Development of Korean Air Quality Prediction System version 1 (KAQPS v1) with focuses on practical issues

Kyunghwa Lee[1,2], Jinhyeok Yu[2], Sojin Lee[3], Mieun Park[4,5], Hun Hong[2], Soon Young Park[2], Myungje Choi[6], Jhoon Kim[6], Younha Kim[7], Jung-Hun Woo[7], Sang-Woo Kim[8] and Chul H. Song[2*]

1. Environmental Satellite Center, Climate and Air Quality Research Department, National Institute of Environmental Research (NIER), Incheon, Republic of Korea
2. School of Earth Sciences and Environmental Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju, Republic of Korea
3. Department of Earth and Atmospheric Sciences, University of Houston, Texas, USA
4. Air Quality Forecasting Center, Climate and Air Quality Research Department, National Institute of Environmemnbntal Research (NIER), Incheon, Republic of Korea
5. Environmental Meteorology Research Division, National Institute of Meteorological Sciences (NIMS), Jeju, Republic of Korea
6. Department of Atmospheric Sciences, Yonsei University, Seoul, Republic of Korea
7. Department of Advanced Technology Fusion, Konkuk University, Seoul, Republic of Korea
8. School of Earth and Environmental Sciences, Seoul National University, Seoul, Republic of Korea

Short title: Air quality prediction system in Korea

Corresponding author: Chul H. Song (chsong@gist.ac.kr)

Abstract

29   For the purpose of providing reliable and robust air quality predictions, an air quality

30   prediction system was developed for the main air quality criteria species in South Korea (PM_{10},

31    $PM_{2.5}$, CO, O_3, and SO_2). The main caveat of the system is to prepare the initial conditions (ICs)

32   of the Community Multi-scale Air Quality (CMAQ) model simulations using observations

33   from the Geostationary Ocean Color Imager (GOCI) and ground-based monitoring networks

34   in northeast Asia. The performance of the air quality prediction system was evaluated during

35   the Korea-United States Air Quality Study (KORUS-AQ) campaign period (1 May–12 June

36   2016). Data assimilation (DA) of optimal interpolation (OI) with Kalman filter was used in this

37   study. One major advantage of the system is that it can predict not only particulate matter (PM)

38   concentrations but also PM chemical composition including five main constituents: sulfate

39   (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+), organic aerosols (OAs), and elemental carbon (EC).

40   In addition, it is also capable of predicting the concentrations of gaseous pollutants (CO, O_3

41   and SO_2). In this sense, this new air quality prediction system is comprehensive. The results

42   with the ICs (DA RUN) were compared with those of the CMAQ simulations without ICs

43   (BASE RUN). For almost all of the species, the application of ICs led to improved performance

44   in terms of correlation, errors, and biases over the entire campaign period. The DA RUN agreed

45   reasonably well with the observations for PM_{10} (IOA = 0.60; MB = -13.54) and $PM_{2.5}$ (IOA =

46   0.71; MB = -2.43) as compared to the BASE RUN for PM_{10} (IOA = 0.51; MB = -27.18) and

47    $PM_{2.5}$ (IOA = 0.67; MB = -9.9). A significant improvement was also found with the DA RUN

48   in terms of bias. For example, for CO, the MB of -0.27 (BASE RUN) was greatly enhanced to

49   -0.036 (DA RUN). In the cases of O_3 and SO_2, the DA RUN also showed better performance

50   than the BASE RUN. Further, several more practical issues frequently encountered in the air

quality prediction system were also discussed. In order to attain more accurate ozone

predictions, the DA of NO_2 mixing ratios should be implemented with careful consideration of

the measurement artifacts (i.e., inclusion of alkyl nitrates, HNO_3, and PANs in the ground-

observed NO_2 mixing ratios). It was also discussed that, in order to ensure accurate nocturnal

predictions of the concentrations of the ambient species, accurate predictions of the mixing

layer heights (MLH) should be achieved from the meteorological modeling. Several

advantages of the current air quality prediction system, such as its non-static free parameter

scheme, dust episode prediction, and possible multiple implementations of DA prior to actual

predictions, were also discussed. These configurations are all possible because the current DA

system is not computationally expensive. In the ongoing and future works, more advanced DA

techniques such as the three-dimensional variational (3DVAR) method and ensemble Kalman

filter (EnK) are being tested and will be introduced to the Korean air quality prediction system

(KAQPS).

Keywords: Air quality prediction; Particulate matter (PM); Geostationary satellite sensor
(GOCI); Air Korea; Data assimilation (DA); Dust episode predictions; NO_2 measurement
artifacts

1. Introduction

Air quality has long been considered an important issue in climate change, visibility, and public health, and it is strongly dependent upon meteorological conditions, emissions, and the transport of air pollutants. Air pollutants typically consist of atmospheric particles and gases such as particulate matter (PM), carbon monoxide (CO), ozone (O_3), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2). These aerosols and gases play important roles in anthropogenic climate forcing both directly (Bellouin et al., 2005; Carmichael et al., 2009; IPCC, 2013; Scott et al., 2014) and indirectly (Bréon et al., 2002; IPCC, 2013; Penner et al., 2004; Scott et al., 2014) in influencing the global radiation budget. Among the various air pollutants, PM and surface O_3 are the most notorious health threats, as has been stated by several previous studies (Carmichael et al., 2009; Dehghani et al., 2017; Khaniabadi et al., 2017).

With the stated importance of atmospheric aerosols and gases, considerable research efforts have been made to monitor and quantify their amounts in the atmosphere through satellite-, airborne-, and ground-based observations as well as chemistry-transport model (CTM) simulations. In South Korea, the Korean Ministry of the Environment (KMoE) provides real-time chemical concentrations as measured by ground-based observations for six criteria air pollutants (PM_{10}, $PM_{2.5}$, O_3, CO, SO_2, and NO_2) at the Air Korea website (https://www.airkorea.or.kr). In addition, the National Institute of Environmental Research (NIER) of South Korea provides air quality predictions using multiple CTM simulations. Air quality predictions are another crucial element for protecting public health through the forecasting of high air pollution episodes in advance and alerting citizens about these high episodes. In this context, reliable and robust air quality forecasts are necessary to avoid any confusion caused by poor predictions given by CTM simulations.

91   Although there are various datasets representing air quality, limitations remain in the

92   observations and model outputs. Specifically, observation data are, in general, known to be

93   more accurate than model outputs, but they have spatial and temporal limitations. These

94   limitations will be overcome by improving spatial and temporal coverage via future

95   geostationary satellite instruments such as the Geostationary Environment Monitoring

96   Spectrometer (GEMS) over Asia, the Tropospheric Emissions: Monitoring of Pollution

97   (TEMPO) over North America, and the Sentinel-4 over Europe. In addition, the TROPOspheric

98   Monitoring Instrument (TROPOMI) on board the Copernicus Sentinel-5 Precursor satellite was

99   successfully launched into low earth orbit (LEO) on 13 October 2017 and are providing

100   information on the chemical composition in the atmosphere with a higher spatial resolution of

101   $3.5 \times 7 \text{ km}^2$.

102   Unlike observation data, models can provide meteorological and chemical information

103   without any spatial and temporal data discontinuity, but they do have an issue of inaccuracy.

104   The major causes of uncertainty in the results of CTM simulations are introduced from

105   imperfect emissions, meteorological fields, initial conditions (ICs), and physical and chemical

106   parameterizations in the models (Carmichael et al., 2008). In order to minimize the limitations

107   and maximize the advantages of observation data and model outputs, there have been numerous

108   attempts to provide accurate and spatially- as well as temporally- continuous information on

109   chemical composition in the atmosphere by integrating observation data with model outputs

110   via data assimilation (DA) techniques.

111   Although the Korean numerical weather prediction (NWP) carried out by the Korea

112   Meteorological Administration (KMA) employs various DA techniques, almost no previous

113   efforts have been made to develop a air quality prediction system with DA in South Korea.

Therefore, in the present study, the air quality prediction system named as Korean Air Quality Prediction System version 1 (KAQPS v1) was developed by preparing ICs via DA for the Community Multi-scale Air Quality (CMAQ) model (Byun and Schere, 2006; Byun and Ching, 1999) using satellite- and ground-based observations for particulate matter (PM) and atmospheric gases such as CO, O_3, and SO_2. The performances of the system were then demonstrated during the period of the Korea-United States Air Quality Study (KORUS-AQ) campaign (1 May – 12 June 2016) in South Korea.

In this study, the optimal interpolation (OI) method with the Kalman filter was applied in order to develop the air quality prediction system, since this method is still useful and viable in terms of computational cost and performance. The performance of the method is almost comparable to that of the three-dimensional variational (3DVAR) method, as shown in Tang et al. (2017). More complex and advanced DA techniques are currently being and will continue to be applied to current air quality prediction systems. These works are now in progress.

In addition, this manuscript also discusses several practical issues frequently encountered in the air quality predictions such as: i) DA of NO_2 mixing ratios for accurate ozone prediction with a careful consideration of measurement artifacts; ii) the issue of the nocturnal mixing layer height (MLH) for nocturnal predictions; iii) predictions of dust episodes; iv) the use of non-static free parameters; and v) the influences of multiple implementations of the DA before the actual predictions.

The details of the datasets and methodology used in this study are described in Sect. 2. The results of the developed air quality prediction system are discussed in Sect. 3, and then a summary and conclusions are given in Sect. 4.

2. Methodology

The air quality prediction system was developed using the CMAQ model along with meteorological inputs provided by the Weather Research and Forecasting (WRF) model (Skamarock et al., 2008). The ICs for the CMAQ model simulations were prepared via the DA method using satellite-retrieved and ground-based observations. The performances of the developed prediction system were evaluated using ground in-situ data. The models, data, and DA technique are described in detail in the following sections.

2.1 Meteorological and chemistry-transport modeling

2.1.1 WRF model simulations

The WRF model has been developed for providing mesoscale numerical weather prediction (NWP). It has also been used to provide meteorological input fields for CTM simulations (Appel et al., 2010; Chemel et al., 2010; Foley et al., 2010; Lee et al., 2016; Park et al., 2014). In this study, WRF v3.8.1 with the Advanced Research WRF (ARW) dynamical core was applied to prepare the meteorological inputs for the CMAQ model simulations. Dynamical and physical configurations for the WRF model simulations were selected as follows: the Yonsei University (YSU) scheme for planetary boundary layer (Hong et al., 2006); the WRF single-moment 6-class (WSM6) scheme for the micro-physics (Hong and Lim, 2006); the Grell-Freitas ensemble scheme for cumulus parameterization (Grell and Freitas, 2014); the Noah-MP land surface model (Niu et al., 2011; Yang et al., 2011); the rapid radiative transfer model for Global Circulation Models (RRTMG) for shortwave/longwave options (Iacono et al., 2008); and the revised MM5 scheme for surface layer options (Jiménez et al., 2012). The National Centers for Environmental Prediction (NCEP) Final (FNL) Operational Global

159 Analysis data on 1° × 1° grids were chosen for the ICs and boundary conditions (BCs) for the

160 WRF simulations. In order to minimize meteorological field errors for the applications of ICs

161 and BCs to the WRF simulations, the objective analysis (OBSGRID) nudging was conducted

162 using the NCEP Automated Data Processing (ADP) global upper-air/surface observational

163 weather data via the Cressman (1959)'s successive correction method. The adjusted

164 meteorological variables were temperature, geopotential height, relative humidity, and

165 zonal/meridional winds.

166 The model domain for the WRF simulations covers Northeast Asia with a horizontal

167 resolution of 15×15 km^2, having a total of 223 latitudinal and 292 longitudinal grid cells. The

168 size of the WRF domain is slightly larger than that of the CMAQ domain, as shown in Fig. 1.

169 The meteorological data have 27 vertical layers from the surface (1000 hPa) to 50 hPa. The

170 WRF meteorological fields (e.g., temperature, pressure, wind, humidity, cloud, etc) were then

171 transformed into the CMAQ-ready format via the Meteorology-Chemistry Interface Processor

172 (MCIP; Otte and Pleim (2010)) v4.3 which is a software to serve for transforming horizontal

173 and vertical coordinates while trying to maintain dynamic consistency between WRF and

174 CMAQ model simulations.

175

2.1.2 CMAQ model simulations

177 The CMAQ v5.1 model was used to estimate the concentrations of the atmospheric

178 chemical species over the domain, as shown in Fig. 1. The CMAQ domain has 204 latitudinal

179 and 273 longitudinal grid cells in total, and also has a 15×15 km^2 horizontal resolution and 27

180 sigma vertical layers. The CMAQv5.1 model was configured to use. Chemical and physical

181 configurations for the CMAQ model simulations were selected as follows: SAPRC07tc for the

182 gas-phase chemical mechanism (Hutzell et al., 2012); AERO6 for aerosol thermodynamics

183 (Appel et al., 2013); Euler Backward Iterative (EBI) chemistry solver (Hertel et al., 1993),

184 which is a numerically optimized photochemistry mechanism solver; M3DRY for dry

185 deposition velocity (Pleim and Xiu, 2003; Xiu and Pleim, 2001); global mass-conserving

186 scheme (YAMO & WRF) for horizontal and vertical advection (Colella and Woodward, 1984);

187 MULTISCALE (Louis, 1979), which is a simple first-order eddy diffusion scheme for

188 horizontal diffusion; and the Asymmetric Convective Model 2 (ACM2; Pleim, 2007a, 2007b)

189 for vertical diffusion.

190   For anthropogenic emissions, KORUS v1.0 emissions (Woo et al., 2012) were used.

191 The KORUS v1.0 emissions cover almost all of Asia, and are based on three emission

192 inventories: the Comprehensive Regional Emissions inventory for Atmospheric Transport

193 Experiment (CREATE) for East Asia excluding Japan; the Model Inter-Comparison Study for

194 Asia (MICS-Asia) for Japan; and the Studies of Emissions and Atmospheric Composition,

195 Clouds and Climate Coupling by Regional Surveys (SEAC4RS) for South and Southeast Asia.

196   Biogenic emissions were prepared by running the Model of Emissions of Gases and

197 Aerosols from Nature (MEGAN v2.1; Guenther et al., 2006; 2012) with a grid size identical to

198 that of the CMAQ model simulations. For the MEGAN simulations, the MODIS land cover

199 data (Friedl et al., 2010) and improved leaf area index (LAI) based on MODIS datasets (Yuan

200 et al., 2011) were utilized. Pyrogenic emissions were obtained from the Fire Inventory from

201 NCAR (FINN; Wiedinmyer et al., 2006, 2011). The lateral BCs for the CMAQ model

202 simulations were prepared using the global model results of the Model for Ozone and Related

203 chemical Tracers version 4 (MOZART-4; Emmons et al., 2010) at every 6 hours. The mapping

204 and re-gridding of the MOZART-4 data were conducted by matching the CMAQ grid

205   information.

206

207   **2.2 Observation data**

208   **2.2.1. Satellite-based observations**

209   A Korean geostationary satellite of Communication, Ocean, and Meteorological

210   Satellite (COMS) was launched on 26 June in 2010 over the Korean Peninsula. The COMS is

211   a geostationary orbit satellite and it is stationed at an altitude of approximately 36,000 km at a

212   latitude of 36°N and a longitude of 128.2°E with a horizontal coverage of 2500×2500 km^2

213   (refer to Fig. 1). Among the three payloads of the COMS, Geostationary Ocean Color Image

214   (GOCI) is the first multi-channel ocean color sensor with visible and near infrared channels.

215   The GOCI instrument provides hourly spectral images with a spatial resolution of 500×500

216   m^2 from 00:30 to 07:30 Coordinated Universal Time (UTC) for eight spectral (6 visible and 2

217   near-infrared) channels at 412, 443, 490, 555, 660, 680, 745, and 865 nm.

218   The Yonsei aerosol retrieval (YAER) algorithm for the GOCI sensor was initially

219   developed by Lee et al. (2010) to retrieve the aerosol optical properties (AOPs) over ocean

220   areas, and was then improved by expanding to consider non-spherical aerosol optical properties

221   (Lee et al., 2012). Choi et al. (2016) further extended the algorithm for application to land

222   surfaces, and the algorithm was referred to as the GOCI YAER version 1 algorithm. With the

223   GOCI YAER algorithm, hourly Aerosol Optical Depths (AODs) at 550 nm were produced over

224   East Asia. Choi et al. (2016) compared the retrieved GOCI AODs with other satellite-retrieved

225   and ground-based observations, and found several errors in the cloud masking and surface

226   reflectances. These errors were corrected in the recently updated second version of the GOCI

227   YAER algorithm (Choi et al., 2018), which used the updated cloud masking and more accurate

228   surface reflectances. In this study, the most recent GOCI AOD products from the GOCI YAER

229 version 2 algorithm were used.

230

2.2.2. Ground-based observations

232   In addition to the satellite data, ground-based observations in South Korea and China

233 were also collected for use in the air quality prediction system for PM and gas-phase pollutants.

234 The orange, red, and blue dots in Fig. 1 represent the ground-based observation sites in China,

235 Air Korea, and super-site stations in South Korea, respectively. These observations provide

236 real-time concentrations of criteria species such as PM_{10}, $PM_{2.5}$, CO, O_3, SO_2, and NO_2.

237   Throughout the period of the KORUS-AQ campaign, ground-based observation data

238 were obtained from 1514 stations in China, 264 Air Korea stations, and seven super-site

239 stations in South Korea. In this study, 80 % of the ground-based observations in China and Air

240 Korea stations in South Korea were randomly selected for use in the prediction system. The

241 other 20 % of the data and super-site observations were used to evaluate the performances of

242 the developed air quality prediction system.

243   In addition, AErosol RObotic NETwork (AERONET) AODs were used to conduct an

244 independent evaluation of the air quality prediction system. AERONET is a federated global

245 ground-based sun photometer network (Holben et al., 1998). Cloud-screened and quality-

246 assured level 2.0 AODs for the AERONET were used in this study.

247

2.3 Air quality prediction system

249   In the present study, the air quality prediction system was developed by adjusting the

250 ICs for the CMAQ model simulations based on DA with satellite-retrieved and ground-

251 measured observations. Two parallel WRF-CMAQ model runs were conducted. The first

252 experiment that involved adjusting ICs via DA is referred to as DA RUN (see Fig. 2). In order

253 to evaluate the prediction system, a second experiment, in which the ICs were originated from

254 the previous CMAQ model simulations without assimilations, was also conducted. This

255 CMAQ run is referred to as BASE RUN.

256

2.3.1. AOD calculations

258 CMAQ AODs are calculated by integrating the aerosol extinction coefficient (σ_{ext})

259 using the following equation:

260

$$\text{AOD}(\lambda) = \int_0^z \sigma_{ext}(\lambda) \ dz \tag{1}$$

262

263 where z represents the vertical height; σ_{ext} is defined as the sum of the absorption coefficient

264 (σ_{abs}) and the scattering coefficient (σ_{sca}); and σ_{abs} and σ_{sca} can be estimated by Eqns (3)

265 and (4), respectively, as shown below:

266

$$\sigma_{ext}(\lambda) = \sigma_{abs}(\lambda) + \sigma_{sca}(\lambda) \tag{2}$$

$$\sigma_{abs}(\lambda) \ [\text{Mm}^{-1}] = \sum_i^n \sum_j^m \{(1 - \omega_{ij}(\lambda)) \cdot \beta_{ij}(\lambda) \cdot f_{ij}(\text{RH}) \cdot [C]_{ij}\} \tag{3}$$

$$\sigma_{sca}(\lambda) \ [\text{Mm}^{-1}] = \sum_i^n \sum_j^m \{\omega_{ij}(\lambda) \cdot \beta_{ij}(\lambda) \cdot f_{ij}(\text{RH}) \cdot [C]_{ij}\} \tag{4}$$

270

271 where i and j denote the particulate species and size bin (or particle mode), respectively; $\omega_{ij}(\lambda)$

272 is the single scattering albedo; $\beta_{ij}(\lambda)$ is the mass extinction efficiency (MEE) of particulate

273 species i for the size bin or particle mode j; $[C]_{ij}$ is the concentration of particulate species

274 including $(\text{NH}_4)_2\text{SO}_4$, NH_4NO_3, black carbon, organic aerosols (OA), mineral dust, and sea-

275   salt aerosols; RH is the relative humidity; $f_{ij}(RH)$ is the hygroscopic factor; and the single

276   scattering albedo (ω) implies to the fraction (portion) of scattering in the total extinction.

277   Using Eqns. (2) - (4), AODs were calculated from the aerosol composition and RH.

278   These have been intensive tests using different β and $f(RH)$ values in the following three

279   previous studies: (1) Chin et al. (2002)'s study with the Goddard Chemistry Aerosol Radiation

280   and Transport (GOCART) model; (2) Martin et al. (2003)'s study with the GEOS-Chem model;

281   and (3) Malm and Hand (2007)'s study with the CMAQ model. Lee et al. (2016) tested these

282   methods, and then found that Chin et al. (2002)'s method reproduced the best results in

283   estimating AODs at 550 nm over East Asia. On the basis of Lee et al. (2016)'s work, σ_{ext} was

284   estimated with the β and $f(RH)$ values suggested by Chin et al. (2002). After that, σ_{ext} was

285   integrated with respect to altitude, in order to calculate the AODs. The calculated AODs were

286   used in the air quality prediction system in order to prepare the ICs for the PM predictions.

287

288   **2.3.2. Data assimilation (DA)**

289   The ground-based observations, together with GOCI-derived AODs, were used to

290   prepare the ICs for the air quality predictions with the CMAQ model simulations. In order to

291   achieve this, the following steps were taken: (i) the CMAQ-calculated concentrations of CO,

292   O_3, and SO_2 were combined with the concentrations of CO, O_3, and SO_2 obtained from ground-

293   based observations in South Korea (Air Korea) and China; (ii) the CMAQ-calculated AODs

294   were assimilated with the GOCI AODs; (iii) the assimilated AODs were converted into PM_{10};

295   (iv) the converted PM_{10} was again assimilated at the surface in South Korea and China; and (v)

296   after the DA at the surface, the ratios of the assimilated species concentrations to the original

297   CMAQ-simulated concentrations were applied so as to the adjust vertical profiles of the

298   chemical species above the surface. In the air quality prediction system, the DA cycle is 24

299   hours and the assimilation takes place every day at 00:00 UTC (refer to Fig. 3).

300   The optimal interpolation (OI) method with the Kalman filter was chosen in the air

301   quality prediction system. The OI method was originally used for meteorological applications

302   (Lorenc, 1986), and has also been used in the assimilations for trace gases (Khattatov et al.,

303   1999, 2000; Lamarque et al., 1999; Levelt et al., 1998). Recently, the OI technique has also

304   been applied to aerosol fields (Collins et al., 2001; Yu et al., 2003; Generoso et al., 2007;

305   Adhikary et al., 2008; Carmichael et al., 2009; Chung et al., 2010; Park et al., 2011; Park et al.,

306   2014; Tang et al., 2015, 2017).

307   Aerosol assimilation using the OI method was first applied by Collins et al. (2001) as

308   follows:

309

310 $$\tau'_m = \tau_m + \mathbf{K}(\tau_o - \mathbf{H}\tau_m) \tag{5}$$

311 $$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{O})^{-1} \tag{6}$$

312 $$\mathbf{O} = [(f_o\tau_o)^2 + (\varepsilon_o)^2]\mathbf{I} \tag{7}$$

313 $$\mathbf{B}(d_x, d_z) = [(f_m\tau_m)^2 + (\varepsilon_m)^2]\exp\left[-\frac{d_x^2}{2l_{mx}^2}\right]\exp\left[-\frac{d_z^2}{2l_{mz}^2}\right] \tag{8}$$

314

315   where τ'_m, τ_m, and τ_o represent the assimilated products by the OI method, the modeled

316   values, and the observed values, respectively; $\mathbf{K}$ is the Kalman gain matrix; $\mathbf{H}$ is the

317   observation operator (or forward operator), which is an interpolator from model to observation

318   space; $\mathbf{B}$ and $\mathbf{O}$ are the background and observation error covariance matrices, respectively;

319     $(\cdot)^T$ denotes the transpose of a matrix; f_o is the fractional error in the observation-retrieved

320   value; ε_o is the minimum root mean square error in the observation-retrieved values; $\mathbf{I}$

321   denotes the unit matrix; f_m is the fractional error in the model estimates; ε_m is the minimum

322   root mean square error in the model estimates; d_x is the horizontal distance between two

323   model grid points; l_{mx} is the horizontal correlation length scale for the errors in the model;

324   d_z is the vertical distance between two model grid points; and l_{mz} is the vertical correlation

325   length scale for the errors in the model. In this work, the OI technique was applied for the DA

326   of atmospheric gaseous species as well as particulate species.

327   Six free parameters (f_m, f_o, ε_m, ε_o, l_{mx}, and l_{mz}) were used to calculate the error

328   covariance matrices of the observations and model, the mathematical formalisms of which are

329   described in Eq. (7) and (8), respectively. Several previous studies have used fixed values for

330   free parameters (Collins et al., 2001; Yu et al., 2003; Adhikary et al., 2008; Chung et al., 2010).

331   These runs are called "static" runs. In contrast to those previous studies, "non-static" free

332   parameters were applied in this study by minimizing the differences between the assimilated

333   values and observations via an iterative process at each assimilation time step. This non-static

334   free parameter scheme is possible due to the fact that the OI technique with the Kalman filter

335   is much less costly in terms of computation time than other DA techniques, such as the 3-D or

336   4-D variational methods. This is another advantage of using the OI technique in this system. It

337   typically takes less than 20 minutes with a workstation environment (dual Intel Xeon 2.40 GHz

338   processor).

339

340   **2.3.3. Allocation of the assimilated PM$_{10}$ & PM$_{2.5}$ into particulate composition**

341   In the procedure of DA, PM$_{10}$ was assimilated in this study, because the PM$_{10}$ data

342   were more plentiful than PM$_{2.5}$. The assimilated PM$_{10}$ then needs to be allocated into the PM

343   composition for the CMAQ-model prediction runs. In order to achieve this, the differences

344   between the assimilated PM_{10} and background PM_{10} (ΔPM_{10}) were first calculated. Then,

345    $\Delta PM_{2.5}$ was estimated using the ratios of $PM_{2.5}$ to PM_{10} from the background CMAQ model

346   runs (i.e., $\Delta PM_{2.5}=\Delta PM_{10}\times PM_{2.5}/PM_{10}$). $\Delta PM_{2.5}$ was then allocated to the $PM_{2.5}$ composition

347   according to the comparison between two $PM_{2.5}$ compositions observed at the seven super-sites

348   and simulated from the CMAQ model runs over South Korea. Both of the compositions are

349   shown in Fig. 4. In Fig. 4, "PM OTHERS" indicates the remaining particulate matter species

350   after excluding sulfate, nitrate, ammonium, organic aerosol (OA), and elementary carbon (EC).

351   The PM OTHERS occupies 25 % of the total $PM_{2.5}$ observed at super-sites. The other

352   fraction, $\Delta PM_{10}\times(1-PM_{2.5}/PM_{10})$, was also distributed into the coarse-mode particles ($PM_{2.5-10}$)

353   as crustal elements.

354

355   **3. Results and discussions**

356   The performances of the air quality prediction system were evaluated by comparing

357   them with ground-based observations from the Air Korea network and super-site stations in

358   South Korea. Several sensitivity analyses were also conducted in order to assess the influences

359   of the DA time-intervals on the accuracy of the air quality prediction.

360

361   **3.1. Evaluation of the air quality prediction system**

362   **3.1.1. Time-series analysis**

363   Figure 5 shows the time-series plots of PM_{10}, $PM_{2.5}$, CO, O_3, and SO_2 concentrations

364   from the BASE RUN and the DA RUN. Here, the observation data (OBS) obtained from the

365   Air Korea network were compared with the results of the two sets of the CMAQ model

366   simulations, i.e., (1) BASE RUN and (2) DA RUN. As mentioned previously, 20% of the Air

367   Korea observations used in the evaluation were randomly selected during the period of the

368   KORUS-AQ campaign. The other 80 % of the Air Korea data were used in the DA at 00:00

369   UTC. For the forecast hours from 01:00 to 23:00 UTC, all of the ground observations (254 Air

370   Korea and seven super-site stations) were used to evaluate the performances of the developed

371   air quality prediction system. As shown in Fig. 5, we achieved some improvements in the

372   prediction performances by applying the ICs to the CMAQ model simulations. The BASE

373   RUN significantly under-predicted PM_{10}, $PM_{2.5}$, and CO while the DA RUN produced

374   concentrations that were more consistent with the observations than those of the BASE RUN.

375   In case of CO, the observed CO mixing ratios were about three times higher than those

376   from the BASE RUN. These large differences are well known, and have been attributed to the

377   underestimated emissions of CO (Heald et al., 2004). However, when the DA was applied, the

378   predictions of the CO mixing ratios improved. Similarly, the performances of the PM_{10} and

379     $PM_{2.5}$ predictions improved with the application of the DA. Unlike PM_{10}, $PM_{2.5}$, and CO, the

380   O_3 mixing ratios and its diurnal trends from both the BASE RUN and DA RUN tend to be well-

381   matched with the observations. By contrast, the poorest performances of the BASE RUN and

382   the DA RUN were shown for SO_2.

383   In addition, a dust event took place between 6 May and 8 May. This event is captured by

384   the DA RUN (check red peaks in Fig. 5(a) and (b)), while the BASE RUN cannot capture this

385   dust event. This demonstrates the capability of the current system to possibly predict dust

386   events in South Korea. In the DA RUN, dust information is provided into the CMAQ model

387   runs through both/either GOCI AOD and/or ground PM observations measured along the dust

388   plume tracks.

389   The effectiveness of the DA with prediction time was also analyzed by calculating the

390   aggregated average concentrations of atmospheric species (see Figs. 6, 7, and 9). Fig. 6 depicts

391   the CMAQ-calculated average concentrations of PM_{10}, $PM_{2.5}$, CO, and SO_2 against the Air

392   Korea observations. Our air quality prediction system re-generated relatively well-matched

393   concentrations for PM_{10}, $PM_{2.5}$, and CO from the DA RUN.

394   Figure 7 shows the case of ozone from the DA RUN by assimilating CMAQ outputs

395   with Air Korea-observed (a) O_3 mixing ratios, and (b) both O_3 and NO_2 mixing ratios for a

396   preliminary test run. The ozone mixing ratios from the DA RUN in Fig. 7(a) were reasonably

397   consistent with the observations at 00:00 UTC, but disagreed with those between 04:00 and

398   09:00 UTC (13:00 and 18:00 KST), when solar insolation is the most intense. This may be

399   attributed to the chemical imbalances between ozone production and ozone destruction (or

400   titration). However, if CMAQ NO_2 was assimilated with ground-based observations in South

401   Korea (Air Korea) and China, the predicted ozone mixing ratios became substantially closer to

402   the observations, as shown in Fig. 7(b). This is clearly due to the fact that NO_x is an important

403   precursor of ozone. In the prediction of the ozone mixing ratios, both 1-hr peak ozone (around

404   15:00 KST) and 8-hr averaged ozone mixing ratios (between 9:00 and 17:00 KST) are

405   important. Fig. 7 clearly shows that the prediction accuracies of both the ozone mixing ratios

406   were improved after the DA of NO_2 mixing ratios.

407   Although the DA for NO_2 provided better ozone predictions, one should take caution

408   in using the NO_2 observations. The NO_2 mixing ratios measured at Air Korea sites are known

409   to be contaminated by other nitrogen gases such as nitric acid (HNO_3), peroxyacetyl nitrates

410   (PANs), and alkyl nitrates (ANs), since the Air Korea NO_2 mixing ratios are measured through

411   a chemiluminescent method with catalysts of gold or molybdenum oxide at high temperatures.

412   These are known to be "NO$_2$ measurement artifacts" (Jung et al., 2017), which is one of the

413   reasons that the DA of NO$_2$ was not shown in Fig. 6. The NO$_2$ mixing ratios are corrected from

414   the Air Korea NO$_2$ data, and are then used to prepare the ICs via the DA for more accurate

415   ozone and NO$_2$ predictions. Currently, such corrections of the observed NO$_2$ mixing ratios are

416   being standardized with more sophisticated year-long NO$_2$ measurements. After the corrections

417   of the NO$_2$ measurement artifacts, more evolved schemes of ozone and NO$_2$ predictions will

418   be possible in the future. As shown in Fig. 7, about a 20% reduction (average fraction of non-

419   NO$_2$ mixing ratios in the observed NO$_2$ mixing ratios) was made for these demonstration runs

420   (Jung et al., 2017)**.**

421   Another practical issue is now discussed. Although the assimilation with the observed

422   NO$_2$ mixing ratios can enhance the accuracy of the predictions of the daytime ozone mixing

423   ratios, the nighttime ozone mixing ratios tend to be consistently over-predicted in the

424   aggregated plot of the ozone mixing ratios at the observation sites (see Fig. 7). This can be

425   caused by underestimated NO$_2$ mixing ratios and thus not enough nighttime ozone titration. As

426   aforementioned, reliable NO$_2$ prediction via the correction of the NO$_2$ measurement artifacts

427   will be made in the future. Another possible reason of the over-predicted ozone mixing ratios

428   during the nighttime can be underestimation of the mixing layer height (MLH). Figure 8 shows

429   a comparison between lidar-measured MLH (black dashed line) and WRF-calculated MLH

430   (with the option of the Yonsei University (YSU) planetary boundary layer scheme by Hong et

431   al. (2006); see red line). As shown in Fig. 8, the nocturnal lidar-measured MLH is about two

432   times higher than the nocturnal WRF-calculated MLH as measured at a lidar site inside the

433   campus of Seoul National University (SNU) in Seoul. Such underestimated MLH in the model

434   tends to compress the ozone molecules within the mixing layer during the nighttime, which

435   leads to consistently over-predicted nocturnal ozone mixing ratios. Based on this discrepancy

436   shown in Fig. 8, more intensive comparison study is being carried out by comparing lidar-

437   measured MLH with model-calculated MLH at multiple sites in South Korea.

438   In this work, the aerosol composition (such as EC, OA, sulfate, nitrate, and ammonium)

439   was further compared with the composition observed at the super-sites shown in Fig. 9. As

440   shown in Fig. 9, agreement was observed between the DA RUN and observations for all of the

441   major PM constituents. Again, a strong capability of our DA system is to improve the

442   predictions of the aerosol composition.

443

444   **3.1.2. Spatial distribution**

445   Figure 10 shows the spatial distributions and bias of PM and chemical species

446   throughout the entire period of the KORUS-AQ campaign over the Seoul Metropolitan Area

447   (SMA). Noticeable improvements are observed to have been achieved in the spatial

448   distributions by applying the ICs into the CMAQ model simulations, particularly for PM_{10} (Fig.

449   10a), $PM_{2.5}$ (Fig. 10b), and CO (Fig. 10c). As shown in Fig. 10, the under-predicted

450   concentrations of PM_{10}, $PM_{2.5}$, and CO were adjusted to concentrations closer to the

451   observations. In case of SO_2 (see Fig. 10d), the DA RUN produced better agreement with the

452   observations than the BASE RUN, but there were still under-predicted SO_2 concentrations over

453   the northeastern part of the SMA.

454   By contrast, relatively lower ozone mixing ratios from the DA RUN against the BASE

455   RUN were found in the southwestern part of the SMA (see Fig. 10e). Due to the nonlinear

456   relationship between NO_x and O_3, high mixing ratios (or emissions) of NO_x in the SMA can

457   lead to depletion of ozone. In these runs, the precursors of ozone such as NO_x and VOCs were

458   excluded in the preparation of the ICs for CMAQ model simulations. Again, this is because the

459   Air Korea NO_2 mixing ratios are contaminated by several reactive nitrogen species, so the data

460   cannot be directly used in the assimilation procedures. In case of VOCs, a limited number of

461   datasets is available in South Korea for the DA. Improvements in the prediction of ozone

462   mixing ratios can be achieved when the NO_2 mixing ratios are corrected and a sufficient

463   number of VOCs data (possibly from satellite data in the future) is available.

464

465   **3.1.3. Statistical analysis**

466   In order to achieve better understanding of the performances of the DA RUN, analyses

467   of statistical variables such as index of agreement (IOA), Pearson's correlation coefficient (R),

468   root mean square error (RMSE), and mean bias (MB) were conducted using observations from

469   the Air Korea stations for PM_{10}, $PM_{2.5}$, CO, SO_2, and O_3 (see Fig. 11). Definitions of the

470   statistical variables are given in Appendix A.

471   After the applications of the ICs, both RMSE and MB became lower, while the

472   correlation coefficient became higher for the entire predictions. In addition, it was found that

473   the differences between the BASE RUN and the DA RUN tended to diminish as the prediction

474   time progressed. The results of the statistical analysis are listed in Table 1. The results of the

475   DA RUN were reasonably consistent with the observations for PM_{10} (IOA = 0.60; R= 0.40;

476   RMSE = 34.87; MB = -13.54) and $PM_{2.5}$ (IOA = 0.71; R= 0.53; RMSE = 17. 83; MB = -2.43),

477   as compared to the BASE RUN for PM_{10} (IOA = 0.51; R= 0.34; RMSE = 40.84; MB = -27.18)

478   and $PM_{2.5}$ (IOA = 0.67; R= 0.51; RMSE = 19.24; MB = -9.9). In terms of bias, an improvement

479   was found for CO: MB = -0.036 for the DA RUN and MB = -0.27 for the BASE RUN.

480   Regarding O_3 and SO_2, the DA RUN showed slightly better performances than the BASE RUN.

481   Table 2 presents the results of the statistical analysis at 00:00 UTC when the DA was

482   conducted, with the results clearly showing how much closer the DA makes the CMAQ-

483   calculated chemical concentrations to the observed concentrations. Collectively, the DA

484   improved model accuracy by a large degree in terms of R, particularly for PM_{10} (R: 0.3→0.75;

485   slope: 0.17→0.66) and O_3 (R: 0.09→0.61; slope: 0.07→0.42). In addition, for all species, MB

486   and RMSE decreased significantly with the DA RUN as compared with the BASE RUN.

487

3.2. Sensitivity test of DA time-interval

3.2.1. AOD

490   In this section, a sensitivity analysis was conducted with different implementation

491   time-intervals of the DA (i.e., 24, 6, and 3 hours) for AOD (refer to Fig. 12). As shown in Fig.

492   12, more frequent implementation of the DA is expected to make the predicted results closer

493   to the observations. Although the DA RUN with a shorter assimilation time-interval tends to

494   produce a better prediction, it is not always the most appropriate choice, since the shorter

495   assimilation time-interval results in increased computational cost. Therefore, an optimized

496   assimilation time-interval should be found to achieve the best performances from the given DA

497   system with the consideration of its own computational ability.

498

3.2.2. PM and gases

500   In addition, sensitivity analyses of the developed air quality prediction system to

501   multiple implementations of the DA with different time-intervals were also investigated for (a)

502   PM_{10}, (b) $PM_{2.5}$, (c) CO, (d) SO_2, and (e) O_3, shown in Fig. 13. Fig. 13 shows a soccer plot

503   analysis for BASE RUN (blue crosses) and DA RUNs with different DA time-intervals of 24

504   hours (OI; red circles), two hours (2-hr OI; black diamonds), and one hour (1-hr OI; dark-green

505   triangles). This set of testing was designed based on the fact that the performances are expected

506   to improve if the DAs are implemented multiple times prior to the actual predictions at 00:00

507   UTC. Here, for the 2-hr OI run, the DA was implemented three times a day at 20:00, 22:00,

508   and 00:00 UTC, while for the 1-hr OI run, the DA was implemented at 22:00, 23:00, and 00:00

509   UTC. The performances of all of the chemical species excluding ozone improved, as expected,

510   with DA RUNs with more frequent and longer DA time-intervals (i.e., three-times

511   implementation with a 2-hr time-interval in our cases). In case of ozone, the best performance

512   was found for the air quality prediction system with the DA time-interval of 24-hr.

513   Unsurprisingly, more frequent DAs prior to the actual prediction mode (i.e., before

514   00:00 UTC in our system) with a longer time-interval (such as 2-hr) will be computationally

515   costly. There will certainly be a "trade-off" between the precision of air quality prediction and

516   the computational cost. The system should be designed under the consideration of these two

517   factors.

518

519   **4. Summary and conclusions**

520   In this study, the air quality prediction system was developed by preparing the ICs for

521   CMAQ model simulations using GOCI AODs and ground-based observations of PM_{10}, CO,

522   ozone, and SO_2 during the period of the KORUS-AQ campaign (1 May – 12 June 2016) in

523   South Korea. The major advantages of the developed air quality prediction system are its

524   comprehensiveness in predicting the ambient concentrations of both gaseous and particulate

525   species (including PM composition) as well as its powerfulness in terms of computational cost.

526   The performances of the developed prediction system were evaluated using near-

527   surface in-situ observation data. The CMAQ model runs with the ICs (DA RUN) showed higher

528   consistency with the observations of almost all of the chemical species, including PM

529   composition (sulfate, nitrate, ammonium, OA, and EC) and atmospheric gases (CO, ozone, and

530   SO_2), than the CMAQ model runs without the ICs (BASE RUN). Particularly for CO, the DA

531   was able to remarkably improve the model performances, while the BASE RUN significantly

532   under-predicted the CO concentrations (predicting about one-third of the observed values). In

533   case of ozone, both the BASE RUN and DA RUN were in close agreement with observations.

534   More reliable predictions of ozone mixing ratios will be achieved via the DA of the observed

535   NO_2 mixing ratios and the corrections of model-simulated mixing layer height (MLH). For SO_2,

536   the performances of both the BASE RUN and the DA RUN were somewhat poor. Regarding

537   this issue, more accurate SO_2 emissions are required to achieve better SO_2 predictions, and

538   these can be estimated through inverse modeling using satellite data (e.g., Lee et al., 2011).

539   The adjustments of both ICs and emissions may be able to improve the performances of the air

540   quality prediction system, and this will be examined in future studies.

541   Moreover, the developed air quality prediction system will be upgraded by using the

542   new observation data that will be retrieved after 2020 from the Geostationary Environment

543   Monitoring Spectrometer (GEMS) with a high spatial resolution of 7×8 km^2 as well as a high

544   temporal resolution of 1-hour over a large part of Asia. In addition, the current DA technique

545   of the OI with the Kalman filter can also be upgraded with the use of more advanced DA

546   methods such as variational techniques of 3DVAR and 4DVAR methods, as well as with the

547   ensemble Kalman filter (EnK) method. These research endeavors are currently underway.

548   In conjunction with improving the air quality modeling system, artificial intelligence

549   (AI)-based air quality prediction systems are also currently being developed in several ways

550   (e.g., Kim et al., 2019). Actually, Kim et al. (2019) developed an AI-based PM prediction

551   system based on a deep recurrent neural network (RNN) in South Korea. The AI-based

552   prediction system was optimized by iterative model trainings with the inputs of ground-

553   observed PM_{10}, $PM_{2.5}$, and meteorological fields including wind speed, wind direction, relative

554   humidity, and precipitation. The AI-based prediction system showed better performances at the

555   several sites than the CMAQ model simulations. However, it works only for the observation

556   sites in South Korea where ground-based observations are available. By taking advantages of

557   both the CTM-based air quality prediction and the AI-based prediction systems, both systems

558   will be eventually combined so as to create a more accurate hybrid air quality prediction system

559   over South Korea. This will be the ultimate goal of the series of our research works.

560

561   **Code and data availability.** WRF v3.8.1 (doi:10.5065/D6MK6B4K) and CMAQ v5.1

562   (doi:10.5281/zenodo.1079909) models are both open-source and publicly available. Source

563   codes for WRF and CMAQ can be downloaded at http://www2.mmm.ucar.edu/wrf/users/

564   downloads.html and https://github.com/USEPA/CMAQ, respectively. Data from the KORUS-

565   AQ field campaign can be downloaded from the KORUS-AQ data archive (http://www-

566   air.larc.nasa.gov/missions/korus-aq). Other data were acquired as follows. Ground-based

567   observation data were downloaded from the Air Korea website (http://www.airkorea.or.kr) for

568   South Korea and https://pm25.in for China. AERONET data were downloaded from

569   https://aeronet.gsfc.nasa.gov. The KAQPS v1 (doi:10.5281/zenodo.3659551) code can be

570 obtained by contacting K. Lee (lkh1515@gmail.com) or from https://github.com/AIR-

571 Codes/KAQPSv1. NCL (2019; doi:10.5065/D6WD3XH5) was used to draw the figures.

572

573 **Author contributions.** KL developed the model code, performed the simulations, and

574 analyzed the results. CHS directed the experiments. JY contributed to shape the research and

575 analysis. SL, MP, HH, and SYP helped analyze the results. MC, JK, YK, JHW, and SWK

576 provided and analyzed data applied in the experiments. KL prepared the manuscript with

577 contributions from all co-authors.

578

579 **Acknowledgments**

580 This research was supported by the National Strategic Project-Fine particle of the National

581 Research Foundation of Korea (NRF) of the Ministry of Science and ICT (MSIT), the Ministry

582 of Environment (MOE), and the Ministry of Health and Welfare (MOHW) (NRF-

583 2017M3D8A1092022). This work was also funded by the GEMS program of the MOE of the

584 Republic of Korea as part of the Eco Innovation Program of KEITI (2012000160004) and was

585 supported by a grant from the National Institute of Environment Research (NIER), funded by

586 the MOE of the Republic of Korea (NIER-2019-01-01-028). Specially thanks to the entire

587 KORUS-AQ science team for their considerable efforts in conducting the campaign.

APPENDIX A: FORMULAS FOR STATISTICAL EVALUATION INDICES

The formulas used to evaluate the performances of the air quality prediction system are defined as follows.

$$\text{Index Of Agreement (IOA)} = 1 - \frac{\sum_1^n (M - O)^2}{\sum_1^n (|M - \overline{O}| + |O - \overline{O}|)^2} \quad (A1)$$

$$\text{Correlation Coefficient (R)} = \frac{1}{(n-1)} \sum_1^n \left(\left(\frac{O - \overline{O}}{\sigma_O} \right) \left(\frac{M - \overline{M}}{\sigma_m} \right) \right) \quad (A2)$$

$$\text{Root Mean Square Error (RMSE)} = \sqrt{\frac{\sum_1^n (M - O)^2}{n}} \quad (A3)$$

$$\text{Mean Bias (MB)} = \frac{1}{n} \sum_1^n (M - O) \quad (A4)$$

$$\text{Mean Normalized Bias (MNB)} = \frac{1}{n} \sum_1^n \left(\frac{M - O}{O} \right) \times 100 \% \quad (A5)$$

$$\text{Mean Normalized Error (MNE)} = \frac{1}{n} \sum_1^n \left(\frac{|M - O|}{O} \right) \times 100 \% \quad (A6)$$

$$\text{Mean Fractional Bias (MFB)} = \frac{1}{n} \sum_1^n \frac{(M - O)}{\left(\frac{M + O}{2} \right)} \times 100 \% \quad (A7)$$

606
$$\text{Mean Fractional Error (MFE)} = \frac{1}{n}\sum_{1}^{n}\frac{|M-O|}{\left(\frac{M+O}{2}\right)} \times 100\ \%\quad\text{(B8)}$$

607

608   In Eqns. (A1) - (A8), M and O represent the model and observation data, respectively. N is the

609   number of data points and σ means the standard deviation. The overbars in the equations

610   indicate the arithmetic mean of the data. The units of RMSE and MB are the same as the unit

611   of data, while IOA and R are dimensionless statistical parameters.

612

613   **References**

614   Adhikary, B., Kulkarni, S., Dallura, A., Tang, Y., Chai, T., Leung, L. R., Qian, Y., Chung, C. E.,
615   Ramanathan, V. and Carmichael, G. R.: A regional scale chemical transport modeling of
616   Asian aerosols with data assimilation of AOD observations using optimal interpolation
617   technique,   Atmospheric   Environment,   42(37),   8600–8615,
618   doi:10.1016/j.atmosenv.2008.08.031, 2008.

619   Appel, K. W., Roselle, S. J., Gilliam, R. C. and Pleim, J. E.: Sensitivity of the Community
620   Multiscale Air Quality (CMAQ) model v4.7 results for the eastern United States to MM5
621   and WRF meteorological drivers, Geoscientific Model Development, 3, 169–188, 2010.

622   Appel, K. W., Pouliot, G. A., Simon, H., Sarwar, G., Pye, H. O. T., Napelenok, S. L., Akhtar,
623   F. and Roselle, S. J.: Evaluation of dust and trace metal estimates from the Community
624   Multiscale Air Quality (CMAQ) model version 5.0, Geoscientific Model Development,
625   6(4), 883–899, doi:10.5194/gmd-6-883-2013, 2013.

626   Bellouin, N., Boucher, O., Haywood, J. and Reddy, M. S.: Global estimate of aerosol direct
627   radiative forcing from satellite measurements, Nature, 438(7071), 1138–1141,
628   doi:10.1038/nature04348, 2005.

629   Bréon, F.-M., Tanré, D. and Generoso, S.: Aerosol Effect on Cloud Droplet Size Monitored
630   from Satellite, Science, 295(5556), 834–838, doi:10.1126/science.1066434, 2002.

631   Byun, D. and Schere, K. L.: Review of the Governing Equations, Computational Algorithms,
632   and Other Components of the Models-3 Community Multiscale Air Quality (CMAQ)
633   Modeling System, Appl. Mech. Rev, 59(2), 51–77, doi:10.1115/1.2128636, 2006.

634   Byun, D. W. and Ching, J. K. S.: Science algorithms of the EPA models-3 community
635   multiscale air quality (CMAQ) modeling system, U.S. Environmental Protection Agency,
636   EPA/600/R-99/030 (NTIS PB2000-100561)., 1999.

637   Carmichael, G. R., Sakurai, T., Streets, D., Hozumi, Y., Ueda, H., Park, S. U., Fung, C., Han,
638   Z., Kajino, M., Engardt, M., Bennet, C., Hayami, H., Sartelet, K., Holloway, T., Wang, Z.,
639   Kannari, A., Fu, J., Matsuda, K., Thongboonchoo, N. and Amann, M.: MICS-Asia II: The
640   model intercomparison study for Asia Phase II methodology and overview of findings,
641   Atmospheric Environment, 42(15), 3468–3490, doi:10.1016/j.atmosenv.2007.04.007,
642   2008.

643   Carmichael, G. R., Adhikary, B., Kulkarni, S., D'Allura, A., Tang, Y., Streets, D., Zhang, Q.,
644   Bond, T. C., Ramanathan, V., Jamroensan, A. and Marrapu, P.: Asian Aerosols: Current
645   and Year 2030 Distributions and Implications to Human Health and Regional Climate
646   Change, Environ. Sci. Technol., 43(15), 5811–5817, doi:10.1021/es8036803, 2009.

647   Chemel, C., Sokhi, R. S., Yu, Y., Hayman, G. D., Vincent, K. J., Dore, A. J., Tang, Y. S., Prain,
648   H. D. and Fisher, B. E. A.: Evaluation of a CMAQ simulation at high resolution over the
649   UK for the calendar year 2003, Atmospheric Environment, 44(24), 2927–2939,
650   doi:10.1016/j.atmosenv.2010.03.029, 2010.

Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., Martin, R. V., Logan, J. A., Higurashi, A. and Nakajima, T.: Tropospheric Aerosol Optical Thickness from the GOCART Model and Comparisons with Satellite and Sun Photometer Measurements, J. Atmos. Sci., 59(3), 461–483, doi:10.1175/1520-0469(2002)059<0461:TAOTFT>2.0.CO;2, 2002.

Choi, M., Kim, J., Lee, J., Kim, M., Park, Y.-J., Jeong, U., Kim, W., Hong, H., Holben, B., Eck, T. F., Song, C. H., Lim, J.-H. and Song, C.-K.: GOCI Yonsei Aerosol Retrieval (YAER) algorithm and validation during the DRAGON-NE Asia 2012 campaign, Atmos. Meas. Tech., 9(3), 1377–1398, doi:10.5194/amt-9-1377-2016, 2016.

Choi, M., Kim, J., Lee, J., Kim, M., Park, Y.-J., Holben, B., Eck, T. F., Li, Z. and Song, C. H.: GOCI Yonsei aerosol retrieval version 2 products: an improved algorithm and error analysis with uncertainty estimation from 5-year validation over East Asia, Atmos. Meas. Tech., 11(1), 385–408, doi:10.5194/amt-11-385-2018, 2018.

Chung, C. E., Ramanathan, V., Carmichael, G., Kulkarni, S., Tang, Y., Adhikary, B., Leung, L. R. and Qian, Y.: Anthropogenic aerosol radiative forcing in Asia derived from regional models with atmospheric and aerosol data assimilation, Atmos. Chem. Phys., 10(13), 6007–6024, doi:10.5194/acp-10-6007-2010, 2010.

Colella, P. and Woodward, P. R.: The Piecewise Parabolic Method (PPM) for gas-dynamical simulations, Journal of Computational Physics, 54(1), 174–201, doi:10.1016/0021-9991(84)90143-8, 1984.

Collins, W. D., Rasch, P. J., Eaton, B. E., Khattatov, B. V., Lamarque, J.-F. and Zender, C. S.: Simulating aerosols using a chemical transport model with assimilation of satellite aerosol retrievals: Methodology for INDOEX, J. Geophys. Res., 106(D7), 7313–7336, doi:10.1029/2000JD900507, 2001.

Cressman, G. P.: An operational objective analysis system, Mon. Wea. Rev., 87(10), 367–374, 1959.

Dehghani, M., Keshtgar, L., Javaheri, M. R., Derakhshan, Z., Conti, O., Gea, Zuccarello, P. and Ferrante, M.: The effects of air pollutants on the mortality rate of lung cancer and leukemia, Molecular Medicine Reports, 15(5), 3390–3397, 2017.

Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S. L. and Kloster, S.: Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4), Geosci. Model Dev., 3(1), 43–67, doi:10.5194/gmd-3-43-2010, 2010.

Foley, K. M., Roselle, S. J., Appel, K. W., Bhave, P. V., Pleim, J. E., Otte, T. L., Mathur, R., Sarwar, G., Young, J. O., Gilliam, R. C., Nolte, C. G., Kelly, J. T., Gilliland, A. B. and Bash, J. O.: Incremental testing of the Community Multiscale Air Quality (CMAQ) modeling system version 4.7, Geosci. Model Dev., 3(1), 205–226, doi:10.5194/gmd-3-

689   205-2010, 2010.

690 Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A. and Huang,
691   X.: MODIS Collection 5 global land cover: Algorithm refinements and characterization
692   of new datasets, Remote Sensing of Environment, 114(1), 168–182,
693   doi:10.1016/j.rse.2009.08.016, 2010.

694 Generoso, S., Bréon, F.-M., Chevallier, F., Balkanski, Y., Schulz, M. and Bey, I.: Assimilation
695   of POLDER aerosol optical thickness into the LMDz-INCA model: Implications for the
696   Arctic aerosol burden, J. Geophys. Res., 112(D2), D02311, doi:10.1029/2005JD006954,
697   2007.

698 Grell, G. A. and Freitas, S. R.: A scale and aerosol aware stochastic convective parameterization
699   for weather and air quality modeling, Atmospheric Chemistry and Physics, 14(10), 5233–
700   5250, doi:https://doi.org/10.5194/acp-14-5233-2014, 2014.

701 Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I. and Geron, C.: Estimates of
702   global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and
703   Aerosols from Nature), Atmospheric Chemistry and Physics, 6(11), 3181–3210, 2006.

704 Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K. and
705   Wang, X.: The Model of Emissions of Gases and Aerosols from Nature version 2.1
706   (MEGAN2.1): an extended and updated framework for modeling biogenic emissions,
707   Geosci. Model Dev., 5(6), 1471–1492, doi:10.5194/gmd-5-1471-2012, 2012.

708 Heald, C. L., Jacob, D. J., Jones, D. B. A., Palmer, P. I., Logan, J. A., Streets, D. G., Sachse, G.
709   W., Gille, J. C., Hoffman, R. N. and Nehrkorn, T.: Comparative inverse analysis of satellite
710   (MOPITT) and aircraft (TRACE-P) observations to estimate Asian sources of carbon
711   monoxide: COMPARATIVE INVERSE ANALYSIS, Journal of Geophysical Research:
712   Atmospheres, 109(D23), doi:10.1029/2004JD005185, 2004.

713 Hertel, O., Berkowicz, R., Christensen, J. and Hov, Ø.: Test of two numerical schemes for use
714   in atmospheric transport-chemistry models, Atmospheric Environment. Part A. General
715   Topics, 27(16), 2591–2611, doi:10.1016/0960-1686(93)90032-T, 1993.

716 Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J.
717   A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I. and Smirnov, A.:
718   AERONET—A Federated Instrument Network and Data Archive for Aerosol
719   Characterization, Remote Sensing of Environment, 66(1), 1–16, doi:10.1016/S0034-
720   4257(98)00031-5, 1998.

721 Hong, S.-Y. and Lim, J.-O. J.: The WRF single-moment 6-class microphysics scheme (WSM6),
722   J. Korean Meteor. Soc., 42(2), 129–151, 2006.

723 Hong, S.-Y., Noh, Y. and Dudhia, J.: A New Vertical Diffusion Package with an Explicit
724   Treatment of Entrainment Processes, Mon. Wea. Rev., 134(9), 2318–2341,
725   doi:10.1175/MWR3199.1, 2006.

Hutzell, W. T., Luecken, D. J., Appel, K. W. and Carter, W. P. L.: Interpreting predictions from the SAPRC07 mechanism based on regional and continental simulations, Atmospheric Environment, 46, 417–429, doi:10.1016/j.atmosenv.2011.09.030, 2012.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A. and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, J. Geophys. Res., 113(D13), D13103, doi:10.1029/2008JD009944, 2008.

IPCC: Climate Change 2013: The Physical Science Basis. The Fifth Assessment Report of the Intergovernmental Panel on Climate Change, , Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Jiménez, P. A., Dudhia, J., González-Rouco, J. F., Navarro, J., Montávez, J. P. and García-Bustamante, E.: A Revised Scheme for the WRF Surface Layer Formulation, Monthly Weather Review, 140(3), 898–918, 2012.

Jung, J., Lee, J., Kim, B. and Oh, S.: Seasonal variations in the NO2 artifact from chemiluminescence measurements with a molybdenum converter at a suburban site in Korea (downwind of the Asian continental outflow) during 2015–2016, Atmospheric Environment, 165, 290–300, doi:10.1016/j.atmosenv.2017.07.010, 2017.

Khaniabadi, Y. O., Goudarzi, G., Daryanoosh, S. M., Borgini, A., Tittarelli, A. and De Marco, A.: Exposure to PM10, NO2, and O3 and impacts on human health, Environmental Science and Pollution Research, 24(3), 2781–2789, doi:10.1007/s11356-016-8038-6, 2017.

Khattatov, B. V., Gille, J. C., Lyjak, L. V., Brasseur, G. P., Dvortsov, V. L., Roche, A. E. and Waters, J. W.: Assimilation of photochemically active species and a case analysis of UARS data, J. Geophys. Res., 104(D15), 18715–18737, doi:10.1029/1999JD900225, 1999.

Khattatov, B. V., Lamarque, J.-F., Lyjak, L. V., Menard, R., Levelt, P., Tie, X., Brasseur, G. P. and Gille, J. C.: Assimilation of satellite observations of long-lived chemical species in global chemistry transport models, J. Geophys. Res., 105(D23), 29135–29144, doi:10.1029/2000JD900466, 2000.

Kim, H. S., Park, I., Song, C. H., Lee, K., Yun, J. W., Kim, H. K., Jeon, M., Lee, J. and Han, K. M.: Development of a daily PM10 and PM2.5 prediction system using a deep long short-term memory neural network model, Atmos. Chem. Phys., 19(20), 12935–12951, doi:10.5194/acp-19-12935-2019, 2019.

Lamarque, J.-F., Khattatov, B. V., Gille, J. C. and Brasseur, G. P.: Assimilation of Measurement of Air Pollution from Space (MAPS) CO in a global three-dimensional model, J. Geophys. Res., 104(D21), 26209–26218, doi:10.1029/1999JD900807, 1999.

Lee, C., Martin, R. V., Donkelaar, A. van, Lee, H., Dickerson, R. R., Hains, J. C., Krotkov, N., Richter, A., Vinnikov, K. and Schwab, J. J.: SO2 emissions and lifetimes: Estimates from

inverse modeling using in situ and global, space-based (SCIAMACHY and OMI) observations, Journal of Geophysical Research: Atmospheres, 116(D6), doi:10.1029/2010JD014758, 2011.

Lee, J., Kim, J., Song, C. H., Ryu, J.-H., Ahn, Y.-H. and Song, C. K.: Algorithm for retrieval of aerosol optical properties over the ocean from the Geostationary Ocean Color Imager, Remote Sensing of Environment, 114(5), 1077–1088, doi:10.1016/j.rse.2009.12.021, 2010.

Lee, J., Kim, J., Yang, P. and Hsu, N. C.: Improvement of aerosol optical depth retrieval from MODIS spectral reflectance over the global ocean using new aerosol models archived from AERONET inversion data and tri-axial ellipsoidal dust database, Atmos. Chem. Phys., 12(15), 7087–7102, doi:10.5194/acp-12-7087-2012, 2012.

Lee, S., Song, C. H., Park, R. S., Park, M. E., Han, K. M., Kim, J., Choi, M., Ghim, Y. S. and Woo, J.-H.: GIST-PM-Asia v1: development of a numerical system to improve particulate matter forecasts in South Korea using geostationary satellite-retrieved aerosol optical data over Northeast Asia, Geosci. Model Dev., 9(1), 17–39, doi:10.5194/gmd-9-17-2016, 2016.

Levelt, P. F., Khattatov, B. V., Gille, J. C., Brasseur, G. P., Tie, X. X. and Waters, J. W.: Assimilation of MLS ozone measurements in the global three-dimensional chemistry transport model ROSE, Geophys. Res. Lett., 25(24), 4493–4496, doi:10.1029/1998GL900152, 1998.

Lorenc, A. C.: Analysis methods for numerical weather prediction, Q.J.R. Meteorol. Soc., 112(474), 1177–1194, doi:10.1002/qj.49711247414, 1986.

Louis, J.-F.: A parametric model of vertical eddy fluxes in the atmosphere, Boundary-Layer Meteorol, 17(2), 187–202, doi:10.1007/BF00117978, 1979.

Malm, W. C. and Hand, J. L.: An examination of the physical and optical properties of aerosols collected in the IMPROVE program, Atmospheric Environment, 41(16), 3407–3427, doi:10.1016/j.atmosenv.2006.12.012, 2007.

Martin, R. V., Jacob, D. J., Yantosca, R. M., Chin, M. and Ginoux, P.: Global and regional decreases in tropospheric oxidants from photochemical effects of aerosols, J. Geophys. Res., 108(D3), 4097, doi:10.1029/2002JD002622, 2003.

NCL: The NCAR Command Language (Version 6.6.2) [Software]. Boulder, Colorado: UCAR/NCAR/CISL/TDD. http://dx.doi.org/10.5065/D6WD3XH5, 2019.

Niu, G.-Y., Yang, Z.-L., Mitchell, Kenneth. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M. and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, J. Geophys. Res., 116(D12109), doi:10.1029/2010JD015140, 2011.

Otte, T. L. and Pleim, J. E.: The Meteorology-Chemistry Interface Processor (MCIP) for the

800 CMAQ modeling system: updates through MCIPv3.4.1, Geoscientific Model
801 Development, 3(1), 243–256, doi:https://doi.org/10.5194/gmd-3-243-2010, 2010.

802 Park, M. E., Song, C. H., Park, R. S., Lee, J., Kim, J., Lee, S., Woo, J.-H., Carmichael, G. R.,
803 Eck, T. F., Holben, B. N., Lee, S.-S., Song, C. K. and Hong, Y. D.: New approach to
804 monitor transboundary particulate pollution over Northeast Asia, Atmos. Chem. Phys.,
805 14(2), 659–674, doi:10.5194/acp-14-659-2014, 2014.

806 Park, R. S., Song, C. H., Han, K. M., Park, M. E., Lee, S.-S., Kim, S.-B. and Shimizu, A.: A
807 study on the aerosol optical properties over East Asia using a combination of CMAQ-
808 simulated aerosol optical properties and remote-sensing data via a data assimilation
809 technique, Atmos. Chem. Phys., 11(23), 12275–12296, doi:10.5194/acp-11-12275-2011,
810 2011.

811 Penner, J. E., Dong, X. and Chen, Y.: Observational evidence of a change in radiative forcing
812 due to the indirect aerosol effect, Nature, 427(6971), 231–234, doi:10.1038/nature02234,
813 2004.

814 Pleim, J. E.: A Combined Local and Nonlocal Closure Model for the Atmospheric Boundary
815 Layer. Part I: Model Description and Testing, J. Appl. Meteor. Climatol., 46(9), 1383–
816 1395, doi:10.1175/JAM2539.1, 2007a.

817 Pleim, J. E.: A Combined Local and Nonlocal Closure Model for the Atmospheric Boundary
818 Layer. Part II: Application and Evaluation in a Mesoscale Meteorological Model, J. Appl.
819 Meteor. Climatol., 46(9), 1396–1409, doi:10.1175/JAM2534.1, 2007b.

820 Pleim, J. E. and Xiu, A.: Development of a Land Surface Model. Part II: Data Assimilation, J.
821 Appl. Meteor., 42(12), 1811–1822, doi:10.1175/1520-
822 0450(2003)042<1811:DOALSM>2.0.CO;2, 2003.

823 Scott, C. E., Rap, A., Spracklen, D. V., Forster, P. M., Carslaw, K. S., Mann, G. W., Pringle, K.
824 J., Kivekäs, N., Kulmala, M., Lihavainen, H. and Tunved, P.: The direct and indirect
825 radiative effects of biogenic secondary organic aerosol, Atmos. Chem. Phys., 14(1), 447–
826 470, doi:10.5194/acp-14-447-2014, 2014.

827 Skamarock, C., Klemp, B., Dudhia, J., Gill, O., Barker, D., Duda, G., Huang, X., Wang, W. and
828 Powers, G.: A Description of the Advanced Research WRF Version 3, ,
829 doi:10.5065/D68S4MVH, 2008.

830 Tang, Y., Chai, T., Pan, L., Lee, P., Tong, D., Kim, H.-C. and Chen, W.: Using optimal
831 interpolation to assimilate surface measurements and satellite AOD for ozone and PM2.5:
832 A case study for July 2011, Journal of the Air & Waste Management Association, 65(10),
833 1206–1216, doi:10.1080/10962247.2015.1062439, 2015.

834 Tang, Y., Pagowski, M., Chai, T., Pan, L., Lee, P., Baker, B., Kumar, R., Delle Monache, L.,
835 Tong, D. and Kim, H.-C.: A case study of aerosol data assimilation with the Community
836 Multi-scale Air Quality Model over the contiguous United States using 3D-Var and

837   optimal interpolation methods, Geosci. Model Dev., 10(12), 4743–4758,
838   doi:10.5194/gmd-10-4743-2017, 2017.

839 Wiedinmyer, C., Quayle, B., Geron, C., Belote, A., McKenzie, D., Zhang, X., O'Neill, S. and
840   Wynne, K. K.: Estimating emissions from fires in North America for air quality modeling,
841   Atmospheric Environment, 40(19), 3419–3432, doi:10.1016/j.atmosenv.2006.02.010,
842   2006.

843 Wiedinmyer, C., Akagi, S., Yokelson, R., Emmons, L., Al-Saadi, J., Orlando, J. and Soja, A.:
844   The Fire INventory from NCAR (FINN): A High Resolution Global Model to Estimate
845   the Emissions from Open Burning, Geoscientific Model Development, 625–641, 2011.

846 Woo, J.-H., Choi, K.-C., Kim, H. K., Baek, B. H., Jang, M., Eum, J.-H., Song, C. H., Ma, Y.-
847   I., Sunwoo, Y., Chang, L.-S. and Yoo, S. H.: Development of an anthropogenic emissions
848   processing system for Asia using SMOKE, Atmospheric Environment, 58, 5–13,
849   doi:10.1016/j.atmosenv.2011.10.042, 2012.

850 Xiu, A. and Pleim, J. E.: Development of a Land Surface Model. Part I: Application in a
851   Mesoscale Meteorological Model, J. Appl. Meteor., 40(2), 192–209, doi:10.1175/1520-
852   0450(2001)040<0192:DOALSM>2.0.CO;2, 2001.

853 Yang, Z.-L., Niu, G.-Y., Mitchell, Kenneth. E., Chen, F., Ek, M. B., Barlage, M., Longuevergne,
854   L., Manning, K., Niyogi, D., Tewari, M. and Xia, Y.: The community Noah land surface
855   model with multiparameterization options (Noah-MP): 2. Evaluation over global river
856   basins, J. Geophys. Res., 116(D12110), doi:10.1029/2010JD015139, 2011.

857 Yu, H., Dickinson, R. E., Chin, M., Kaufman, Y. J., Holben, B. N., Geogdzhayev, I. V. and
858   Mishchenko, M. I.: Annual cycle of global distributions of aerosol optical depth from
859   integration of MODIS retrievals and GOCART model simulations, J. Geophys. Res.,
860   108(D3), 4128, doi:10.1029/2002JD002717, 2003.

861 Yuan, H., Dai, Y., Xiao, Z., Ji, D. and Shangguan, W.: Reprocessing the MODIS Leaf Area
862   Index products for land surface and climate modelling, Remote Sensing of Environment,
863   115(5), 1171–1187, doi:10.1016/j.rse.2011.01.001, 2011.

864

865

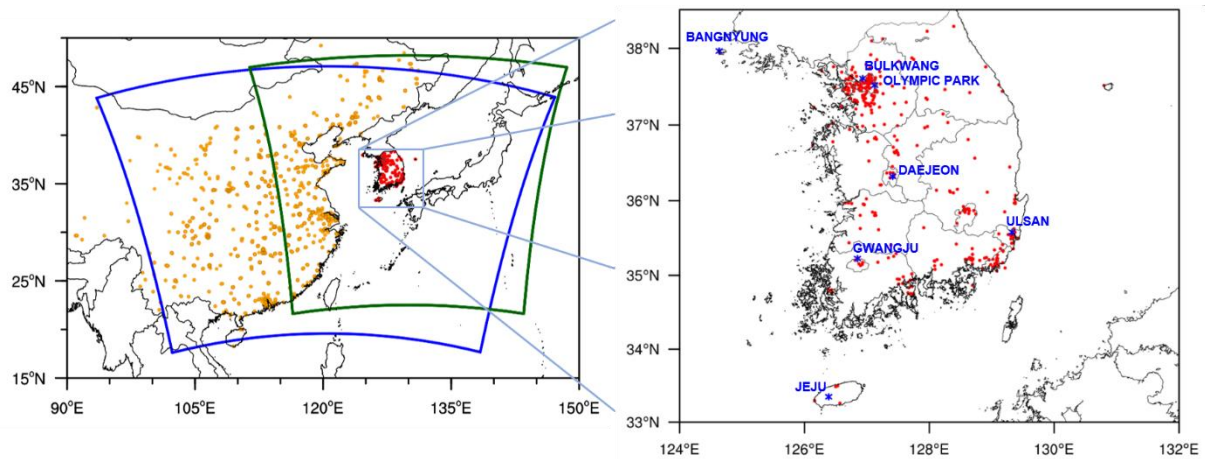
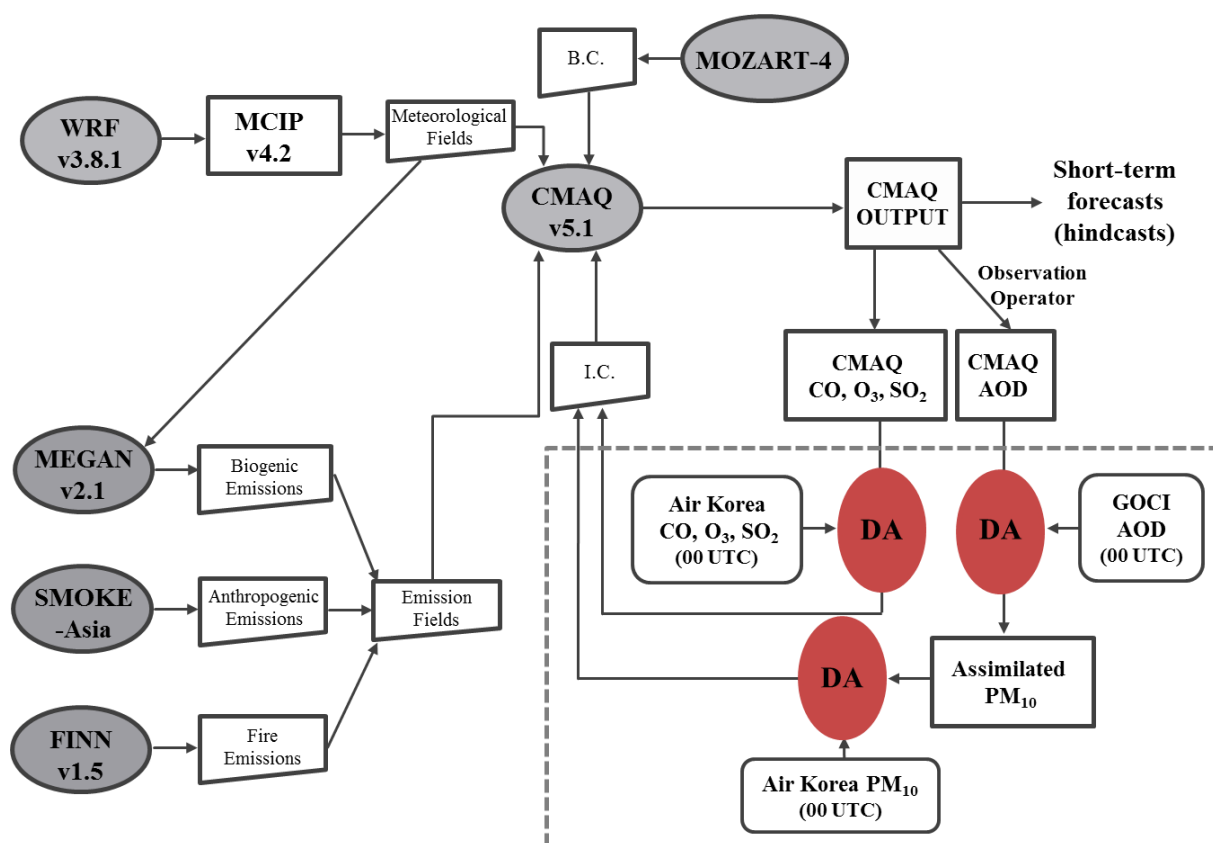
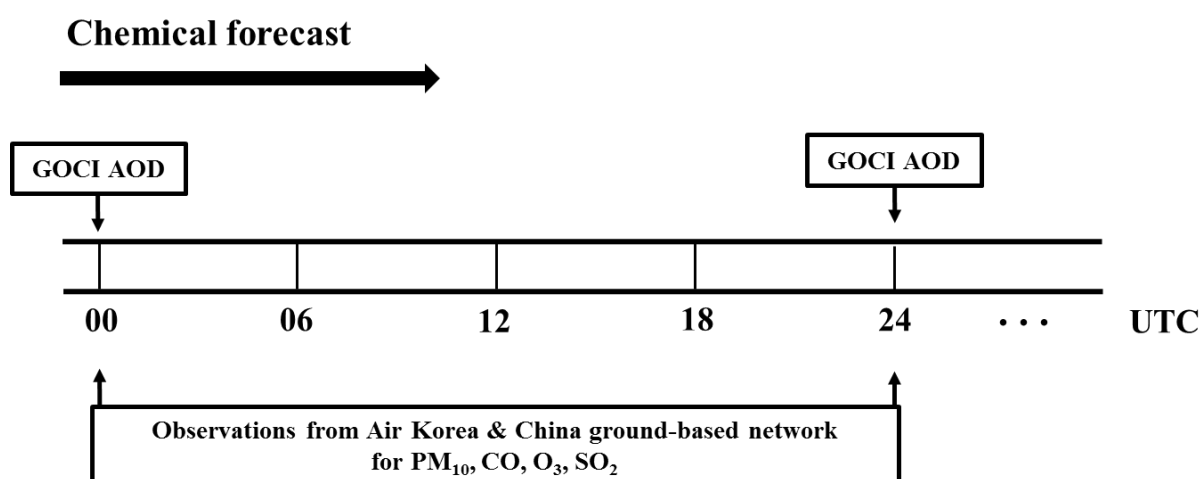
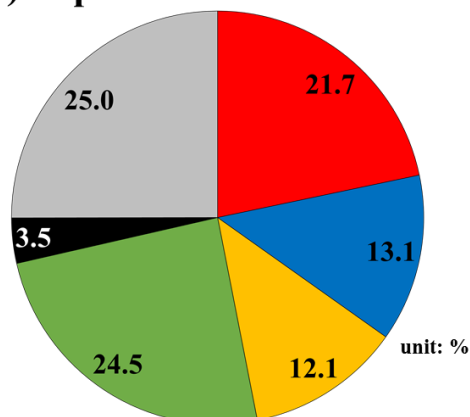



Figure 1. Domains of GOCI sensor (dark green line) and CMAQ model simulations (blue line). Red-colored dots denote the locations of Air Korea sites in South Korea. Orange-colored dots represent the locations of ground-based observation stations in China. Blue stars show the locations of seven super-sites in South Korea. During the KORUS-AQ campaign, observation data were obtained from 1514 stations in China as well as 264 Air Korea and seven super-site stations in South Korea. NCL (2019) was used to draw this figure.

872

873

874 **Figure 2.** Schematic diagram of the Korean air quality prediction system developed in this
875 study. The initial conditions (ICs) of the CMAQ model simulations are prepared by assimilating
876 CMAQ outputs with satellite-retrieved and ground-measured observations. The data process
877 for preparing the ICs is shown in the box with gray-dashed lines.

878

879 **Figure 3.** Schematic diagram of the Korean air quality prediction system for particulate matter
880 (PM) and gas-phase pollutants. The data assimilation (DA) cycle is 24 hours for both PM and
881 gas-phase pollutants such as CO, O_3, and SO_2. The DA of NO_2 is excluded in the current study,
882 the reason for which is discussed in the text.

(a) Super-site

(b) CMAQ

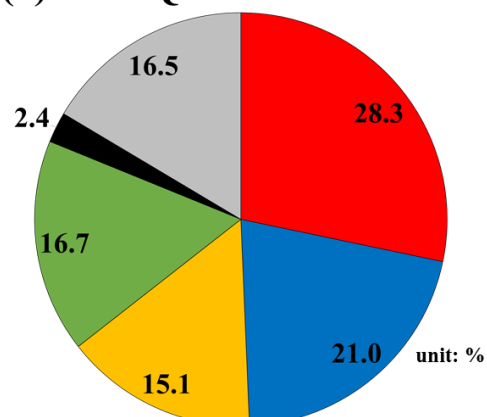
Average PM$_{2.5}$ = 28 μg/m^3 Average PM$_{2.5}$ = 19.9 μg/m^3

■ **SULFATE** ■ **NITRATE** ■ **AMMONIUM** ■ **OA** ■ **EC** □ **OTHERS**

Figure 4. Average PM$_{2.5}$ composition (a) observed at the super-site stations and (b) simulated by the CMAQ model during the KORUS-AQ campaign. The averaged PM$_{2.5}$ measured from the super-sites and calculated from the CMAQ model simulations over the period of the KORUS-AQ campaign are 28 μg/m^3 and 19.9 μg/m^3, respectively. The mass of organic aerosols (OAs) was calculated by multiplying organic carbon mass by 1.6.

889

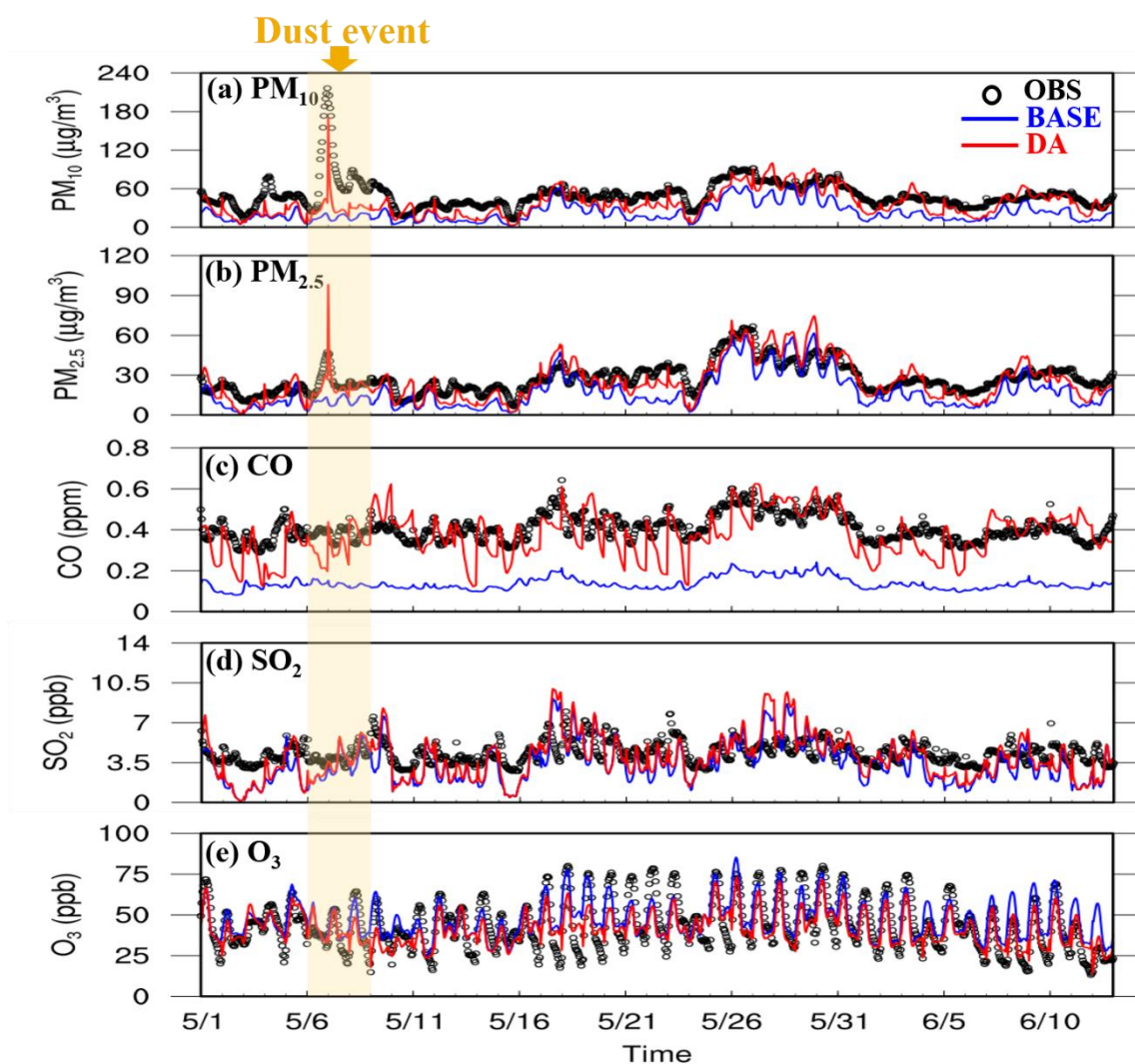
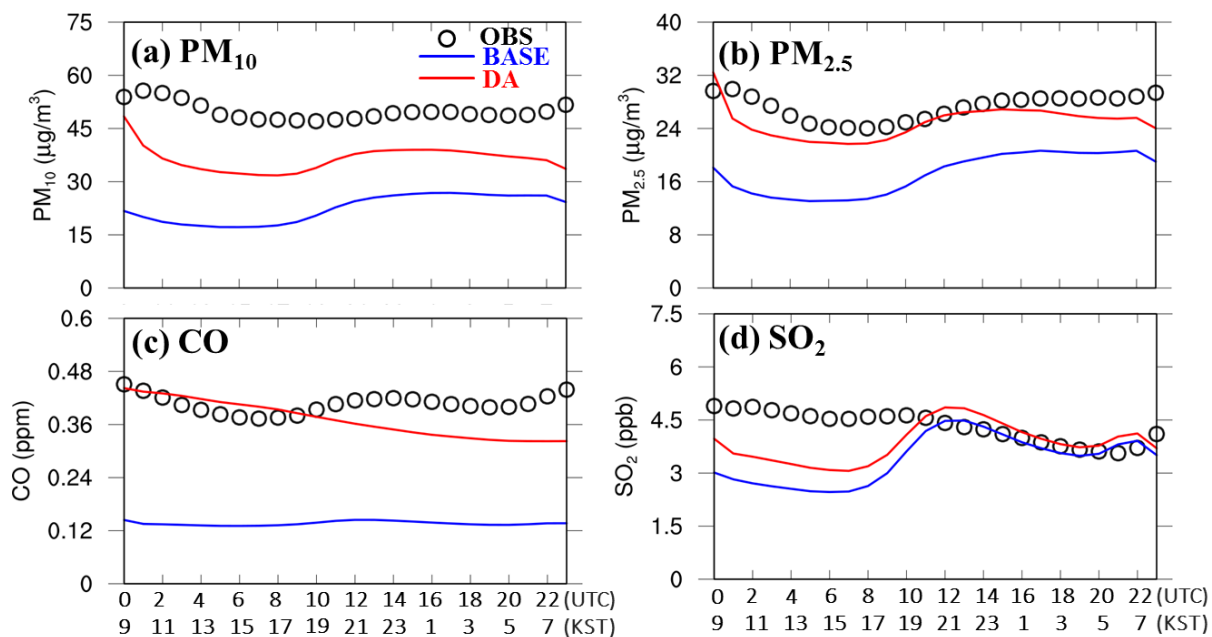



Figure 5. Time-series plots of hourly (a) PM_{10}, (b) $PM_{2.5}$, (c) CO, (d) SO_2, and (e) O_3 concentrations at 264 Air Korea stations. Black open circles (OBS) represent the observed concentrations. Blue and red lines show the results simulated from the BASE RUN and DA RUN over South Korea, respectively.

894

895 **Figure 6.** Aggregated average concentrations of (a) PM_{10}, (b) $PM_{2.5}$, (c) CO, and (d) SO_2 at
896 264 Air Korea stations over the KORUS-AQ campaign period. Open black circles denote the
897 observations obtained from 264 Air Korea stations in South Korea. Blue and red lines represent
898 the predicted concentrations from the BASE RUN and DA RUN, respectively. The DA was
899 conducted at 00:00 UTC every day throughout the KORUS-AQ campaign period.

900

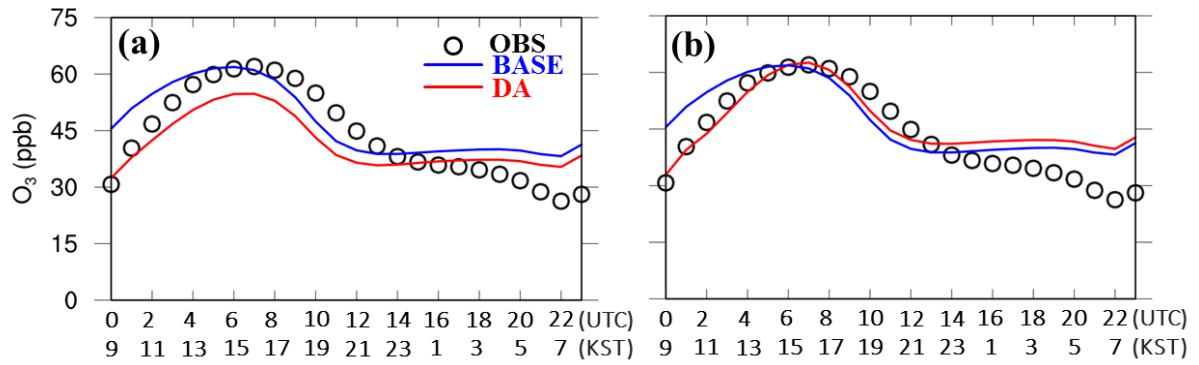



Figure 7. Comparison of CMAQ-simulated O_3 mixing ratios (BASE RUN with blue lines and DA RUN with red lines) with O_3 mixing ratios from Air Korea stations (open black circles). DA RUN was carried out by assimilating CMAQ outputs with Air Korea observations using (a) only O_3 mixing ratios and (b) both O_3 and NO_2 mixing ratios.

905

906

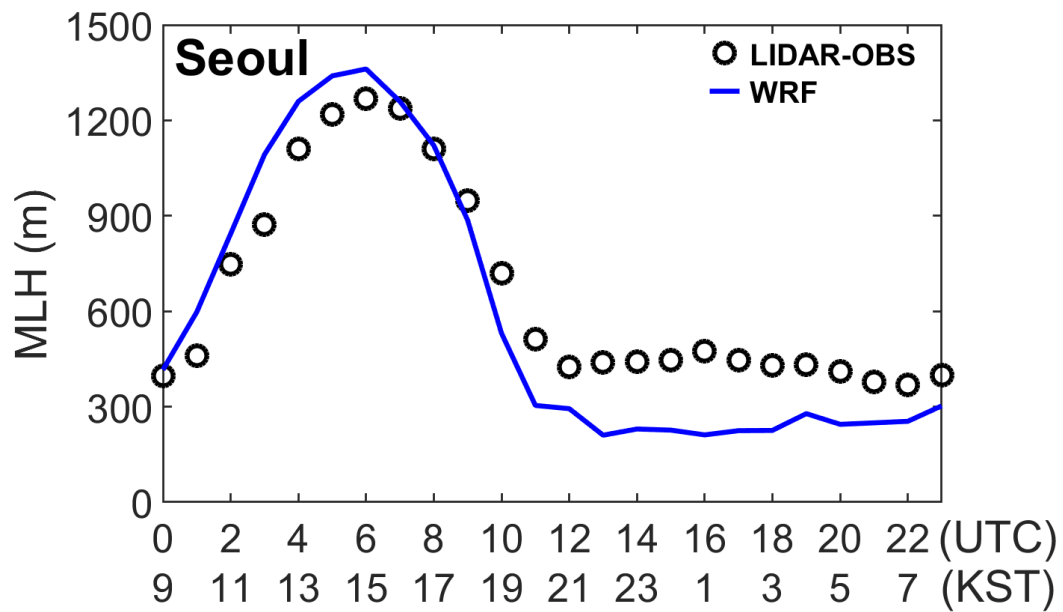



Figure 8. Comparison of WRF-simulated mixing layer height (MLH) (denoted by blue-dashed
line) with lidar-measured MLH (denoted by open black circles) at Seoul National University
(SNU) in Seoul. KST stands for Korean standard time.

910

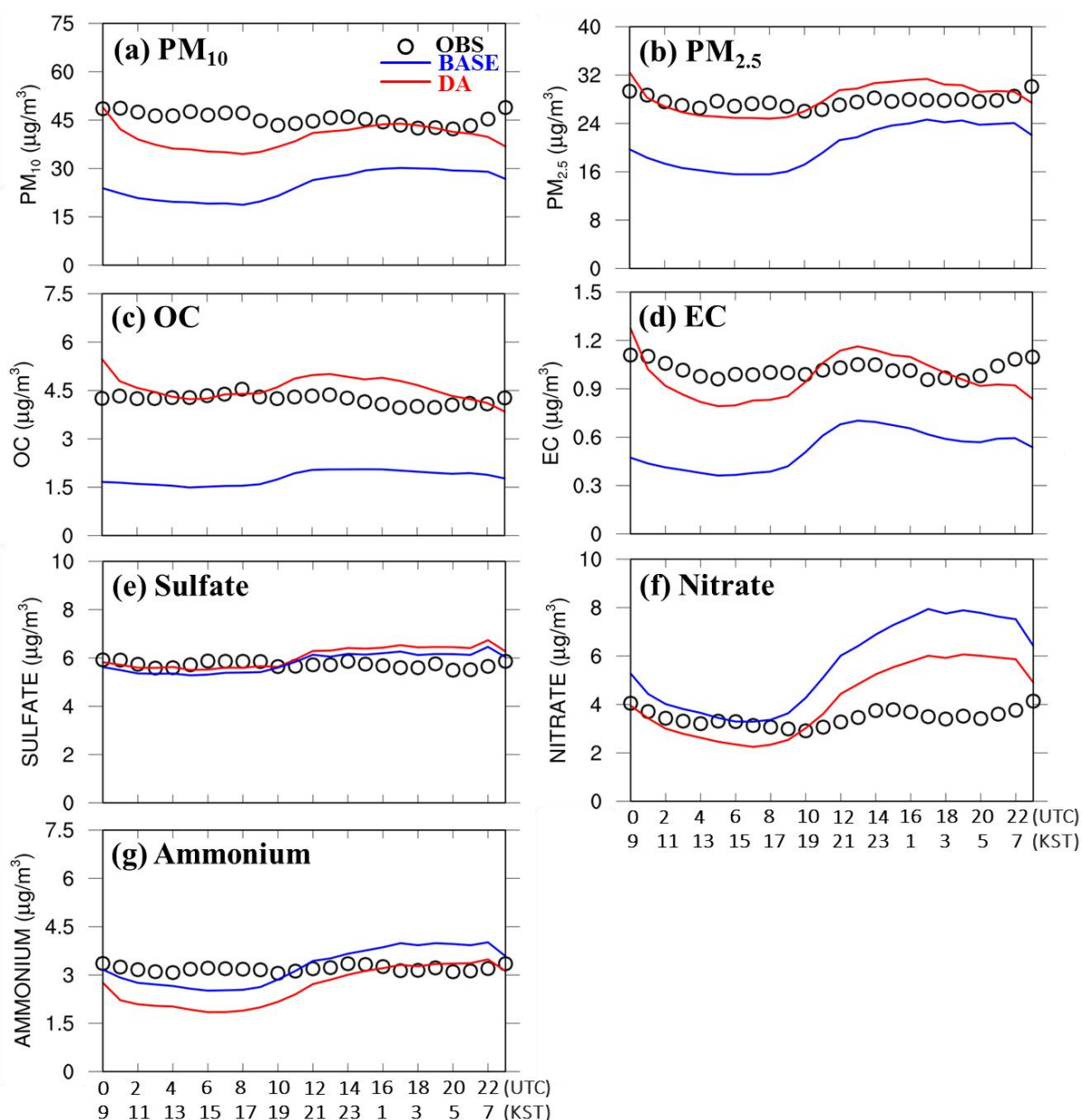



Figure 9. Aggregated average concentrations of (a) PM$_{10}$, (b) PM$_{2.5}$, (c) OC, (d) EC, (e) sulfate, (f) nitrate, and (g) ammonium as predicted by CMAQ model during the period of the KORUS-AQ campaign. The others are the same as those shown in Fig. 7, except for the fact that the observation data used here were obtained from the seven super-site stations in South Korea.

915

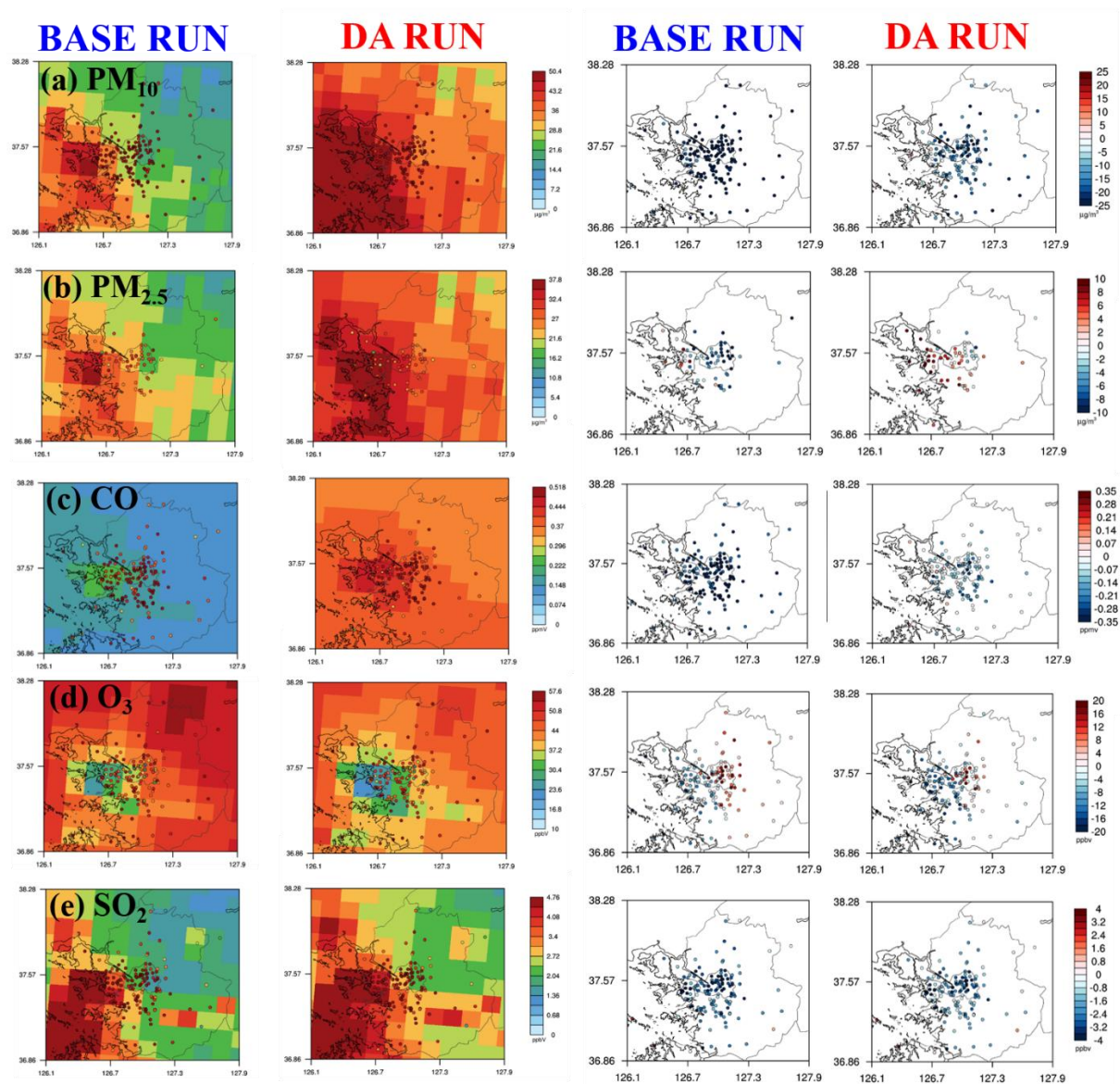



Figure 10. Spatial distributions (first and second columns) and bias (third and fourth columns) of (a) PM₁₀, (b) PM₂.₅, (c) CO, (d) SO₂, and (e) O₃ over Seoul Metropolitan Area (SMA) for the entire period of the KORUS-AQ campaign. Colored circles of first and second columns represent the concentrations of the air pollutants observed at the Air Korea stations in the SMA.

920

921

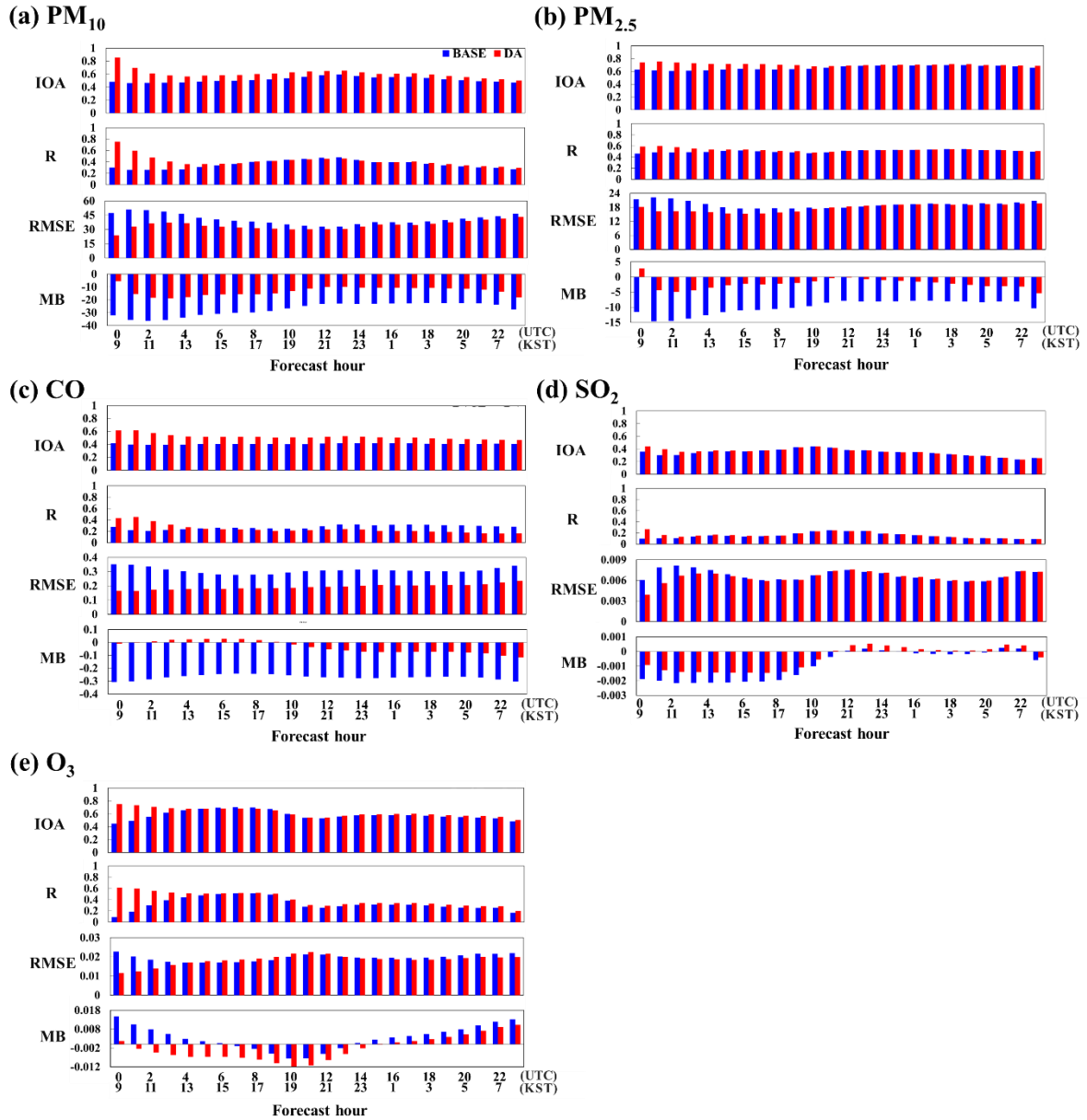



Figure 11. Time-series plots of four performance metrics (IOA, R, RMSE, and MB) for (a) PM_{10}, (b) $PM_{2.5}$, (c) CO, (d) SO_2, and (e) O_3 forecasts. The DA was conducted at 00:00 UTC. The units of RMSE and MB are $\mu g/m^3$ and ppmv for PM concentrations and for gaseous species, respectively. The definitions of the four performance metrics are shown in Appendix A.

926

927

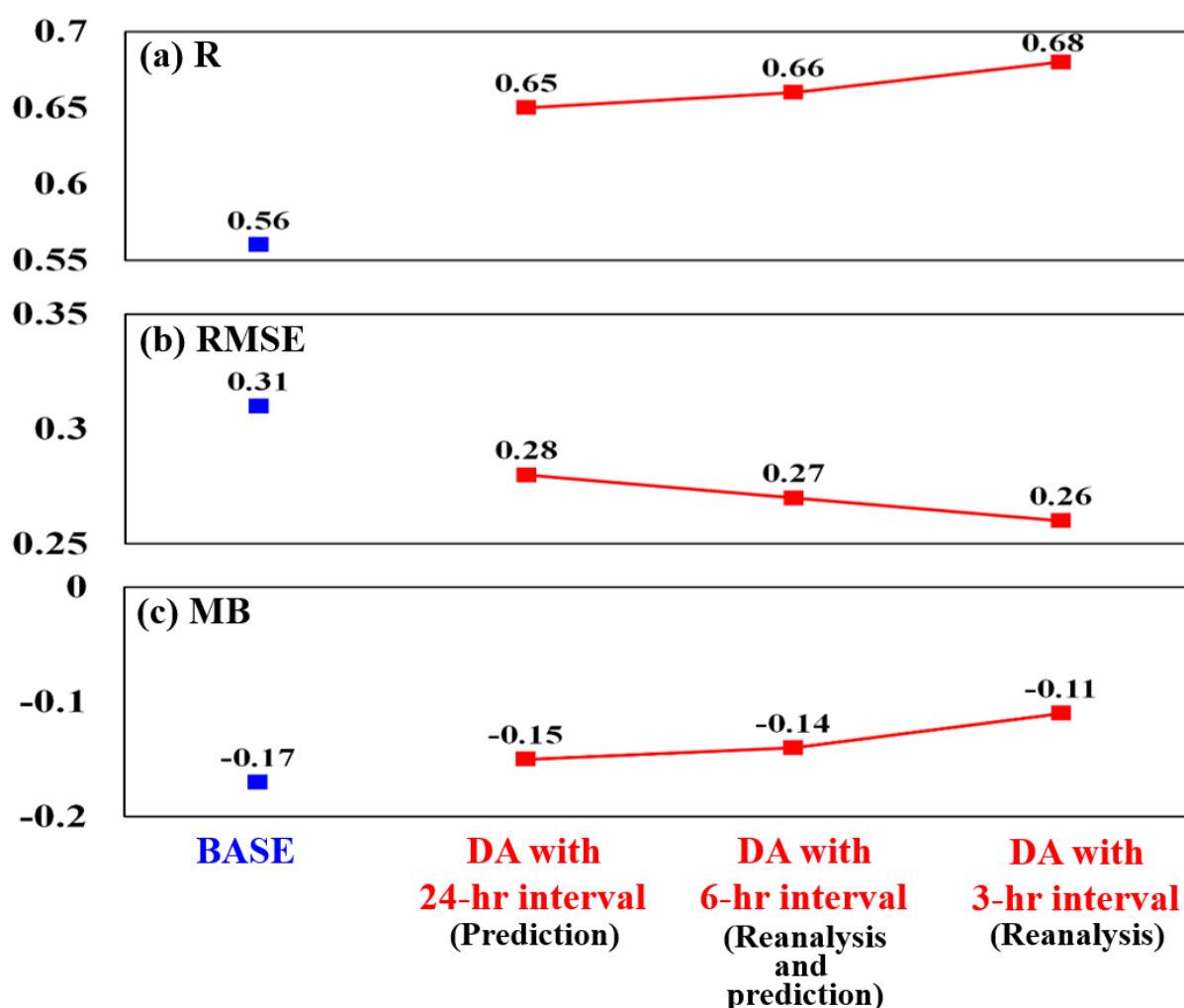



Figure 12. Variations of three performance metrics (R, RMSE, and MB) with time-intervals of
data assimilations. For these tests, the GOCI AODs were used in the DA to update the initial
conditions of the CMAQ model simulations. The results from the three CMAQ model
simulations were compared with AERONET AODs ("ground truth"). The two blue squares
represent the performances from the BASE RUNs and the red squares indicate the
performances from the DA RUNs. The three experiments were carried out with the assimilation
time-intervals of 24, 6, and 3 hours (hr), respectively. Here, the DA RUN with the 24-hr time-
interval is referred to as "air quality prediction", and the DA RUNs with the 6-hr and 3-hr time-
interval are referred to as "air quality reanalysis".

937

938

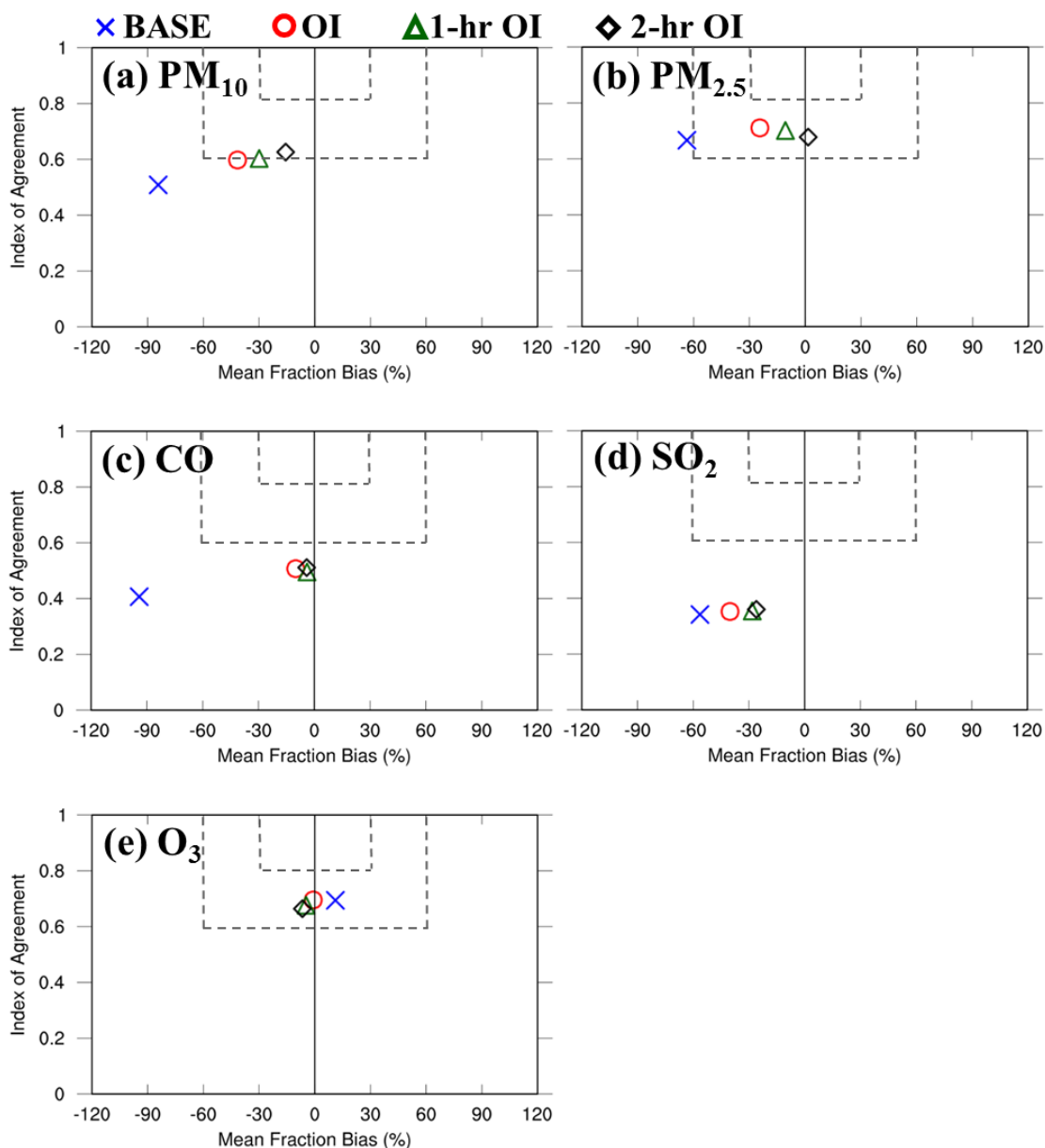



Figure 13. Soccer plot analyses for (a) PM_{10}, (b) $PM_{2.5}$, (c) CO, (d) SO_2, and (e) O_3. The CMAQ-predicted concentrations were compared with the Air Korea observations. Blue crosses, red circles, dark-green triangles, and black diamonds represent the performances calculated from the BASE RUN, the DA RUNs with the OI system, the 1-hour (hr) OI system, and the 2-hr OI system, respectively.

944 **Table 1.** Statistical metrics from BASE RUN and DA RUN with Air Korea observations over
945 the entire period of the KORUS-AQ campaign.

	PM_{10}		$PM_{2.5}$		CO		SO_2		O_3	
	BASE RUN	DA RUN	BASE RUN	DA RUN	BASE RUN	DA RUN	BASE RUN	DA RUN	BASE RUN	DA RUN
N	101852		65383		101764		101764		101836	
IOA	0.51	0.60	0.67	0.71	0.41	0.51	0.34	0.35	0.69	0.70
R	0.34	0.40	0.51	0.53	0.28	0.21	0.14	0.15	0.50	0.52
RMSE	40.8	34.87	19.2	17.83	0.31	0.19	0.0068	0.0066	0.020	0.02
MB	-27.2	-13.54	-9.9	-2.43	-0.27	-0.04	-0.0009	-0.0004	0.003	-0.0024
ME	30.1	24.20	15.3	13.48	0.27	0.15	0.004	0.0034	0.015	0.015
MNB	-50.0	-18.17	-30.1	5.32	-62.0	3.14	3.1	17.77	48.0	30.22
MNE	60.7	52.35	62.6	62.77	62.9	40.67	93.1	93.56	70.2	61.34
MFB	-84.3	-41.61	-63.6	-24.41	-94.1	-10.00	-56.4	-40.20	11.1	-0.82
MFE	91.1	62.32	81.6	60.01	94.9	39.49	91.4	82.91	40.7	40.64

946

947 **Table 2.** Statistical metrics from BASE RUN and DA RUN with Air Korea observations at
948 00:00 UTC when the DA was conducted during the KORUS-AQ campaign.

	PM_{10}		$PM_{2.5}$		CO		SO_2		O_3	
	BASE RUN	DA RUN	BASE RUN	DA RUN	BASE RUN	DA RUN	BASE RUN	DA RUN	BASE RUN	DA RUN
N	1057		695		1024		1007		1043	
IOA	0.48	0.86	0.63	0.74	0.41	0.62	0.36	0.44	0.45	0.75
R	0.30	0.75	0.46	0.59	0.28	0.43	0.097	0.27	0.09	0.61
RMSE	47.2	23.92	21.5	18.21	0.35	0.16	0.0061	0.0039	0.023	0.012
MB	-32.2	-5.46	-11.5	2.80	-0.31	-0.01	-0.0019	-0.0009	0.015	0.002
ME	34.5	16.03	17.2	13.25	0.31	0.12	0.0039	0.0023	0.018	0.009
MNB	-54.9	-0.53	-33.2	26.17	-64.3	9.69	-20.1	7.35	100.4	27.45
MNE	64.0	36.07	63.1	59.77	64.8	30.69	86.7	55.27	107.8	43.81
MFB	-92.8	-13.38	-67.3	0.56	-98.7	1.81	-75.9	-17.39	43.7	12.16
MFE	98.8	38.41	84.3	48.30	99.1	27.14	99.9	56.23	52.9	31.53

949