, the CPU time consumed by WRF-CMAQ simulating one-day meteorology and air quality was shortest, followed by WRF-CHIMERE, and WRF-Chem. Compared with simulations without aerosol feedbacks, the processing time of WRF-CMAQ with ARI enabled increased by 0.22–0.34 hours per day, while increases in the running time of WRF-Chem and WRF-CHIMERE were not significant (0.02–0.03 hours per day). The CPU time for both WRF-Chem and WRF-CHIMERE with both ARI and ACI effects enabled was slightly increased, and the increase in CPU time for the former (0.25 hours per day) was larger than that for the latter (0.11 hours per day). Compared with WRF-CMAQ and WRF-Chem, the CPU time of WRF-CHIMERE showed obvious seasonal differences, with the time in winter and spring being significantly longer than that in summer and autumn. These differences can be partially explained by the choice of main configurations, including model resolution, model version, and parametrization schemes (cloud microphysics, PBL, cumulus, surface layer, land surface, gas-phase chemistry, and aerosol mechanisms).
Table 5. Summary of running time for different coupled models.

<table>
<thead>
<tr>
<th>Month</th>
<th>WRF-CMAQ (hour)</th>
<th>WRF-Chem (hour)</th>
<th>WRF-CHIMERE (hour)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NO</td>
<td>ARI</td>
<td>NO</td>
</tr>
<tr>
<td>Jan.</td>
<td>0.37</td>
<td>0.59</td>
<td>0.69</td>
</tr>
<tr>
<td>Feb.</td>
<td>0.35</td>
<td>0.60</td>
<td>0.68</td>
</tr>
<tr>
<td>Mar.</td>
<td>0.39</td>
<td>0.65</td>
<td>0.70</td>
</tr>
<tr>
<td>Apr.</td>
<td>0.37</td>
<td>0.67</td>
<td>0.67</td>
</tr>
<tr>
<td>May</td>
<td>0.39</td>
<td>0.71</td>
<td>0.61</td>
</tr>
<tr>
<td>June</td>
<td>0.40</td>
<td>0.74</td>
<td>0.66</td>
</tr>
<tr>
<td>July</td>
<td>0.36</td>
<td>0.69</td>
<td>0.65</td>
</tr>
<tr>
<td>Aug.</td>
<td>0.38</td>
<td>0.68</td>
<td>0.66</td>
</tr>
<tr>
<td>Sept.</td>
<td>0.37</td>
<td>0.63</td>
<td>0.64</td>
</tr>
<tr>
<td>Oct.</td>
<td>0.38</td>
<td>0.62</td>
<td>0.66</td>
</tr>
<tr>
<td>Nov.</td>
<td>0.36</td>
<td>0.58</td>
<td>0.68</td>
</tr>
<tr>
<td>Dec.</td>
<td>0.35</td>
<td>0.57</td>
<td>0.63</td>
</tr>
</tbody>
</table>

5 Conclusions

Applications of two-way coupled meteorology and air quality models have been performed in eastern China in recent years, but no research focused on the comprehensive assessments of multiple coupled models in this region. To the best of our knowledge, this is the first time to conduct comprehensive inter-comparisons among the open-sourced two-way coupled meteorology and air quality models (WRF-CMAQ, WRF-Chem, and WRF-CHIMERE). This study systemically evaluated the hindcast simulations for 2017 and explored the impacts of ARI and/or ACI on model and computational performances in eastern China.

After detailed comparisons with ground-based and satellite-borne observations, the evaluation results showed that three coupled models perform well for meteorology and air quality, especially for surface temperature (with R up to 0.97) and PM$_{2.5}$ concentrations (with R up to 0.68). The effects of aerosol feedbacks on model performances varied depending on the two-way coupled models, variables, and time scales. There were around 20%–70% increase of computational time when these two-way coupled models enabled aerosol feedbacks against simulations without aerosol-radiation-cloud interactions. It is noteworthy that all three coupled models could well reproduce the spatiotemporal distributions of satellite-retrieved CO column concentrations but not for ground-observed CO concentrations.

With inter-comparisons, some uncertainty sources can be ascertained in evaluating aerosol feedback effects. As numerous schemes can be combined in configurations of different coupled models, here we only evaluated simulations with specific settings. Future comparison works with considering more combinations of multiple schemes within the same or different coupled models need to be conducted. Among the three coupled models, the numerical representations for specific variable in same scheme are diverse, e.g., treatments of cloud cover and cloud optical properties in the Fast-JX photolysis scheme. More accurate representations of photolysis
processes should be taken into account to reduce the evaluation uncertainties. In addition, FDDA nudging technique can attenuate the ARI effects during severe air polluted episodes, and optimal nudging coefficients among different regions need to be determined. Last but not least, the actual mechanisms underlying ACI effects are still unclear, and the new advances in the measurements and parameterizations of CCN/IN activations and precipitation need to be timely incorporated in coupled models.

Code availability

The source codes of the two-way coupled WRF v4.1.1-CMAQ v5.3.1, WRF-Chem v4.1.1, and WRF v3.7.1-CHIMERE v2020r1 models are obtained from https://github.com/USEPA/CMAQ, https://github.com/wrf-model/WRF, and https://www.lmd.polytechnique.fr/chimere, respectively (last access: November 2020). The related source codes, configuration information, namelist files and automated run scripts of these three two-way coupled models are archived at Zenodo with the associated DOI: https://doi.org/10.5281/zenodo.7901682 (Gao et al., 2023a; link: https://zenodo.org/record/7901682).

Data availability

The meteorological ICs and BCs used for three coupled models can be obtained at https://doi.org/10.5281/zenodo.7925012 (Gao et al., 2023b; link: https://zenodo.org/record/7925012). The Chemical ICs and BCs used for WRF-CMAQ, WRF-Chem and WRF-CHIMERE are available at https://doi.org/10.5281/zenodo.7932390 (Gao et al., 2023c; link: https://zenodo.org/record/7932390), https://doi.org/10.5281/zenodo.7932936 (Gao et al., 2023d; link: https://zenodo.org/record/7932936), and https://doi.org/10.5281/zenodo.7933641 (Gao et al., 2023e; link: https://zenodo.org/record/7933641), respectively. The emission data used for WRF-CMAQ, WRF-Chem and WRF-CHIMERE can be downloaded from https://doi.org/10.5281/zenodo.7932430 (Gao et al., 2023f; link: https://zenodo.org/record/7932430), https://doi.org/10.5281/zenodo.7932734 (Gao et al., 2023g; link: https://zenodo.org/record/7932734), and https://doi.org/10.5281/zenodo.7931614 (Gao et al., 2023h; link: https://zenodo.org/record/7931614), respectively. The DOIs and links regarding the output data of each simulation scenario are presented in Table S9. All data used to create figures and tables in this study are provided in an open repository on Zenodo (https://doi.org/10.5281/zenodo.7750907, Gao et al., 2023i; link: https://zenodo.org/record/7750907).

Author contributions

CG, ZX, AX performed the majority of the source code configuration of WRF-CMAQ, WRF-Chem and WRF-CHIMERE, designed the numerical simulations to carry them out, related analysis, figure plotting, and paper writing. QT, HZ, SZ, GY, MZ and XS were involved with the original research plan and made suggestions for the
paper writing.

Competing interests

The contact author has declared that neither they nor their co-authors have any competing interests.

Acknowledgements

The authors are very grateful to David Wong, Chun Zhao and Laurent Menut who provided detailed information on the two-way coupled WRF-CMAQ, WRF-Chem and WRF-CHIMERE models, respectively.

Financial support

This study was financially sponsored by the Youth Innovation Promotion Association of Chinese Academy of Sciences, China (grant nos. 20222230), the National Key Research and Development Program of China (grant nos. 2017YFC0212304 & 2019YFE0194500), the Talent Program of Chinese Academy of Sciences (Y8H1021001), and the National Natural Science Foundation of China (grant nos. 42171142 & 41771071).

References


Briant, R., Tuccella, P., Deroubaix, A., Khvorostyanov, D., Menut, L., Mailler, S., and Turquety, S.: Aerosol-radiation interaction modelling using online coupling between the WRF 3.7.1...


Gao, C., Xiu, A., Zhang, X., Tong, Q., Zhao, H., Zhang, S., Yang, G., Zhang, M., Xie, S.: Data used to create figures and tables in the GMD manuscript "Inter-comparison of multiple two-way coupled meteorology and air quality models (WRF v4.1.1-CMAQ v5.3.1, WRF-Chem v4.1.1 and WRF v3.7.1-CHIMERE v2020r1) in eastern China", Zenodo [data set], https://doi.org/10.5281/zenodo.7750907, 2023i.


Gao, M., Han, Z., Tao, Z., Li, J., Kang, J.-E., Huang, K., Dong, X., Zhuang, B., Li, S., and Ge, B.:


Jacobson, M. Z.: Developing, coupling, and applying a gas, aerosol, transport, and radiation model to study urban and regional air pollution, 1994.


