Chemical data assimilation with aqueous chemistry in WRF-Chem coupled with WRFDA (V4.4.1)

Soyoung Ha

[1]**National Center for Atmospheric Research, Boulder, CO, USA**

Rajesh Kumar

[1]**National Center for Atmospheric Research, Boulder, CO, USA**

Gabriele Pfister

[1]**National Center for Atmospheric Research, Boulder, CO, USA**

Yonghee Lee

[2]**Air Quality Forecasting Center, National Institute of Environmental Research, Incheon, Republic of Korea**

Daegyun Lee

[2]**Air Quality Forecasting Center, National Institute of Environmental Research, Incheon, Republic of Korea**

Hyun Mee Kim

[3]**Department of Atmospheric Sciences, Yonsei University, Seoul, Republic of Korea**

Young-Hee Ryu

[3]**Department of Atmospheric Sciences, Yonsei University, Seoul, Republic of Korea**

Key Points:

- The WRF-Chem/WRFDA 3D-Var system (V4.4.1) is extended for chemical data assimilation with aqueous chemistry using the RACM-MADE-VBS scheme.
- Surface $PM_{2.5}$, PM_{10}, SO_2, NO_2, O_3, and CO concentrations are assimilated in the coupled system.
- The inclusion of aqueous-phase aerosols in air quality cycling enhances the simulation of aerosol wet deposition in cloudy conditions.

Corresponding author: Soyoung Ha, syha@ucar.edu

Abstract

This study introduces a new chemistry option in the Weather Research and Forecasting model data assimilation (WRFDA) system, coupled with the WRF-Chem model (Version 4.4.1), to incorporate aqueous chemistry (AQCHEM) in the assimilation of ground-level chemical measurements. The new data assimilation capability includes the integration of aqueous-phase aerosols from the Regional Atmospheric Chemistry Mechanism (RACM) gas chemistry, the Modal Aerosol Dynamics Model for Europe (MADE) aerosol chemistry, and the Volatility Basis Set (VBS) for secondary organic aerosol (SOA) production. The RACM-MADE-VBS-AQCHEM scheme facilitates aerosol-cloud-precipitation interactions by activating aerosol particles in cloud water during the model simulation. With the goal of enhancing air quality forecasting in cloudy conditions, this new implementation is demonstrated in the weakly coupled three-dimensional variational data assimilation (3D-Var) system through regional air quality cycling over East Asia. Surface particulate matter (PM) concentrations and four gas species (SO_2, NO_2, O_3, and CO) are assimilated every 6 h for the month of March 2019. The results show that including aqueous-phase aerosols in both the analysis and forecast can represent aerosol wet removal processes associated with cloud development and rainfall production. During a pollution event with high cloud cover, simulations without aerosols defined in cloud water exhibit significantly higher values for liquid water path (LWP), and surface PM_{10} ($PM_{2.5}$) concentrations are overestimated by a factor of 10 (3) when wet scavenging processes dominate. On the contrary, aqueous chemistry proves to be helpful in simulating the wet deposition of aerosols, accurately predicting the evolution of surface PM concentrations without such overestimation.

Plain Language Summary

Major air pollution events over the Korean peninsula are often observed in association with significant cloud cover, especially over the Yellow Sea to the west of the peninsula. Cloudy conditions pose challenges for both remote sensing observations and model predictions, but the inclusion of aqueous-phase (or cloud-borne) aerosols in the WRF-Chem/WRFDA system improves the simulation of aerosol wet scavenging, leading to improved predictions of surface particulate matter concentrations that were otherwise substantially overestimated.

Keywords

Aerosol data assimilation, aqueous chemistry, wet deposition

1 Introduction

Poor air quality, characterized by high concentrations of particulate matter (PM) at ground level, is often accompanied by extensive cloud cover Eck et al. (2018, 2020), posing challenges for both observation and prediction. Given the short lifetime of aerosol species (up to one week) and the large uncertainties in modeling atmospheric composition, improving initialization is crucial for enhancing short-range air quality forecasting. Data assimilation (DA) incorporates available observations into a numerical prediction model to produce initial conditions that can lead to accurate forecasts. The quality of the initial condition, or the analysis, largely depends on the quantity of reliable observations and the accuracy of the forecast model.

Various efforts have been recently made towards chemical data assimilation, but the utilization of advanced chemistry schemes remains limited, especially for the prediction of particulate matter concentrations (Chen et al., 2019; Sun et al., 2020; Ha, 2022). Hence, it is fair to say that chemical data assimilation, especially aerosol data assimi-

lation focusing on aerosol species, is still in its early stages, lagging behind meteorological or oceanographic applications (Baklanov et al., 2014; Bocquet et al., 2015). Major challenges specific to aerosol data assimilation can be described as follows: 1) The limited information content of atmospheric composition observations, which often lack accuracy and coverage, especially in cloudy conditions. 2) Large uncertainties or systematic errors in chemical transport models, partly due to significant uncertainties in forcing parameters such as emissions and partly due to imperfect representation of complex chemical processes and their interactions with the atmospheric environment. 3) Surface PM concentrations, major indicators of the air quality index, are only computed diagnostically at the end of the model integration to account for all contributions from multiple aerosol species. Therefore, the data assimilation system should handle all the prognostic aerosol species that contribute to the estimation of PM concentrations and determine how to distribute analysis increments in ground PM concentrations back to the three-dimensional aerosol variables. Since the number of aerosol species predicted in the model is typically larger than the number of observed variables, it becomes an underconstrained problem where a unique solution is not guaranteed. Moreover, the use of sophisticated chemistry schemes means a large number of prognostic variables, rendering aerosol analysis a high-dimensional problem that requires substantial computational resources. This makes the three-dimensional variational data assimilation (3D-Var) algorithm still attractive due to its speed and operational simplicity, despite limitations such as static background error covariance and the ignorance of model errors. 4) The model configuration for aerosol chemistry coupled with meteorology, the first step toward aerosol cycling, is typically more complicated than that of a weather prediction model alone. In an online coupled system, the dynamics and physics configurations for meteorology can greatly impact the reliability and performance of chemical simulations. Therefore, careful consideration should be given to the weather component of the configuration as well. 5) Most interfaces to input data and chemical processes are highly customized for specific gas and aerosol chemistry schemes. This is because each chemistry parameterization defines its own unique set of prognostic variables that are not interchangeable with other schemes (Pfister et al., 2020). Consequently, each chemistry scheme requires its own preparation of input forcing data, such as anthropogenic, biogenic and biomass burning emissions. In the context of data assimilation, this also entails developing a new interface within the variational data assimilation system. This includes creating new observation operators (that compute the model correspondents from the aerosol species defined in the scheme), as well as their tangent linear and adjoint models, and estimating background error covariance.

Attempts to develop coupled data assimilation between chemistry and meteorology for regional forecast applications have been limited, despite the widespread recognition of high correlations between the two components in the modeling community (Baklanov et al., 2017). In the context of 3D-Var, fully (or strongly) coupled data assimilation is still challenging because many factors contributing to cross-covariance between meteorological and chemical variables are highly variable in time and space. When it comes to real observations, developments and research on coupled data assimilation have primarily focused on trace gases such as ozone and carbon monoxide, often using simplified background error covariance (Ménard et al., 2019). On the other hand, in weakly coupled data assimilation, the coupling occurs during the model integration by using a two-way coupled forecast model, but not through the analysis. Observations are assimilated in each component (e.g., meteorology or chemistry) to update the analysis variables independently, which are then used together to initialize the coupled modeling system for prediction. The background error covariance is estimated from the forecasts produced by the fully coupled model, capturing the coupling aspects from the model simulations rather than the cross-covariance component. As a result, the direct influence of observations is limited to each component of the model.

Aerosol particles play a crucial role in various key processes related to atmospheric chemistry and physics (Rosenfeld et al. (2008); Stevens and Feingold (2009); Tao et al. (2012); Baklanov et al. (2014)). These particles directly scatter and absorb incoming solar radiation, leading to changes in the atmospheric radiation reaching the Earth's surface (e.g., "aerosol direct effects"). They are either suspended in the air or attached to hydrometeors such as cloud droplets or ice crystals, acting as cloud condensation nuclei (CCN) or ice nuclei (IN). This interaction modifies the formation, lifetime, and optical properties of clouds, including cloud albedo, as well as precipitation rates, indirectly affecting the radiative transfer (e.g., "aerosol indirect effects"). During cloud processes, aerosol particles undergo physical and chemical changes in their composition and mass concentrations. They are also redistributed by clouds and convection, which serve as transport media over time scales ranging from minutes to hours (Ervens, 2015). Aerosol-cloud interactions, which encompass both the effects of aerosols on clouds and the cloud effects on aerosol particles, can play a crucial role in daily air quality, especially in regions with polluted aerosols and cloudy conditions. These interactions can influence both in-cloud and below-cloud wet scavenging of aerosol particles, leading to changes in PM concentrations reaching the ground.

From a data assimilation perspective, cloudy conditions are particularly challenging because ground-based or remote-sensing retrievals are often missing or significantly degraded in quality due to cloud contamination. When only limited observations are available for data assimilation, the quality of the analysis depends heavily on the numerical prediction system, especially its systematic errors. Variational data assimilation algorithms typically assume a perfect model without any systematic errors. Although it is not straightforward to detect and correct model error sources (Dee and Da Silva (1998); Dee (2005)), the analysis can benefit from advanced features that can enhance physical and chemical mechanisms, thereby improving the model performance. Eck et al. (2018, 2020) have reported that major air pollution events in East Asia are frequently associated with significant cloud cover, highlighting the importance of aerosol-cloud interactions compared to pristine conditions. This study is motivated by the haze events in Korea, which coincide with extensive cloud cover and rainfall.

The Weather Research and Forecasting model coupled with Chemistry (WRF-Chem; Grell et al. (2005)) facilitates simulations of real-time interactions between aerosol, cloud, radiation, and precipitation for regional applications. To account for aerosol effects on clouds and simulate aerosol-cloud interactions within the system, it is necessary to define cloud-borne aerosols (particles attached to cloud droplets) as well as interstitial aerosols (particles suspended in the air). When aerosol particles are represented in the aqueous phase (or in cloud water), cloud droplet number concentrations are prognostically treated through processes such as droplet activation, scavenging, and resuspension, allowing for real-time feedback between aerosols and clouds. Wet deposition, a major sink process for aerosol particles, involves the transport and removal of soluble or scavenged constituents by precipitation. It encompasses in-cloud scavenging and removal by rain and snow (rain-out), release by evaporation of rain and snow, and below-cloud scavenging by precipitation falling through without formation of precipitation (wash-out) (Seinfeld & Pandis, 2006). Subgrid-scale convective transport and in-cloud scavenging can be activated for aerosols in both the interstitial and aqueous phases (as well as tracers). However, to simulate wet scavenging by grid-resolvable precipitation and below-cloud scavenging through cloud chemistry, one needs to choose a chemistry option that includes aqueous chemistry with aerosols in the aqueous phase (as well as the interstitial phase) in the WRF-Chem model. Under cloudy conditions, the absence of these features can impact cloud formation, growth, and wet deposition of air pollutants, resulting in systematic forecast errors in surface PM concentrations.

Tuccella et al. (2015) implemented aqueous chemistry (AQCHEM) in the Regional Atmospheric Chemistry Mechanism (RACM; Stockwell et al. (1997)) gas-phase chem-

istry, coupled with the Modal Aerosol Dynamics Model for Europe (MADE; Ackermann et al. (1998)) inorganic aerosol mechanism and a secondary organic aerosol (SOA) scheme based on a four-bin volatility basis set (VBS) (Ahmadov et al., 2012) in the WRF-Chem model. They demonstrated that the RACM-MADE-VBS-AQCHEM scheme, coupled with cloud microphysics and radiation parameterization schemes, could characterize aerosol-cloud feedbacks, reducing large uncertainties in the prediction of microphysical and optical properties of clouds.

The WRFDA system has been recently updated to incorporate the RACM-MADE-VBS scheme (chem_opt=108) in the WRF-Chem model for chemical data assimilation (Ha, 2022). A case study conducted during the Korea–United States Air Quality (KORUS-AQ) period demonstrated that the 3D-Var aerosol analysis resulted in systematic improvements in the prediction of surface PM concentrations over Korea. However, forecast errors tended to increase in cloudy conditions.

This study further extends the WRFDA 3D-Var system to incorporate aerosol-cloud interactions with aqueous chemistry using the RACM-MADE-VBS-AQCHEM scheme (chem_opt=109) in aerosol analysis and forecast cycling. The objective is to enhance the short-range prediction of surface PM concentrations over South Korea, particularly during cloudy conditions characterized by wet scavenging of polluted aerosols.

Section 2 provides an overview of the WRF-Chem cycling system, describing the new implementation in the WRFDA system, which includes the introduction of new forward operators and background error statistics designed for aerosols in the aqueous phase. In Section 3, cycling experiments are conducted, and the forecast results are verified against independent observations to evaluate the reliability and performance of the system. Finally, in Section 4, we draw conclusions, discuss the limitations of this study, and provide suggestions for future research.

2 The WRF-Chem cycling system

The Weather Research and Forecasting (WRF) system consists of three main components: the WRF Preprocessing System (WPS), the WRF model coupled with Chemistry (WRF-Chem), and the WRF Data Assimilation (WRFDA) system. Through cycling (e.g., conducting analysis and forecast consecutively), the observed information is incorporated into the WRF-Chem model at certain time intervals (ex. every 6 h) to initialize the simulations. By pulling out the model trajectory towards observations every cycle, we re-initiate dynamical, physical and chemical mechanisms to be close to the observed state. This is a unified system in a sense that the forecast error is incorporated into the analysis (through background error covariance) and the prediction is initiated from its own analysis every time, thereby the error of the model itself is constantly reflected throughout the cycling system. Figure 1 shows a flowchart of the WRF-Chem/WRFDA cycling system with chemical data assimilation. Dotted lines imply optional input data while solid lines the mandatory inputs for WRF-Chem/WRFDA cycling, accompanied by typical input file names (with no specification of domain ID or time) used in the WRF system. As a first step, WPS is run to configure the model domain using geographical data for land use and soil categories (geogrid.exe), ungrib meteorological data (e.g., the UK Met Office analysis; UM MET) in the grib format (ungrib.exe), and transform the three-dimensional data into the WRF domain (metgrid.exe). Through the WRF initialization step (real.exe), the data is then converted to the initial condition (wrfinput) and lateral boundary condition (wrfbdy) files for meteorological variables in each domain. For chemical simulations, emissions data should be prepared based on the wrfinput file to define land use categories consistent with those used for the meteorological initial condition. As soon as the WRF-Chem model starts, atmospheric physics and chemistry parameterizations are initialized based on the land use categories (e.g., mminlu) in the lookup tables such as LANDUSE.TBL and VEGPARM.TBL. It is thus critical to use the same

wrfinput file in producing all the emissions data. By default, WRF-Chem regional simulations use an idealized gas profile for some chemical species at the lateral boundaries, as used in this study. It should be noted that the diagram shown here is not meant to describe all the possible data input in the WRF-Chem model. Optional input data, such as an upper boundary condition for some gas species, biomass burning (e.g., fire), or aircraft emissions data, are not included because they were not considered here.

Without data assimilation, wrfinput ($\mathbf{x_b}$) and wrfbdy files are used directly to initialize the model simulation, bypassing the WRFDA processes (e.g. da_wrfvar.exe and da_update_bc.exe). But if one wants to update the initial condition in the variational data assimilation, at least three input files are required for each model domain - a first guess ($\mathbf{x_b}$; wrfinput or fg), background error covariance ($\mathbf{B}$; be.dat), and observations ($\mathbf{y}$; ob.ascii or ob.bufr) that usually come with the specification of observation errors ($\mathbf{R}$). Before incorporating observations into the DA system, data collection and processing should be carefully carried out, including data quality check. Since WRF Version 4 (including WRFDA), simultaneous data assimilation has been available for a few chemical options in WRF-Chem to update meteorological and chemical fields at the same time. In the current implementation, chemical observations are designed to be available in ascii format (ob_chemsfc.ascii), separate from meteorological (MET) data provided in BUFR format. When data assimilation (da_wrfvar.exe) is run for each domain, the initial condition is updated as the analysis ($\mathbf{x_a}$) in each domain. The lateral boundary condition in the mother domain also needs to be updated (through da_update_bc.exe) to be consistent with the analysis in the boundary zone.

To compute the background error covariance ($\mathbf{B}$), WRF-Chem forecasts should be run in advance, typically cycling without data assimilation using the same model configuration for a long period of time (at least for one month). In the National Meteorological Center (NMC) method (Parrish and Derber (1992)), forecast differences between two different forecast leads at the same validation time are used to estimate the background error covariance for all the analysis variables in each domain.

Once the WRF-Chem model is integrated from the initial condition, the output forecast reached in the next cycle is reused to provide the next first guess with the simulated chemical species (e.g., wrf_chem_input). By repeating the WRF initialization, WRFDA, and WRF-Chem simulations with the recycled chemical species at the cycling frequency, WRF-Chem/WRFDA cycling can be carried out continuously.

2.1 WRF-Chem for aerosol effects

The WRF-Chem model has long been used to study a wide range of atmospheric phenomena associated with atmospheric chemistry and aerosols over regional domains (e.g., Ntelekos et al. (2009), Grell and Baklanov (2011), Pfister et al. (2011), Ahmadov et al. (2012), Saide et al. (2012), Yang et al. (2015)). The online-coupled chemical transport model numerically solves for the concentration of chemical species through various processes. Included are emissions, advection, vertical mixing with dry deposition, convective transport, gas chemistry, aerosol chemistry, cloud chemistry (for activated aerosols in cloud water), and wet scavenging. At the end of each time step, PM concentrations are computed diagnostically as the sum of all aerosol species defined in the chemistry option.

In the WRF-Chem model, aerosol effects are simulated through various processes, including activation, resuspension, aqueous reactions, and wet removal of aerosol particles. These processes are mostly controlled through namelist options for each model domain. As such, we specify the namelist parameters in parentheses corresponding to the description of each process. To account for aerosol-cloud interactions, or aerosol indirect effects, the direct effects of aerosols on incoming solar radiation must be activated (e.g., aer_ra_feedback = 1), which involves relating aerosol sizes and compositions to aerosol

optical properties (Fast et al., 2006). In this study, aerosol particles within a certain size range or mode are assumed to have the same composition. This allows for the averaging of refractive indices of spherical particles over all the species within each mode (e.g., volume averaging; aer_op_opt = 1). Subsequently, extinction coefficients due to aerosol scattering and absorption are calculated using Mie theory (Ghan et al., 2001) and incorporated into shortwave and photolysis schemes.

While aerosol direct effects can be accounted for by all the aerosol chemistry schemes when using either the RRTMG or the Goddard shortwave radiation scheme, aerosol indirect effects are only supported by a limited number of modal and sectional aerosol chemistry options in WRF-Chem. For cloud microphysics, it is recommended to employ double-moment schemes such as Lin (Lin et al., 1983) or Morrison (Morrison et al., 2009) to take indirect aerosol effects into account. In this study, the Morrison two-moment scheme is utilized, which predicts the mass and number concentrations of five species (i.e., cloud droplets, cloud ice, snow, rain, and graupel). It should be noted that the new implementation in DA is applicable to any double-moment scheme without any modifications. The double-moment microphysics accounts for the autoconversion of cloud droplets to rainwater based on the droplet number concentrations and interacts with prognostic aerosols, altering their size and composition through aqueous processes and wet scavenging (Yang et al., 2011). To represent aerosol effects on cloud chemistry and grid-scale precipitation (cldchem_onoff = 1 and wetscav_onoff = 1, respectively), it is necessary to simulate an aerosol activation process that enables aerosol particles to grow by water condensation, forming cloud droplets based on supersaturation and particle size (Abdul-Razzak & Ghan, 2002). For aerosol indirect effects, cloud droplet number concentrations should be prognostically treated (progn = 1). Meanwhile, the Grell-Devenyi cumulus scheme simulates convective precipitation, which is recognized by atmospheric radiation and photolytic processes in the model. It also parameterizes convective transport, enabling the displacement of chemical species. Although aqueous chemistry in subgrid-scale convection (conv_tr_aqchem = 1) and wet scavenging within subgrid-scale clouds (conv_tr_wetscav = 1) can be accounted for as part of the cumulus parameterization, it does not explicitly consider cloud-borne aerosols or their impact on cloud chemistry. For brevity, we refer to aqueous chemistry (AQCHEM) only when aerosols are defined in the aqueous phase for the chosen chemistry scheme (e.g., chem_opt = 109) and aerosol indirect effects are simulated.

Tuccella et al. (2015) implemented the RACM-MADE-VBS-AQCHEM scheme (chem_opt=109) for the simulation of aerosol-cloud-radiation interactions, following Fast et al. (2006) and Chapman et al. (2009), with simple aqueous reactions. The MADE-VBS aerosol scheme defines the particle size distribution as a superposition of three log-normal modes: an Aitken mode with a median diameter of 0.01 μm, an accumulation mode ranging between 0.01 and 1 μm, and a coarse mode for particles typically larger than 1 μm (with a median around 10 μm). All aerosol particles are assumed to be spherical and internally mixed (Aquila et al., 2011). The aerosol species treated are sulfate ($SO_4^=$), nitrate (NO_3^+), ammonium (NH_4^+), elemental carbon (EC), primary organic matter (POA), anthropogenic and biogenic secondary organic aerosol (SOA), chloride (Cl), sodium (Na), unspeciated $PM_{2.5}$, unspeciated coarse fraction of PM_{10} (antha), soil dust (soila), and sea salt (seas). The unspeciated $PM_{2.5}$ includes the fine fraction of sea salt and mineral dust aerosols.

For aqueous processes, each aerosol species is defined in the aqueous (or cloud-borne) phase as well as in the interstitial (or non-activated) state. The number and mass concentrations of activated aerosols are calculated for each mode in the presence of water supersaturation. In this study, we identified and addressed several bugs in the RACM-MADE-VBS-AQCHEM scheme in the previous versions of the WRF-Chem model (e.g., prior to V4.4.1), primarily related to a simple sulfuric oxidation. These bug fixes were incorporated into the released version of WRF V4.4.1. Also, some deposition variables have been added for diagnostics purposes specific to this study.

2.2 WRFDA updates for aqueous chemistry

An interface between the WRF-Chem model and the WRFDA 3D-Var system in version 4.4.1 has been extended for the RACM-MADE-VBS-AQCHEM option to assimilate ground-level measurements of $PM_{2.5}$, PM_{10}, SO_2, NO_2, O_3, and CO concentrations. The RACM-MADE-VBS scheme without aqueous chemistry (chem_opt=108) was previously implemented in the WRF-Chem/WRFDA system by Ha (2022). In this study, the interface is further expanded to include aqueous-phase aerosols (e.g., chem_opt=109) in the aerosol analysis.

2.2.1 Cost function

In the 3D-Var system, the cost function $J(\mathbf{x})$ is minimized to find an optimal solution for the model state ($\mathbf{x}$) that best fits to all the observations ($\mathbf{y}$) available at the analysis time. The minimization is performed based on the background and observation error covariance matrices ($\mathbf{B}$ and $\mathbf{R}$, respectively), assuming Gaussian error distributions (Lorenc, 1986). In the incremental formulation (Courtier et al., 1994) adopted in WRFDA, analysis increments $\delta\mathbf{x}(= \mathbf{x} - \mathbf{x_b})$ are computed at each iteration using the background forecast ($\mathbf{x_b}$) from the previous analysis ($\mathbf{x_a}$) or the previous iteration step. The control vector ($\mathbf{v}$) is defined as $\delta\mathbf{x} = \mathbf{B^{1/2}v}$ to construct the cost function as below.

$$J(\mathbf{v}) = \frac{1}{2}\mathbf{v^T v} + \frac{1}{2}(\mathbf{d} - \mathbf{HB^{1/2}v})^{\mathbf{T}}\mathbf{R^{-1}}(\mathbf{d} - \mathbf{HB^{1/2}v}) \quad (1)$$

where the innovation vector is defined as $\mathbf{d} = \mathbf{y} - \mathbf{H}(\mathbf{x_b})$ and the observation operator $\mathbf{H}$ transforms the model states ($\mathbf{x}$) to the observed quantities ($\mathbf{y}$) at observation locations. In the chemical data assimilation presented in this study, all the chemical species defined in the model states ($\mathbf{x}$) are also used as control variables ($\mathbf{v}$), which are the same as analysis variables. These control variables have univariate error covariances, meaning that there are no static cross-correlations between chemical species or between chemical and meteorological variables. A list of 32 three-dimensional aerosol species defined in the analysis includes the following: aerosol sulfate (so4ai and so4aj), nitrate (no3ai and no3aj), ammonium (nh4ai and nh4aj), chloride (clai and claj), primary organic matter (orgpai and orgpaj), elemental carbon (eci and ecj), sodium (naai and naaj), unspeciated $PM_{2.5}$ (p25ai and p25aj), 4-bin anthropogenic and biogenic SOA (asoa1i, asoa1j, asoa2i, asoa2j, ..., bsoa4i, bsoa4j). Each variable name in the parenthesis ends with i or j to indicate Aitken or accumulation mode. Also included are three coarse-mode variables - non-reactive anthropogenic primary aerosol (antha), marine aerosol concentration (seas), soil-derived aerosol particles such as dust (soila). In summary, there are 35 aerosol species defined in three modes (= 16 Aitken + 16 accumulation + 3 coarse modes), and their aqueous phase counterparts are also defined with 'cw' added to their names. Four gas species, namely SO_2, NO_2, O_3, and CO, can be also assimilated at the ground level, depending on the assimilation option. This results in a total of 74 three-dimensional chemical species in the control vector.

2.2.2 Forward operator and observation errors

In the assimilation of surface PM observations, $\mathbf{H}$ is calculated as the sum of each aerosol species defined in the control vector ($\mathbf{v}$), interpolated at each observation site. This approach follows the way the MADE-VBS aerosol scheme in the model estimates PM concentrations based on individual aerosol species, ensuring that they are treated consistently between the analysis and forecast. For the activation of aqueous chemistry, $\mathbf{H}(\mathbf{x})$ is extended to include cloud-borne (activated) as well as interstitial (non-activated) aerosol species in all three modes. The $PM_{2.5}$ concentrations in the model space ($y_{pm_{2.5}}$) are computed as the sum of all the aerosol species listed above in accumulation (j) and

Aitken (i) modes.

$$\mathbf{y}_{pm_{2.5}} = \rho_d \sum_{p=1}^{N} (\sum_{m=i}^{j} \mathbf{y}_m^p + \sum_{m=i}^{j} \mathbf{y}^{*p}{}_m), \tag{2}$$

where N is 32, ρ_d dry air density ([kg m^{-3}]) for unit conversion from aerosol mixing ratios ([μg kg-1]) to mass concentrations ([μg m^{-3}]), $\mathbf{y}_m^p$ and $\mathbf{y}^{*p}{}_m$ representing each aerosol species in the interstitial and aqueous phases, respectively.

When assimilating PM_{10} alone, the model correspondent is computed by adding three coarse-mode variables - antha, seas, and soila - to the simulated $PM_{2.5}$. For the aqueous chemistry option, the three coarse-mode variables in the aqueous phase (anthcw, seascw, and soilcw) are also included in the observation operator. If PM_{10} is assimilated together with $PM_{2.5}$, the residuals from (PM_{10} - $PM_{2.5}$) are assimilated as the sum of the three coarse-mode aerosols, following Ha (2022). It should be noted that aerosol number concentrations, which are not directly associated with PM mass concentrations, are not included as analysis variables or in the observation operators in this study.

The assimilation of trace gases is straightforward because each gas species is explicitly defined in the model prognostic variables. The control variables for the assimilation are the same four gas species (SO_2, NO_2, O_3, and CO). The observation operator for trace gases involves a simple horizontal interpolation of the corresponding variable at the lowest model level.

The observation error covariance matrix $\mathbf{R}$ remains unchanged, regardless of the inclusion of aqueous chemistry. It uses the same uncorrelated observation errors for each observation (y_o). Following Ha (2022), the observation errors for surface PM consist of measurement errors (ϵ_o) and representative errors (ϵ_r): $\epsilon_y = \sqrt{{\epsilon_o}^2 + {\epsilon_r}^2}$ where $\epsilon_o = 1.5{+}0.0075{*}y_o$ and $\epsilon_r = \gamma\epsilon_o\sqrt{\frac{\Delta x}{L}}$. Here, γ is set to be 0.5, Δx is grid spacing (27 km for domain 1 and 9 km for domain 2), and the scaling factor L (defined as 3 km).

For system reliability, data quality control (QC) is performed by applying maximum thresholds to observation values and innovations ((o − f)'s) during the assimilation process. Surface $PM_{2.5}$ and PM_{10} observations are rejected if they exceed 300, 500 μg m^{-3}, respectively, or if they differ from their corresponding model equivalents (e.g., $\mathbf{H}(\mathbf{x_b})$) by more than 100 μg m^{-3}. Regarding gas species, the maximum threshold values are set at 2 ppmv for observed SO_2, NO_2, and O_3, and 50 ppmv for CO. They are also rejected if their innovations exceed the threshold values of 0.2 ppmv for SO_2, NO_2, and O_3, and 20 ppmv for CO. It is noted that these threshold values are set to be conservative for the sake of the system reliability, especially in the operational environment. Although none of the observations were rejected based on the thresholds during the month-long period in our case study, it could limit the applicability of WRFDA to heavy pollution events or wildfire episodes. In the future, it would be nice to move these parameters out of the codes and place them in the namelist such that users can easily modify them for their specific applications. In the current implementation, ground-level gas-phase pollutants are assimilated together as a group rather than individually, using the corresponding model variables as their analysis (or control) variables.

2.2.3 Background error covariance

In the WRFDA system, the square root of the $\mathbf{B}$ matrix ($\mathbf{B} = \mathbf{B^{1/2}}(\mathbf{B^{1/2}})^{\mathbf{T}}$) is decomposed into a series of sub-matrices, eliminating the need of computing the inverse of the large $\mathbf{B}$ matrix, as below.

$$\mathbf{B^{1/2}} = \mathbf{U_p S U_v U_h} \tag{3}$$

Here, the matrix $\mathbf{U_p}$ is called physical or balance transformation (via linear regression), $\mathbf{S}$ a diagonal matrix of forecast error standard deviation, $\mathbf{U_v}$ the vertical transform, and $\mathbf{U_h}$ the horizontal transform matrix.

In this study, the WRF-Chem model is configured with two domains at grid resolutions of 27 and 9 km, respectively, in a one-way nesting mode, as illustrated in Fig. 2 (a). Vertically, a total of 31 model levels are used up to 50 hPa, with upper level jets located around level 23 (∼12 km) and low level jets (LLJ) situated around level 9 (∼1.2 km). At Seoul (37.5N, 127.0E), for instance, a total of 8 model levels are configured under 1 km, with the lowest level at 72 meters (height above mean sea level). Chemical simulations are cycled without data assimilation every six hours using the RACM-MADE-VBS-AQCHEM scheme, starting from 21 Feb to 31 March 2019, producing 48-hour forecasts from 00Z each day for the month of March. The differences between the 24-hour and 48-hour forecasts at the same validation time are then computed as a proxy for forecast errors in each domain. In total, 31 sample forecasts in March, following an 8-day spin-up in Feb 2019, were used to construct the $\mathbf{B}$ matrix through the NMC method. To generate the $\mathbf{B}$ matrix in the WRFDA system, the GENBE 2.0 software (Descombes et al., 2015) is extended to incorporate all the aerosols in the aqueous phase, as specified in the chosen chemistry scheme. In line with previous 3D-Var studies (Kumar et al. (2019) and Ha (2022)), we binned all the grid points together for each latitudinal band (every 2°) and each model level, but with no longitudinal dependencies in the background error covariance. To examine the impact of aqueous-phase aerosols in the B matrix, we conducted one-month cycling forecasts with chem_opt = 108 and 109, referred to as NO_AQ and AQ, respectively.

Figure 3 illustrates the comparison of the square root of the background error covariance ($\mathbf{B^{1/2}}$) between the two experiments for each aerosol component in the interstitial state. As the control vector ($\mathbf{v}$) is multiplied by $\mathbf{B^{1/2}}$ to convert it to analysis increments ($\delta\mathbf{x}$) after the minimization process, this figure shows the relative weights of each species, indicating their contributions to atmospheric constituents and their vertical distribution when assimilating ground-level PM concentrations. It is evident that there is considerable variability in the vertical distribution of aerosol species, and their vertical structure undergoes distinct changes when simulating aerosols in cloud water (AQ). These changes in the interstitial state can be interpreted in conjunction with the background error covariance ($\mathbf{B^{1/2}}$) for aqueous-phase aerosols (AQ case), as displayed in Fig. 4. Sulfate aerosols are found to have the largest weights in the accumulation mode (cwj) and Aitken mode (cwi), while sea salt aerosols dominate the coarse mode (cw). These findings are consistent with the significant reductions observed for the same species in the interstitial phase across the entire troposphere, as depicted in Fig. 3. These changes can be attributed to the aqueous chemistry implemented in the WRF-Chem model, which primarily involves the oxidation of sulfur dioxide (SO_2) to dissolved sulfur in oxidation state 4, S(IV). It is also noteworthy that aerosols in cloud water ("cw") are concentrated below level 15 (e.g., below 5 km), with peak concentrations estimated around level 5, likely in association with low-level clouds. In contrast, most interstitial aerosols (as shown in Fig. 3) tend to increase as the level goes down, but their distribution extends up to level 24. Only dust aerosols (soila in Fig.3 (l)) dominate at higher altitudes, with a peak around level 13 (around 3 km), possibly due to long-range transport of dust during the spring month. Another thing to note is that organic aerosols (both primary and secondary) are not updated through aqueous chemistry in the current version, showing little changes in the aqueous phase, as indicated by the pink solid line for "POA" and blue dotted line for "SOA" in Fig. 4 (a) and (b). The specific mechanisms underlying all the changes in each species are not fully understood, but it is apparent that the weights for soluble species are partitioned between the interstitial (dry) and the aqueous phase (cloud water), changing the structure of the background error covariance with aqueous chemistry. This highlights the complex interactions and transformations that occur between different phases of aerosols and the role of aqueous chemistry in modifying their behavior and representation in the model.

3 Chemical analysis and forecast cycling

3.1 Cycling experiments

To examine the impact of chemical data assimilation with aqueous chemistry, WRF-Chem and WRFDA cycling experiments are conducted every 6 h from Feb 21 to Mar 31, 2019 over the East Asian region (with a 27-km grid resolution) nested down to the Korean peninsula (at a 9-km resolution). This study uses the Morrison two-moment scheme (Morrison et al., 2009) for cloud microphysics, Grell-3 for cumulus parameterization (Grell & Dévényi, 2002), the YSU scheme (Hong et al., 2006) for the planetary boundary layer (PBL), and the rapid radiative transfer model for GCMs (RRTMG) for both shortwave and longwave radiation (Iacono et al., 2008). As described in the previous section, direct aerosol effects are accounted for through interactions with atmospheric radiation and photolysis while indirect aerosol effects are represented through interactions with cloud microphysics. Dust and sea salt emissions are simulated online, following the GOCART mechanism (e.g, dust_opt = 13 and seas_opt = 2). Photolysis rates of chemical species are computed in a simplified version of the National Center for Atmospheric Research (NCAR) Tropospheric Ultraviolet-Visible (TUV) model (phot_opt=1) (Madronich, 1987). A list of the physics and chemistry schemes used in this study is summarized in Table S1.

The anthropogenic emissions data for chemical species defined in the RACM-MADE-VBS scheme are obtained from the National Institute of Environment Research (NIER), which operates daily air quality forecasting in South Korea. These data are provided at a single level and do not include information on plume rise or vertical distribution. Biogenic emissions, on the other hand, are generated online using the Model of Emission of Gases and Aerosol from Nature (MEGAN; Version 2) (Guenther et al., 2006), but biomass burning emissions are not used in this study. All WRF files, including anthropogenic and biogenic emissions, are processed based on MODIS land use dataset (Friedl et al., 2002).

The initial and lateral boundary conditions for meteorological variables are derived from global analyses and forecasts from the UK Met Office's Unified Model (UM) operated by the Korea Meteorological Administration (KMA) every 6 hours. However, the chemical lateral boundary conditions for the outer domain are not considered, while the chemical boundaries for the inner domain (D2) are provided during the one-way nested model simulations. Upper boundary conditions are also not provided in this particular study. The impacts of chemical boundary conditions on internal physical mechanisms, such as wet scavenging, are reserved for future studies.

Hourly surface observations of $PM_{2.5}$, PM_{10}, SO_2, NO_2, O_3, and CO are collected from 379 South Korean sites operated by AIRKOREA (http://www.airkorea.or.kr, last access: 27 April 2023) and 765 Chinese sites of the China National Environmental Monitoring Center (CNMEC; http://www.cnemc.cn, last access: 27 April 2023) within model domain 1 (D1). Since measurements are mainly concentrated in large cities, Korean sites are randomly divided into assimilation and verification datasets. Each dataset is then averaged over the 9-km model grid. As a result, 279 Korean sites are processed into 219 stations for assimilation, while the remaining 100 sites are averaged to form 71 independent observations for evaluation over South Korea. The Chinese data are used for both assimilation and verification without any additional data processing because the focus of this study is to examine the aerosol impact over South Korea. Figure 2 depicts the surface network used for assimilation (panel (a)) and for evaluation (light blue dots in (b)). In panel (b), three surface stations of the EANET (Acid Deposition Monitoring Network in East Asia; https://www.eanet.asia; last access: 13 June 2023) are also represented by a pink "+" symbol along with their respective site names, which are utilized for evaluating the monthly wet deposition simulations. For meteorological data assimilation, all the conventional observations available in the National Centers for Environmental Prediction (NCEP) prepbufr data (https://rda.ucar.edu/datasets/ds337.0/; last

access: 13 June 2023) are employed. To verify the model performance against weather observations, a total of 699 surface Automatic Weather System (AWS; https://www.weather.go.kr/weather/observatio last access: 13 June 2023) sites in South Korea (marked as blue dots in Fig. 2c) are employed.

As summarized in Table S2, two baseline experiments are performed with and without aqueous chemistry (NODA and NODA_AQ, respectively). Using the background error statistics computed from each of these experiments, two corresponding DA cycling runs are then conducted with the same model configuration, assimilating surface concentrations of all six air pollutants as well as conventional meteorological data (DA and DA_AQ, respectively).

Figure 5 illustrates time series of observation-minus-background (o-b; blue solid lines) and observation-minus-analysis (o-a; red dashed lines) averaged over 219 Korean stations (red dots in Fig. 2) for $PM_{2.5}$ (top) and PM_{10} (bottom) at the ground level in the DA_AQ cycling experiment in domain 2. The light blue shading indicates the standard deviation in (o-b)'s across the stations. Although the deviations of 6-hour background forecasts from observations exhibit some fluctuations with cycles, they remain stable, and the analyses closely match the observations, confirming that the system runs reliably throughout the entire period. As for four gas species, they tend to be slightly overestimated during the cycling period, but runs stably as well, as shown in the supplement (Fig. S1). As this study focuses on the prediction of surface particulate matter, our discussion will remain focused on surface $PM_{2.5}$ and PM_{10} concentrations.

Figure 6 compares the analysis and background forecast profiles of each aerosol component, averaged over the 71 verification sites in Korea, for the month of March 2019, in two DA experiments (DA and DA_AQ). Aerosols are combined for each species regardless of the aerosol phase or mode. Panels (a) - (i) show the sum of Aitken and accumulation mode particles, and cloud-borne (cw) aerosols are also included in the corresponding chemical species in the case of DA_AQ. In the 3D-Var analysis, analysis increments in PM concentrations are distributed across aerosol species based on the background error covariance, resulting in the vertical structure of each species generally following their background error structures illustrated in Figs. 3 and 4. Therefore, even though only surface concentrations are assimilated, their impact goes up to the boundary layer. Comparing the analysis with and without aqueous chemistry (red and black lines, respectively), sulfate (a), ammonium (b), sea salt (k), and soil dust (l) are reduced with aqueous chemistry throughout the troposphere, while nitrate (b) and chloride (i) are mostly increased in the low troposphere. Primary organic carbon (f), elemental carbon (g), and the unspeciated coarse fraction of PM_{10} (j) tend to be redistributed toward the surface with aqueous chemistry. Regarding analysis increments, DA with aqueous chemistry seems to produce slightly larger increments than without AQ, especially in coarse-mode sea salt (k) in the boundary layer, followed by nitrate (b) and sulfate aerosols (a) in the accumulation mode. Averaged over the month of March 2019, the analyses of most aerosol species show clear differences depending on whether aqueous chemistry is activated (e.g., red vs. black). However, the analysis increments are relatively small in both DA experiments, implying that the physics mechanism plays a more crucial role in simulating atmospheric composition than the initialization through data assimilation.

3.2 Air pollution events in cloudy conditions

Air pollutants transported to the Korean peninsula are susceptible to the moist environments above the Yellow Sea, which lies between Korea and China. The extent to which aerosol particles interact with moisture or cloud droplets is subject to the moving speed and direction of the synoptic weather systems, such as fronts or troughs, that traverse the sea. To examine the effect of data assimilation with aqueous chemistry, an air pollution case with substantial cloud cover and wet scavenging was selected during

the cycling period. Figure 7 (a) illustrates the Level 3 (gridded) daily mean aerosol optical depth (AOD) retrieved from the Visible Infrared Imaging Radiometer Suite (VIIRS) aboard the Suomi National Polar-Orbiting Partnership (Suomi NPP) spacecraft (https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/viirs/, last access: 27 April 2023) on 19 March 2019. The total column AOD indicates high aerosol loading over Korea, especially along the west coast near the Seoul Metropolitan Area. On that day, a surface high pressure center is situated southwest of Jeju island (around 33.4°N, 126.5°E), and the associated anticyclonic circulation over the sea brings air pollutants and moisture into South Korea along with southwesterly winds (not shown). The areas without colors in the satellite image are commonly affected by cloud contamination, resulting in missing data.

On the following day, a low surface pressure system further approaches, leading to significant cloud cover over a wide area surrounding the Korean peninsula. As shown in Fig. 7 (b), the low cloud top pressure retrieved from MODIS sensors onboard the Aqua satellite (DOI: 10.5067/MODIS/MYD06_L2.061) indicates the development of convective clouds (shown in white) over the Yellow Sea near Seoul, South Korea. This cloudy condition leads to the absence or poor quality of most remote sensing retrievals across the entire region, making in-situ surface measurements the primary source of observed information. In association with the extensive cloud cover and long-range transport embedded in the synoptic atmospheric flows, air quality experiences large variations before and after the two days. Figure 8 displays the boxplots representing surface PM_{10} (top) and $PM_{2.5}$ (bottom) concentrations for 24-hour forecasts from the analysis at 00 UTC every day for five consecutive days (17-22 March 2022). Each boxplot indicates either daily observations or daily forecasts from each cycling experiment at 71 verification sites. The time series of observations (shown in a blue boxplot) illustrates that both PM_{10} and $PM_{2.5}$ concentrations increase until 19 March 2019, accompanying the low pressure system with extensive clouds, and then gradually decreases in the following days.

The largest differences between experiments and observations are found in surface PM_{10} concentrations on 20 March 2019, depending on whether aqueous chemistry (AQ) is employed. The experiments with AQ (e.g., NODA_AQ in red and DA_AQ in purple) accurately simulate the decreasing trend in mass concentrations owing to the wet scavenging of aerosol particles. However, the experiments without AQ (e.g., NODA in orange and DA in green) predict substantial increases in concentrations. This implies that the absence of aqueous chemistry in the model can lead to a statistically significant overestimation of surface concentrations. Without the capability to simulate below-cloud scavenging of air pollutants by precipitation, the model misses the observed wash-out feature, predicting an opposite trend to the observed one. Surface $PM_{2.5}$ concentrations exhibit similar trends, but the differences between experiments are much smaller. Also, the differences in experiments without DA (e.g., NODA-NODA_AQ) are larger than those with DA (e.g., DA-DA_AQ) on 20 March, indicating that model forecasts are more vulnerable to model errors when initialized without data assimilation.

These daily trends can provide a statistical overview of each experiment's behavior, but it would be also interesting to investigate hourly variations associated with the wet scavenging process. Figure 9 depicts 48-hour forecasts from the 00 UTC analysis for each experiment, comparing them to the observed surface PM_{10} (top) and $PM_{2.5}$ (bottom) concentrations averaged over the 71 verification sites. Hourly rainfall accumulations (top) and mean sea level pressure (bottom) are also plotted as gray dotted lines with the right y-axis, averaged over 699 automated weather stations in South Korea. The time series of hourly rainfall (in the top panel) shows light precipitation reported overnight (from 10 to 15 UTC; 19-24 KST), while a surface low pressure system (in the bottom panel) passes over Korea. As aerosol particles are removed through wet scavenging, the observed PM concentrations continue to decrease until the next morning. The experiments with aqueous chemistry (NODA_AQ represented by an orange plus sign and DA_AQ

marked as a red circle) accurately capture the decreasing concentrations for most of the 48-hour forecast period. In contrast, experiments without aqueous chemistry (NODA in green "x" and DA in a blue square) predict substantial increases in hourly concentrations even after the rainfall. This is likely because the model simulations without AQ assume all the aerosol particles as suspended in the air without any loss through washout. In this particular case, data assimilation does not make any significant differences, even at the analysis time, suggesting that the forecast error is primarily driven by model errors related to the parameterized wet scavenging mechanism. And it should also be noted that the difference between NODA and NODA_AQ in surface PM_{10} concentrations (green "x" vs. orange "+"), as the mean over all the Korean sites, corresponds to the prediction of air quality changing from very unhealthy (> 150 μg kg-1) to clear condition (< 30 μg kg-1), which is almost comparable to a situation of rain or no rain. This suggests that wet deposition with aqueous chemistry should be included to correctly simulate the evolution of aerosols and prevent false alarms in air quality forecasting, especially when associated with precipitating clouds.

These distinctive differences are also observed in the meteorological variables. Figure 10 displays the same 48-hour forecasts in AOD, liquid water path (LWP), planetary boundary layer height (PBLH), and hourly rain accumulations. Note that this figure is not intended for evaluation, but rather for demonstrating the sensitivity of these meteorological fields to the wet scavenging of aerosols in the online-coupled system. The column-integrated AOD is expected to follow the hourly trends in surface PM_{10} concentrations in each experiment. As observed in the cloud image shown in Fig. 7 (b), the LWP ($=\int_0^z \rho q_c dz$, where ρ stands for dry density, q_c cloud water mixing ratio, dz height differences between two adjacent levels in the model) increases overnight. However, depending on whether aqueous chemistry is activated with aqueous-phase aerosols, the simulated LWP can vary by as much as three times. The PBL height is also considerably different between the experiments, in association with the development of clouds and rain that affect the wet removal of aerosol species. During the nighttime (around 12 UTC; 21 KST), the height of the stable boundary layer exceeds 1 km in experiments without aqueous chemistry. Since we lack in-situ measurements for the PBLH, we are unable to directly verify these modeling behaviors. However, it is apparent that the simulated PBL height becomes systematically higher under stable boundary conditions (e.g., at night) when aqueous-phase aerosols are not considered in cloud chemistry and wet scavenging. In terms of hourly total precipitation, which includes both convective and nonconvective rainfall amounts, wet deposition of aerosols does not significantly alter the predicted rainfall in the cycling experiments. But the total precipitation is mostly consistent with the observation (represented by a "x" marker) for this particular case of high cloud cover producing light precipitation. Whether or not aqueous-phase aerosols are represented, it not only affects the vertical mixing and the structure of the boundary layer but also affects surface conditions. While surface winds are not very sensitive to the use of aqueous chemistry in this particular study, which is partly attributed to the large uncertainties in simulating the nocturnal stable boundary layer, there are noticeable changes in 2-m temperature and relative humidity, up to 5%, especially during the nighttime (not shown). As Saide et al. (2015) pointed out, aerosols can play an important role in modifying severe weather conditions or outbreaks. But in the weakly coupled DA system used in this study, aerosol and weather data assimilation only indirectly affect each other through aerosol feedbacks in the forecast model, and the assimilation of surface weather observations is not effectively performed owing to the specification of large observation errors. A thorough investigation of the influence of aerosol data assimilation on meteorological conditions and the optimization of weather data assimilation is left behind for future studies.

To further explore the relationship between the vertical distribution of PM concentrations and the development of clouds and rain, we examine hourly PM_{10} concentrations with a few meteorological variables at Seoul (37.5°N, 127.0°E) in 24-hour forecasts from the 00 UTC analysis on 20 March 2019 for each experiment, as shown in Fig.

11. In the WRF-Chem model V4.4.1, since aerosols can act as cloud condensation nuclei, but not as ice nuclei, the simulation of aerosol-cloud-precipitation interactions is limited to warm rain processes (Tuccella et al., 2015). Also, the background error covariance defines most of the aerosol impacts below the model level 20 (as depicted in Figs. 3 and 4), so it is desirable to focus on the vertical structure up to level 20. Contours of cloud water and rain water mixing ratios are represented in white and pink, respectively, and the LWP is overlaid as a black dashed line with the y-axis on the right. In the comparison, it is evident that the use of aqueous chemistry can play an important role in the formation and development of clouds as well as the wet deposition of aerosols. Without AQ and DA (Fig. 11 (a)), the simulation is initiated with high concentrations of PM_{10} in the boundary layer at 00 UTC (09 KST), which remains for most of the day (until 08 UTC; 17 KST). Clouds start to form around level 15 ($\sim$5 km) in the late afternoon and undergo some autoconversion and accretion processes (pink contours), but they mostly persist through the night, moving down to the ground. Autoconversion refers to the process where cloud droplets collide and coalesce to form raindrops, while accretion denotes the collection of cloud droplets by falling raindrops. With the development of low clouds in this simulation, the air quality is predicted to reach very unhealthy conditions (surface $PM_{10} > 150\ \mu g\ m^{-3}$) for the following hours starting from 18 UTC (03 KST). In Fig. 11 (b), data assimilation effectively mitigates the overestimation of low-level PM concentrations from the initial time to the late afternoon. But it cannot compensate for the model error resulting from the absence of the wet removal mechanism in the later forecast lead times, where a similar pattern of high PM concentrations is simulated. In Fig. 11 (c), however, the activation of aerosols in clouds drastically changes the model behavior to enhance rain water mixing ratios in the low troposphere (between levels 8-13, corresponding to 1-3 km) starting from the late afternoon (around 08 UTC) and the wet removal of aerosol particles in the troposphere (around level 10) for the next 6 hours or so. In the case of DA_AQ (Fig. 11 (d)), data assimilation initially suppresses the PM overestimation again. Other than that, the vertical structure and temporal evolution remain similar to the case without DA, highlighting the strong impact of aqueous chemistry interacting with cloud microphysics and wet deposition processes on air quality forecasting. As cloud droplets play an important role in aqueous-phase reactions, an accurate simulation of LWP is crucial. The time series of LWP reveals that the LWP can be simulated up to three times larger when aerosols in clouds are not considered. The enhanced mid-level clouds with little autoconversion of cloud to rain water droplets are closely tied to the overprediction of PM concentrations.

The same figure for $PM_{2.5}$ concentrations, along with the PBL height, is provided in the supplementary material (Fig. S2). The temporal and vertical distributions of $PM_{2.5}$ concentrations are similar to those of PM_{10}, with data assimilation suppressing initial concentrations and aqueous chemistry contributing to the wet removal of aerosols overnight. In the AQ experiments, cloud-borne sulfate aerosols are well overlaid with cloud water mixing ratios (QCLOUD), illustrating their formation within the clouds through the process of aqueous chemistry. Also, the temporal variation of the boundary layer height during nighttime is largely influenced by the activation of aqueous chemistry. While the maximum height in Seoul remains relatively consistent across the experiments, ranging from 1.1 to 1.5 km, the timing of the peak PBL height varies significantly. Both AQ experiments (NODA_AQ and DA_AQ) simulate the deepest boundary layer around 08 UTC (17 KST), followed by a sharp decrease, which is a more realistic representation compared to the experiments without AQ, where the boundary layer continues to grow until the evening (11 UTC or 20 KST). Overall, it is apparent that clouds and precipitation can greatly alter the evolution of aerosols in the atmosphere, and aerosols can also exert large influences on the cloud formation and development. Notably, whether aerosol wet scavenging or removal is simulated in the model determines the prediction of PM concentrations between heavy or no pollution events at the ground level. The significance of aerosol wet scavenging by clouds aligns with findings from previous studies (Ryu et al., 2022a, 2022b).

It is challenging to directly evaluate the aerosol wet removal mechanism at the process level, and a comprehensive investigation of all the physics mechanisms in the coupled system is beyond the scope of this study. However, we can compare our simulations with the monthly observations provided by the Acid Deposition Monitoring Network in East Asia (EANET). The EANET data includes monthly (dry and wet) deposition at three Korean sites, as marked in Fig.2 (b). To facilitate the comparison, we have modified the model codes to produce wet deposition for sulfate, nitrate, and ammonium aerosols in the RACM-MADE-VBS-AQCHEM scheme during the model integration. Figure 12 shows that the daily rainfall predictions in both NODA_AQ and DA_AQ are in good agreement with the observation in Gangwha, but are overestimated in Imsil and underestimated in Jeju. Wet deposition at each site largely follows the bias in rainfall, with the largest underestimation simulated in Jeju, except for sulfate in DA_AQ. Overall, data assimilation tends to increase wet deposition at all three sites, slightly improving the underestimation in NODA_AQ.

The equilibrium between the gas and aqueous phase varies across aerosol species depending on their solubility. The activation of aerosols is determined based on the hygroscopicity of each aerosol component, and wet deposition is applied to individual aerosol species. Hence, the processes of aerosol activation, scavenging, and chemical reactions through aqueous chemistry can lead to changes in the aerosol size distribution and the atmospheric composition. To investigate the impact of aqueous-phase aerosols on the aerosol size distribution, Fig. 13 presents the vertical distribution of aerosol mass concentrations by mode. Here, the 6-hour background forecasts from each experiment are averaged over the 71 verification sites in domain 2. The top panel shows the monthly mean for March 2019, while the bottom panel compares the forecasts after the aerosol wet removal associated with grid-scale precipitation. In the far right in (d) and (h), PM_{10} and $PM_{2.5}$ are also displayed for each experiment. In the month-long statistics, data assimilation mostly increases aerosol species in the accumulation mode, especially in the boundary layer (panel (b)), leading to an increase in surface $PM_{2.5}$ concentrations by up to 29%. On the other hand, the use of aqueous chemistry seems to have a greater impact (compared to data assimilation) on the simulation of coarse-mode aerosols. It reduces the mass concentrations in the mid-to-low troposphere, contributing to the decrease of PM_{10} concentrations in NODA_AQ and DA_AQ, as indicated by the lines with dots in panel (d). In the Aitken mode, both aqueous chemistry and data assimilation vertically redistribute aerosol species, resulting in higher concentrations near the surface. After the wet removal associated with the rainfall event, however, the inclusion of aqueous chemistry considerably decreases aerosol species in all three modes, regardless of data assimilation. This leads to reduced PM concentrations, as illustrated in the bottom panel for 00 UTC 21 March 2019. It is worth noting that without AQ, the DA experiment (represented by blue dashed lines) attempts to reduce PM concentrations in both accumulation and coarse modes but data assimilation alone is unable to effectively reduce the overestimation when there is a large model error due to the missing mechanism for the below-cloud wet scavenging.

In terms of forecast skills over the entire month, the root-mean-square-error (RMSE) of 24-hour forecasts in the 9-km domain (D2) in NODA is around 24 and 37 μg kg^{-1} for surface $PM_{2.5}$ and PM_{10} concentrations, respectively, verified against independent observations (e.g., at the 71 verification sites). Data assimilation improves the errors up to $\sim$3 μg kg^{-1}, but the use of aqueous chemistry does not change the errors in a statistically significant way. In terms of systematic normalized mean bias of 24-h forecasts (e.g., $NMB = \sum(f-o)/\sum o \times 100\%$), however, surface $PM_{2.5}$ concentrations without data assimilation are significantly underestimated (up to -24 %) in both NODA and NODA_AQ, which is improved by assimilation up to -4 %. Surface PM_{10} concentrations are somewhat overestimated by data assimilation in DA (15 %), which is reduced to approximately 5% when aqueous chemistry is activated (e.g., DA_AQ).

4 Conclusions and discussion

Aerosol-cloud interactions can play a crucial role in daily air quality, especially in regions with polluted aerosols under cloud conditions. Major air pollution events in East Asia are frequently associated with extensive cloud cover, which leads to the absence or poor quality of most remote-sensing retrievals across the entire region, making in-situ surface measurements the primary source of observed information.

With the goal of enhancing air quality forecasting in cloudy conditions associated with aerosol wet removals, the WRF-Chem/WRFDA system is extended for the RACM-MADE-VBS-AQCHEM scheme (chem_opt = 109) to assimilate surface measurements of $PM_{2.5}$, PM_{10}, SO_2, NO_2, O_3, and CO concentrations. The reliability and the effects of data assimilation using aqueous chemistry are demonstrated through regional air quality cycling where chemical and weather observations are concurrently assimilated over East Asia every 6 h from February 21 to March, 2019.

By introducing aerosols in the aqueous (or cloud water) phase in WRFDA, the regional cycling system using the online-coupled forecast model could represent the below-cloud wet scavenging of aerosol particles, leading to changes in PM concentrations reaching the ground. The use of aqueous chemistry for aerosol-cloud interactions requires a double-moment microphysics, for which the Morrison two-moment scheme is employed along with Grell-3 cumulus and the RRTMG short- and long-wave schemes.

For the entire cycling period, a time series of innovations ((o-b)'s) in surface PM concentrations remains stable, confirming the successful implementation of the data assimilation system for the particular chemistry option (chem_opt = 109).

Although only surface mass concentrations are assimilated, the impact was recognized throughout the troposphere based on the background error statistics of each aerosol species. Both aqueous chemistry (AQ) and data assimilation (DA) systematically changed the atmospheric composition and its vertical structure, increasing nitrate in the accumulation mode and sea salt aerosols in the coarse mode near the surface. During the month of March 2019, data assimilation tend to considerably increase aerosol species in the accumulation mode within the boundary layer while aqueous chemistry significantly reduces coarse-mode aerosol particles up to mid troposphere. When accounted for in clouds, sulfate aerosols experience large increases in the aqueous phase, in association with the oxidation in low-level clouds.

In a pollution event with high cloud cover, data assimilation with aqueous chemistry was particularly helpful in simulating wet deposition of aerosols to accurately predict the evolution of surface PM concentrations. As the activation, resuspension, and wet scavenging processes in association with cloud chemistry can be all simulated only when aerosols in cloud water are defined through aqueous chemistry, DA without aqueous chemistry treated all the aerosols as interstitial (e.g., suspended in the air) even when precipitation occurred, leading to a significant overestimation of surface PM concentrations. At that time, large LWP was also produced over a wide range of the domain, demonstrating that the formation and development of clouds were also largely affected by aerosols in the aqueous phase.

The use of aqueous chemistry in the aerosol cycling system is beneficial from several perspectives. First, when air pollutants are transported with significant cloud cover, aerosols can be considerably affected by clouds, and in turn, they can influence cloud development and properties. As shown in the presented case study, such interactions can affect wet scavenging of aerosol particles, significantly reducing false alarms in surface PM concentrations. Second, cloudy conditions are generally hard to observe, especially with remote-sensing data, making data assimilation more challenging or susceptible to cloud contamination. In the strong constraint variational data assimilation system, where model errors are not taken into account, the solution (e.g., analysis) might be subopti-

mal when model errors become large. In other words, reducing model errors through advanced physics mechanisms (such as aqueous chemistry along with cloud chemistry and wet scavenging) can make data assimilation more effective, as the assumption of no model errors is violated to a lesser extent. Lastly, the effects of the analysis can accumulate over time through cycling to make systematic improvements in air quality forecasting.

Aqueous chemistry currently implemented in the WRF-Chem model is designed for warm-rain processes by treating aerosols only in the cloud-water phase. And aqueous chemistry implemented for the RACM-MADE-VBS scheme in WRF-Chem does not account for all the complex aqueous-phase reactions, either. For example, the contribution to SOA concentration by cloud chemistry is missing, and aerosol interactions with ice nuclei are not taken into account in this version of the model (Tuccella et al., 2015). As such, the aqueous chemistry used in the chemical option might be overly simple to represent all the physical processes for indirect aerosol effects, especially for mixed-phase convective clouds with nonprecipitating supercooled liquid water near cloud tops (Rosenfeld et al., 2008).

In the spring case examined here, however, the new chemistry option was clearly helpful to simulate the reduction of PM concentrations due to wet removal of aerosol particles. In cloudy conditions that do not result in precipitation, however, enhanced aerosol concentrations in the atmosphere can act to reduce the mean size of cloud droplets and suppress coalescence and warm-rain processes, while enhancing the growth of large hail and cold-rain processes. Those cases cannot be simulated in the model with such simple aqueous chemistry, which can mislead the analysis. As the strong-constraint 3D-Var system used in this study does not include any specific model error term, model errors are not investigated nor discussed in detail, but there is room for further improvements for the RACM-MADE-VBS-AQCHEM option in the model to account for more sophisticated aerosol effects in clouds, radiation, and precipitation.

On the other hand, in the WRFDA system developed for this study, aerosol number concentrations are not included as part of analysis (or control) variables so that they are not changed through the assimilation. To fully describe aerosol impacts on clouds or to handle complex cases with mix-phased clouds or cold-rain processes, we might need to consider developing the assimilation system to reflect the changes in aerosol number concentrations per aerosol (size) mode. Recently, cloud properties and/or atmospheric constituents have increasingly been measured or derived from multiple platforms on ground- and space-born satellites. Two notable examples are the National Aeronautics and Space Administration (NASA)'s Tropospheric Emissions Monitoring of Pollution instrument (TEMPO; https://tempo.si.edu/index.html, last access: 5 July 2023) and South Korea's Geostationary Environment Monitoring Spectrometer (GEMS; Kim et al. (2020)) onboard the Geostationary Korean Multi-purpose Satellite 2B (GEO-KOMPSAT-2B; https://nesc.nier.go.kr/product/vi last access: 5 July 2023). Needless to say, these data can be extremely valuable to not only evaluate the modeling system but also to better initialize the model through data assimilation.

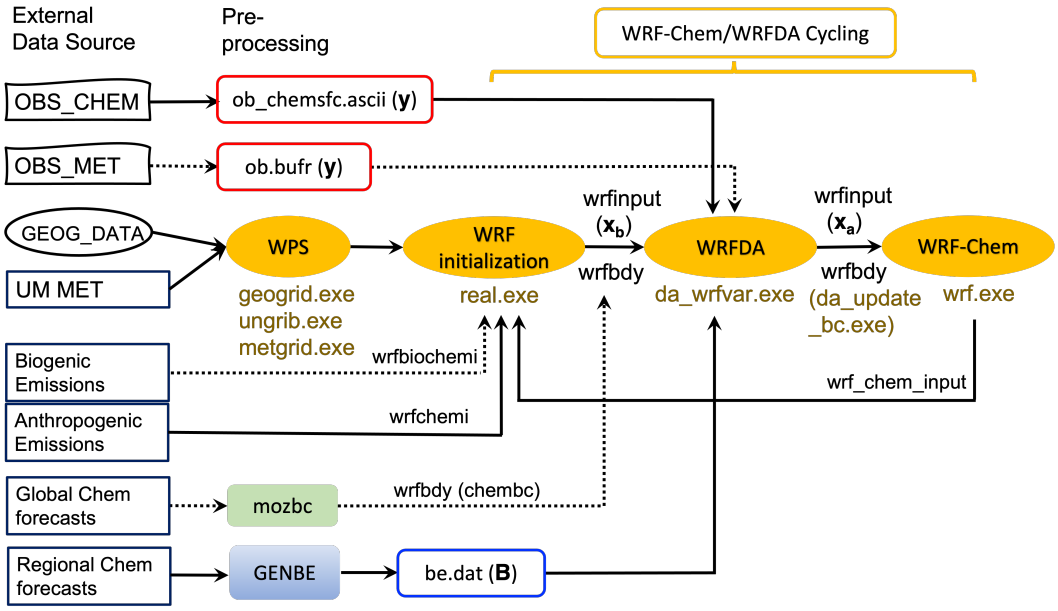


Figure 1. Flowchart of the WRF-Chem/WRFDA cycling system for chemical data assimilation. Gridded input data is indicated by rectangular boxes on the left, and all the software packages are filled in colors. Dotted lines imply optional input data, while solid lines are mandatory inputs for WRF-Chem/WRFDA cycling, accompanied by typical input file names (without specification of domain ID and time) used in the WRF system.

5 Open Research

All meteorological observations assimilated in this study are obtained from the National Centers for Environmental Prediction (NCEP) prepbufr data (https://rda.ucar.edu/datasets/ds337.0/) while the monthly wet deposition observations at Korean EANET (Acid Deposition Monitoring Network in East Asia) sites are read from the website (https://monitoring.eanet.asia/document/public/index/) for March 2019. Chemical measurements at the surface are collected from http://www.airkorea.or.kr for Korean sites and from http://www.cnemc.cn for Chinse sites. The observation data processed for the experiments presented in this study are available in Ha (2023b). The emissions data used in the case study are accessible in Ha (2023a). The base version (V4.4.1) of the WRF system is publicly released in https://github.com/wrf-model/WRF/releases/tag/v4.4.1, and the updated codes introduced for the new features in WRF-Chem/WRFDA are accessible in Ha (2023c). Figure 1 is produced in Microsoft Powerpoint and all other figures are produced in Python V3.8.13. VIIRS Level 3 gridded AOD data in netCDF4 are downloaded and displayed using the scripts provided in https://www.star.nesdis.noaa.gov/smcd/emb/viirs_aerosol/sof which is publicly available.

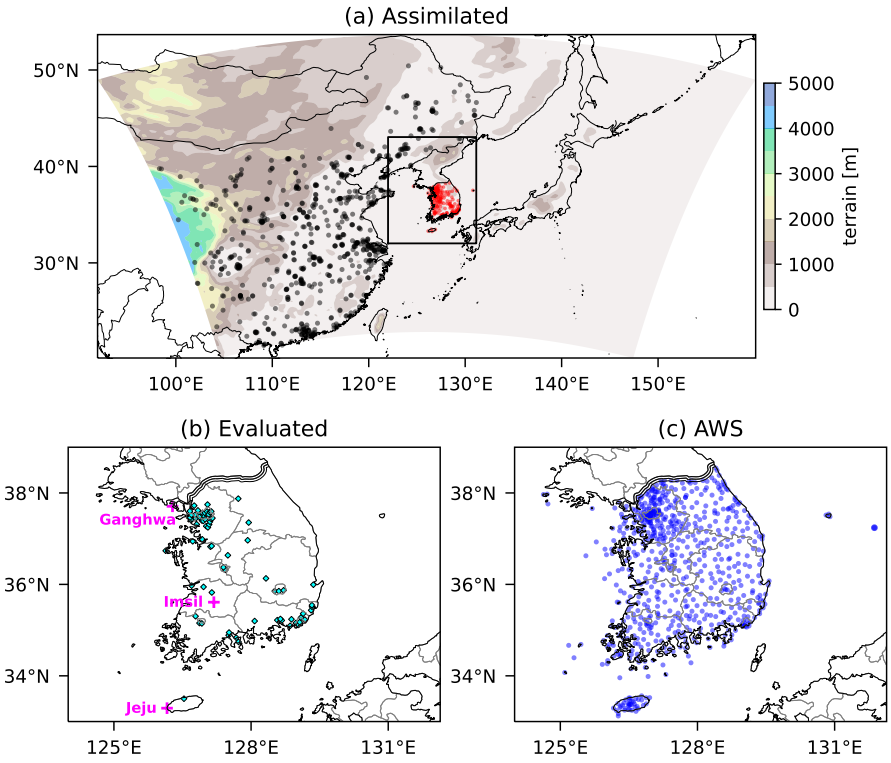


Figure 2. Surface observing network used in this study: a) Assimilated observation sites are marked as dots in red (black) for Korean (Chinese) sites, with terrain height colored in domain 1 (at 27 km resolution) and a black box over the Korean peninsula indicating domain 2 (D2; 9 km resolution) b) 71 evaluation sites are marked in light blue dots while three EANET sites - Ganghwa, Imsil, and Jeju - are displayed in pink ("+" sign) along with their site names and c) the Korean automated weather stations (AWS) measuring meteorological variables at the surface are shown in blue dots.

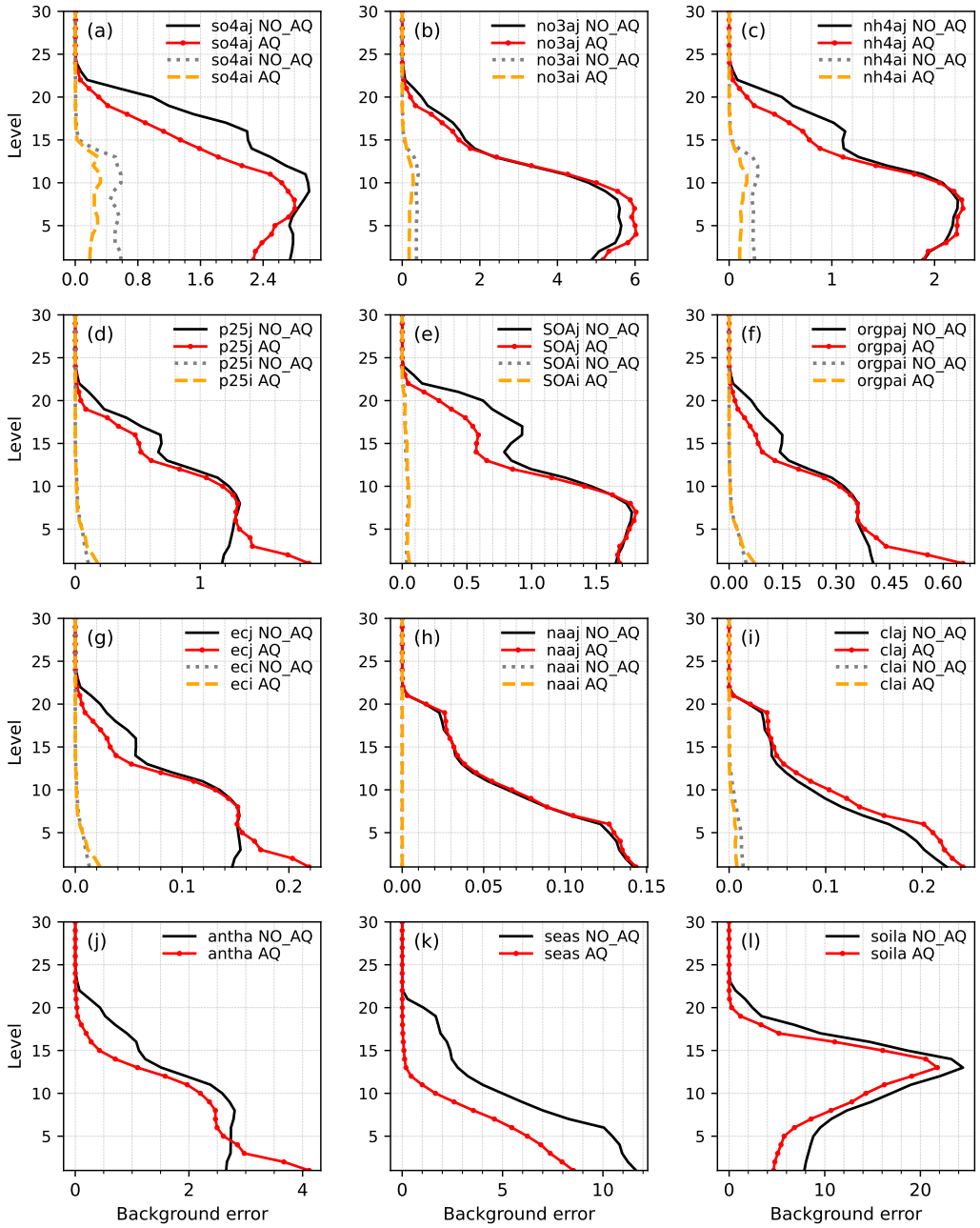


Figure 3. Vertical profile of each aerosol species in background error standard deviation estimated with and without aqueous chemistry (AQ and NO_AQ, respectively) over domain 2 (D2). Accumulation mode aerosols in AQ (NO_AQ) are depicted in red lines with dots (black solid lines) while Aitken mode aerosols in AQ (NO_AQ) in dashed orange (dotted gray) lines.

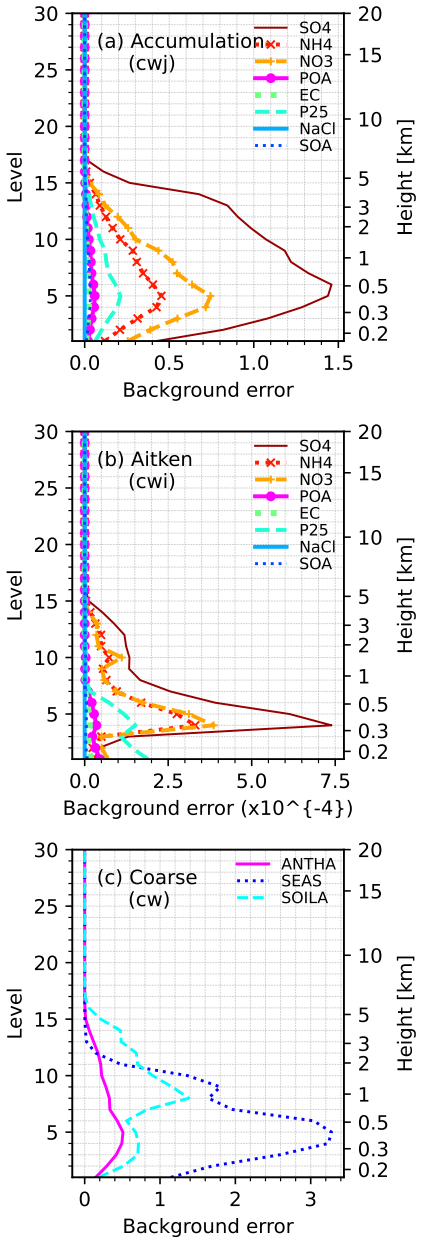


Figure 4. Vertical profile of background error standard deviation with aqueous chemistry (AQ) for aerosol species in the aqueous or cloud water ("cw") phase in domain 2. Domain-averaged height ([km]), corresponding to the model levels, is displayed on the right y-axis.

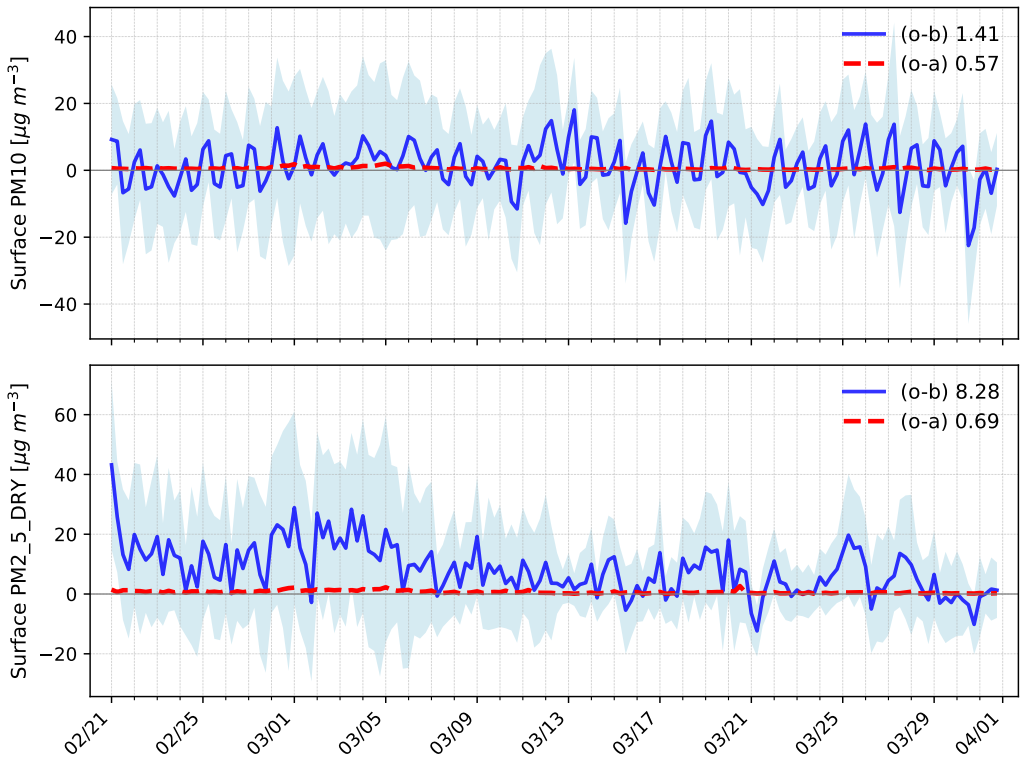


Figure 5. Time series of (o-a; red dashed lines) and (o-b; blue solid lines) in $PM_{2.5}$ (top) and PM_{10} concentrations (bottom) on the ground in DA_AQ in domain 2. The lines indicate the mean over all the assimilated stations in South Korea while the light blue shading area shows standard deviation in (o-b)'s across the surface stations. The numbers in the legend indicate the mean over the entire cycling period from Feb 21 to March 31, 2019.

Table S1. Physics and chemical options

Scheme	Option	Namelist_option
Chemistry driver	RACM-MADE-VBS-AQCHEM	chem_opt = 109
Photolysis	Madronich	phot_opt = 1
Grid-scale Wet scavenging	On	wetscav_onoff = 1
Grid-scale Cloud chemistry	On	cldchem_onoff = 1
Convective transport and wet deposition	On	conv_tr_wetscav = 1
Convective-scale aqueous chemistry	On	conv_tr_aqchem = 1
Dust emissions	On	dust_opt = 13
Sea salt emissions	On	seas_opt = 2
Prognostic Number Concentration	On	progn = 1
Microphysics	Morrison 2-moment	mp_physics = 10
Longwave radiation	RRTMG	ra_lw_physics = 4
Shortwave radiation	RRTMG	ra_sw_physics = 4
Surface layer	Monin–Obukhov	sf_sfclay_physics = 1
Land surface	Noah	sf_surface_physics = 1
Boundary layer	YSU	bl_pbl_physics = 1
Cumulus parametrization	Grell-3D	cu_physics = 5

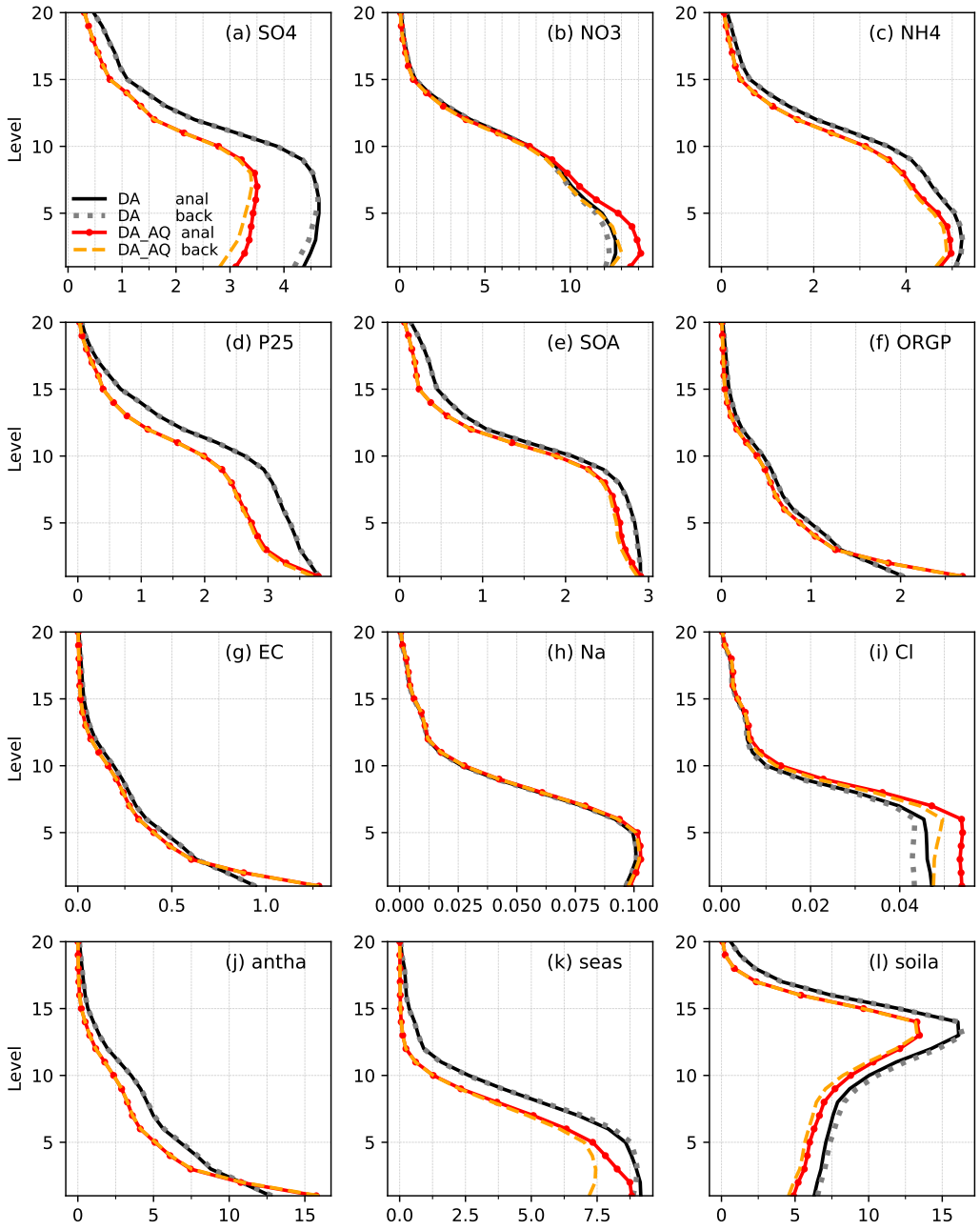


Figure 6. Vertical profile of each aerosol species ([μg/kg-dryair]) in the analysis and background forecast in domain 2 with and without aqueous chemistry (DA_AQ and DA, respectively), averaged over 71 verification sites in Korea (light blue dots in Fig.2) from 6-hr cycling for the period of 1-31 March 2019.

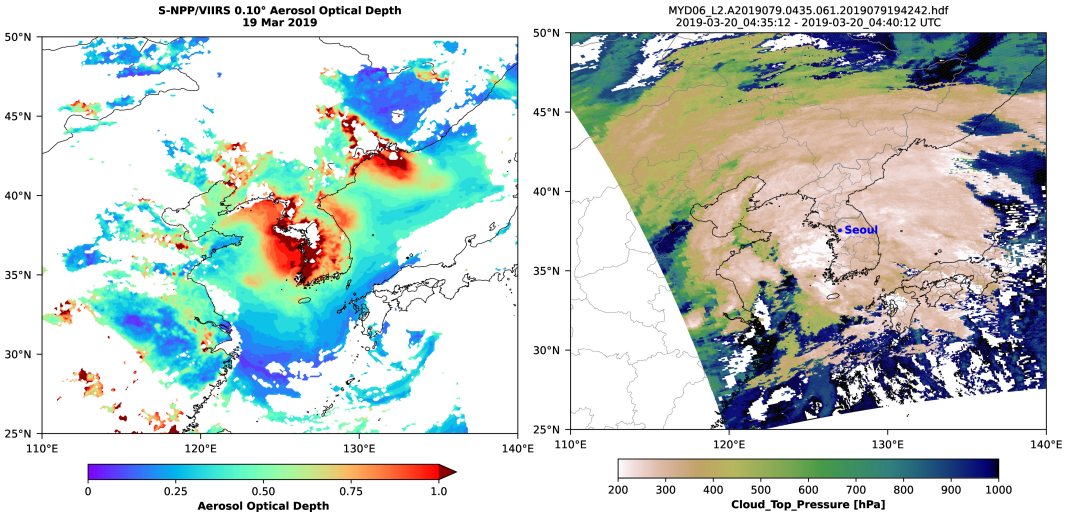


Figure 7. (a) Aerosol optical depth retrieved from VIIRS onboard the Suomi NPP as a daily mean gridded data (level 3) on Mar 19, 2019 (b) Level 2 cloud top pressure retrieved from the MODIS sensors onboard the Aqua satellite, merged between 04:35:12 and 04:40:12 UTC on 20 March 2019.

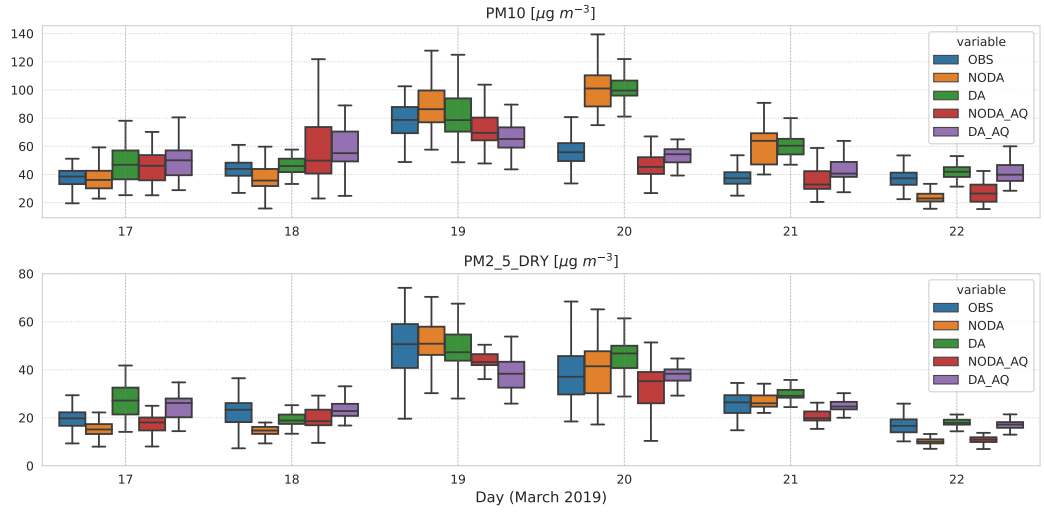


Figure 8. Time series of boxplots for daily PM simulations in each experiment compared to observations. Each day displays 24-hour forecasts initiated from the 00Z analysis in each experiment at 71 verification sites in Korea.

Table S2. Experiments. DA stands for data assimilation and AQ aqueous chemistry.

Experiment	chem_opt	Assimilation
NODA	108	None
DA	108	CHEM+MET
NODA_AQ	109	None
DA_AQ	109	CHEM+MET

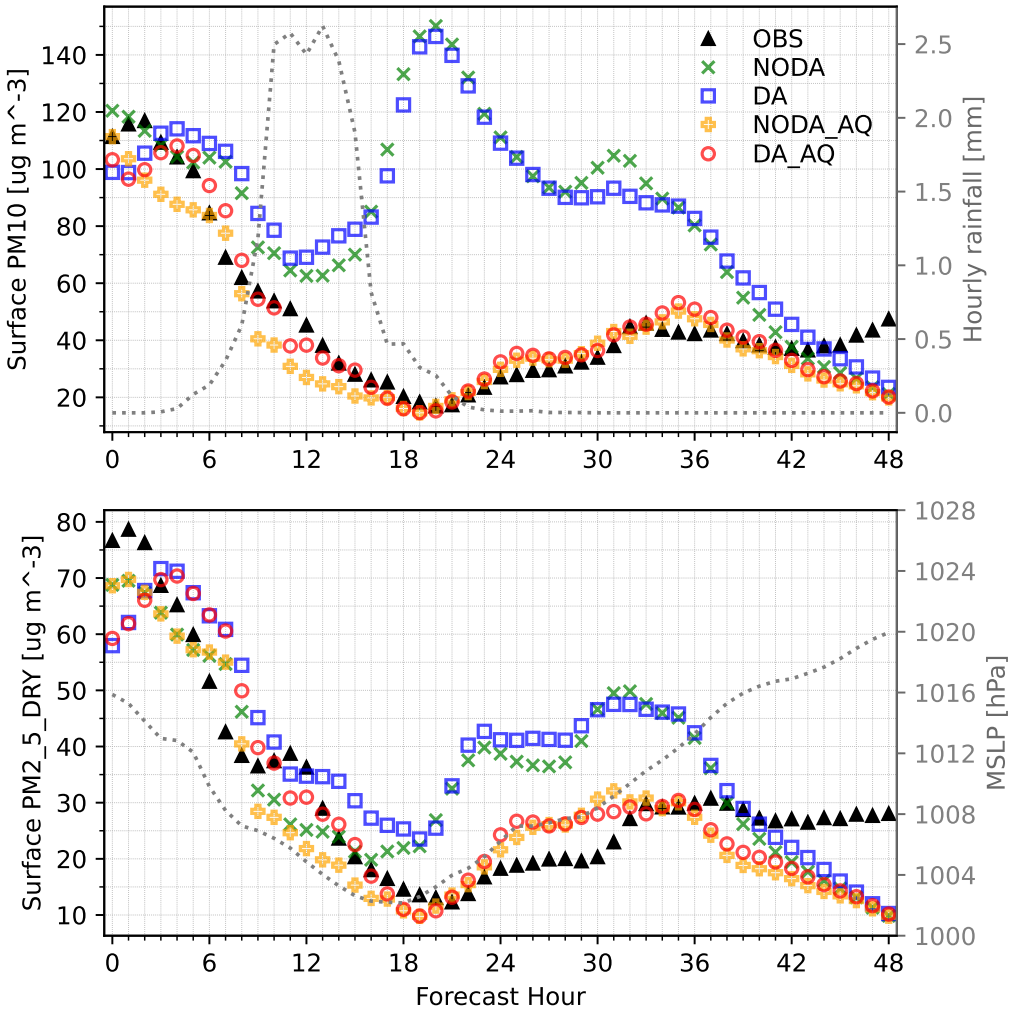


Figure 9. Time series of (top) surface PM_{10} and (bottom) $PM_{2.5}$ concentrations for 48 h forecasts from 00 UTC 20 March 2019, averaged over 71 Korean verification sites (marked in Fig. 2 (b)). In-situ observations (OBS; black triangle) are compared with hourly forecasts from their 00Z analysis in each cycling experiment. Gray dotted lines with the right y-axis are (top) hourly rainfall (mm) and mean sea level pressure (hPa) observations (bottom) averaged over 699 AWS sites over South Korea (marked in Fig. 2 (c)).

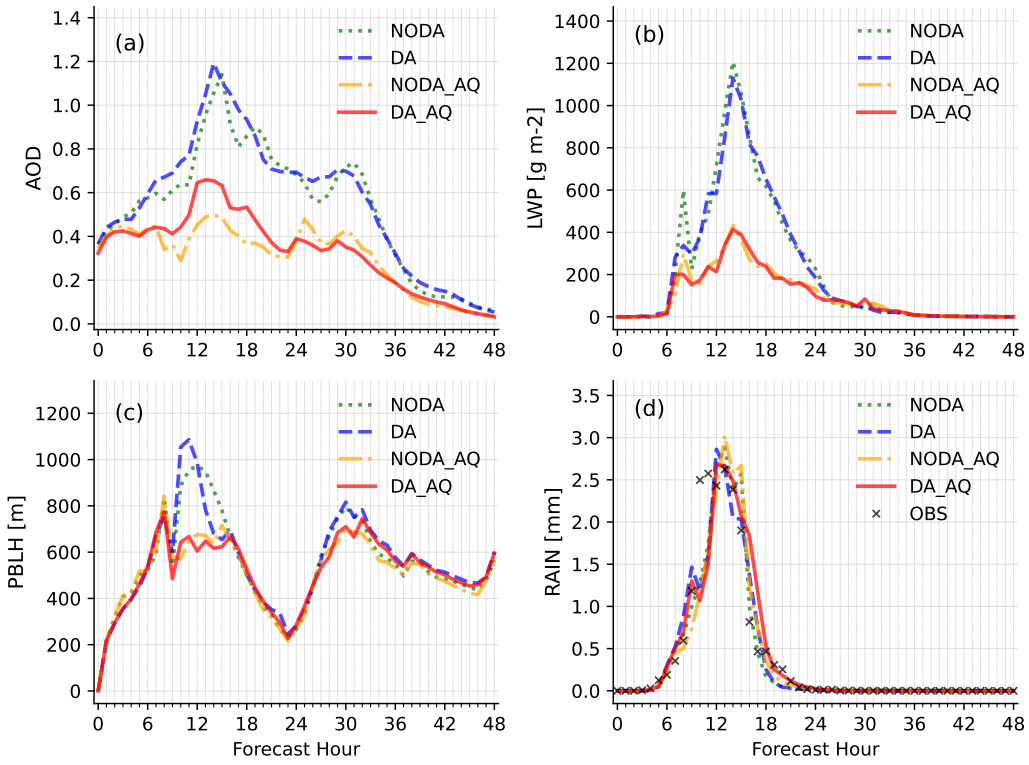


Figure 10. Same as Fig.9, but for (a) AOD (b) LWP (c) PBLH and (d) hourly rain accumulation for each experiment. In (d), hourly observations averaged over 699 AWS sites are also marked as a "x" symbol.

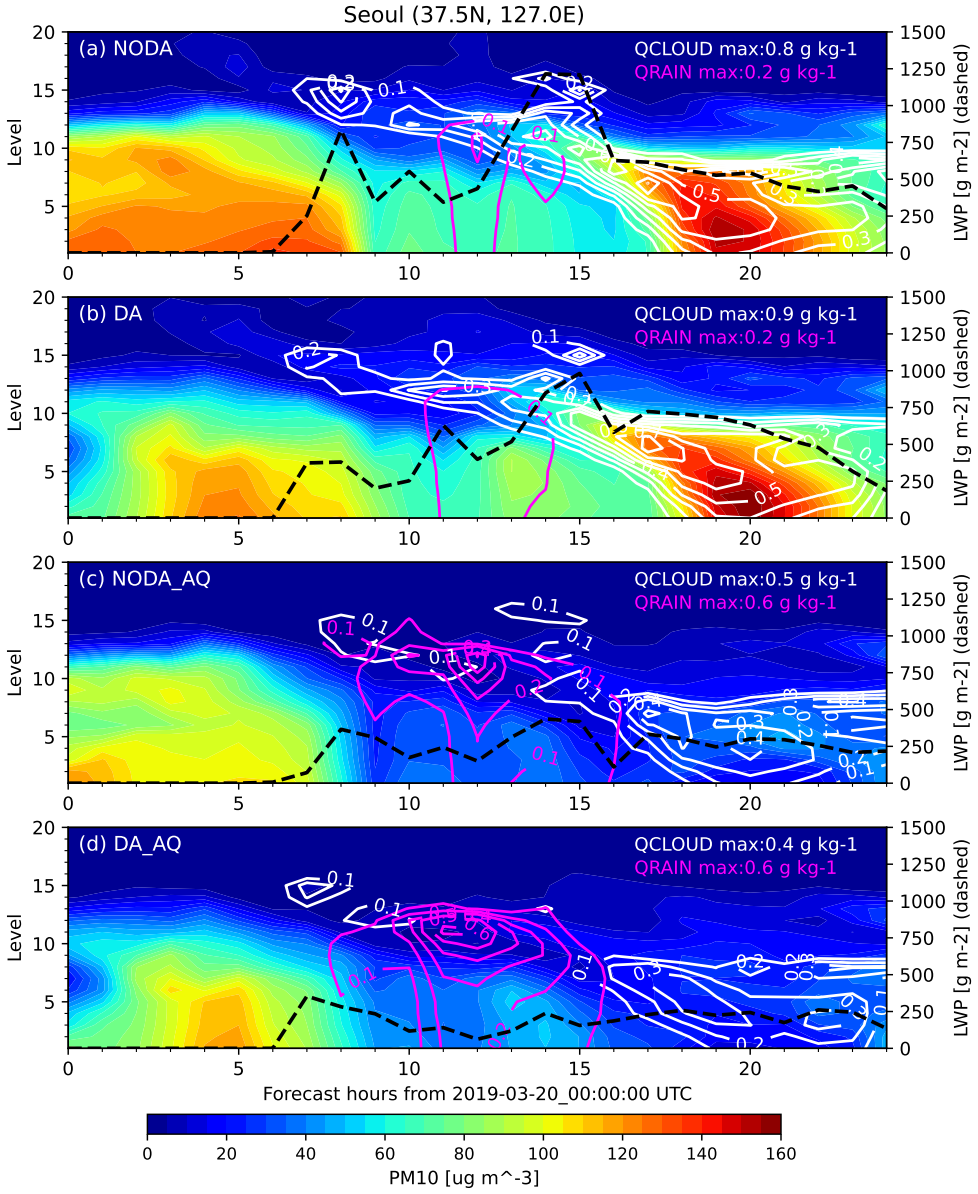


Figure 11. Time series of 24 h forecasts of PM_{10} concentrations (filled) simulated at Seoul, South Korea (in model levels up to 20) in each experiment. Cloud and Rain water mixing ratios (QCLOUD and QRAIN ([g kg-1])) are contoured in white and pink, respectively. Liquid Water Path (LWP) is also plotted in black dashed lines with the right y-axis.

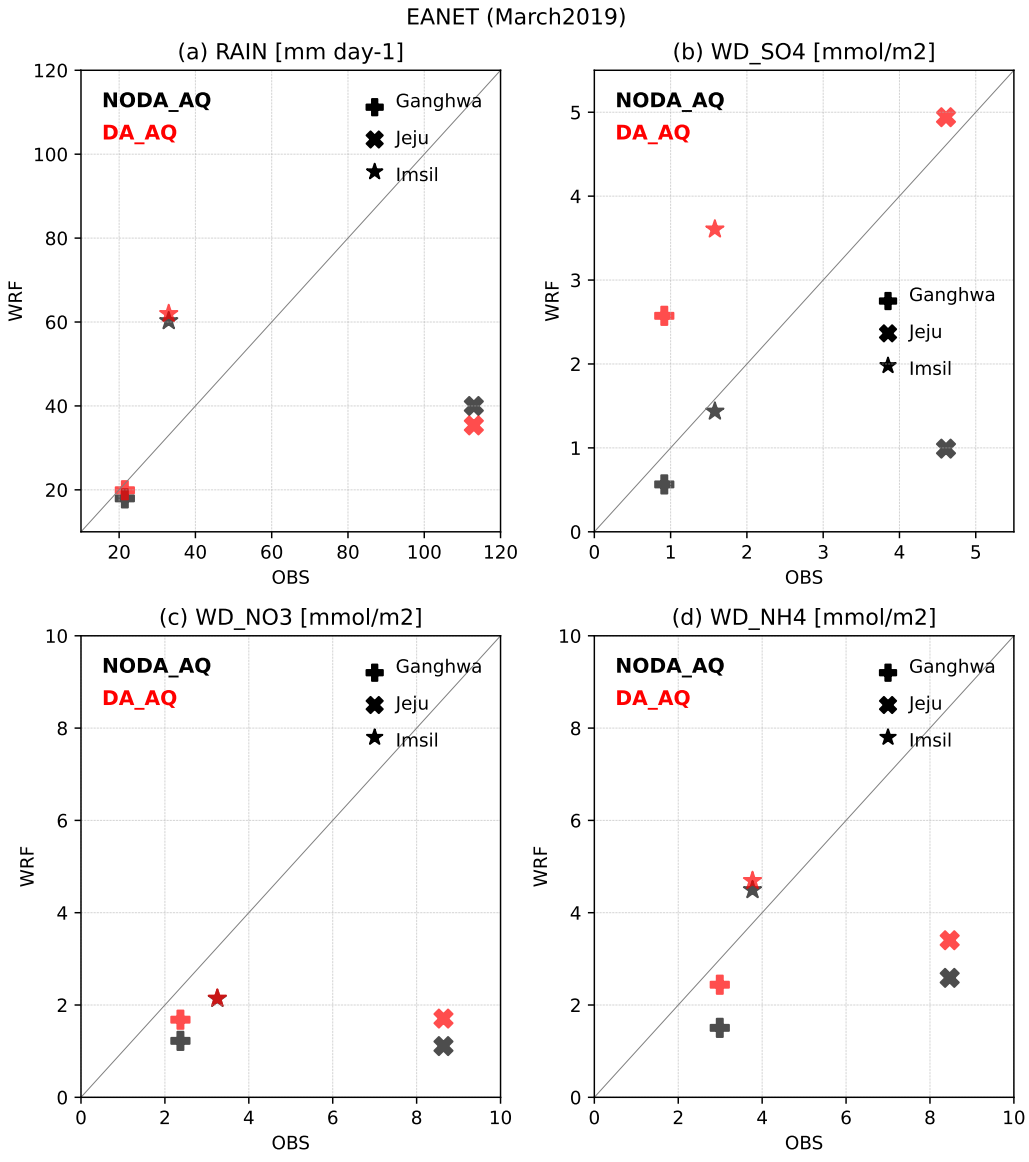


Figure 12. Monthly data for March 2019, comparing two AQ experiments with measures from three EANET sites (marked in Fig. 2 (b)) in (a) daily rain accumulation and wet deposition in (b) SO4 (c) NO3 and (d) NH4.

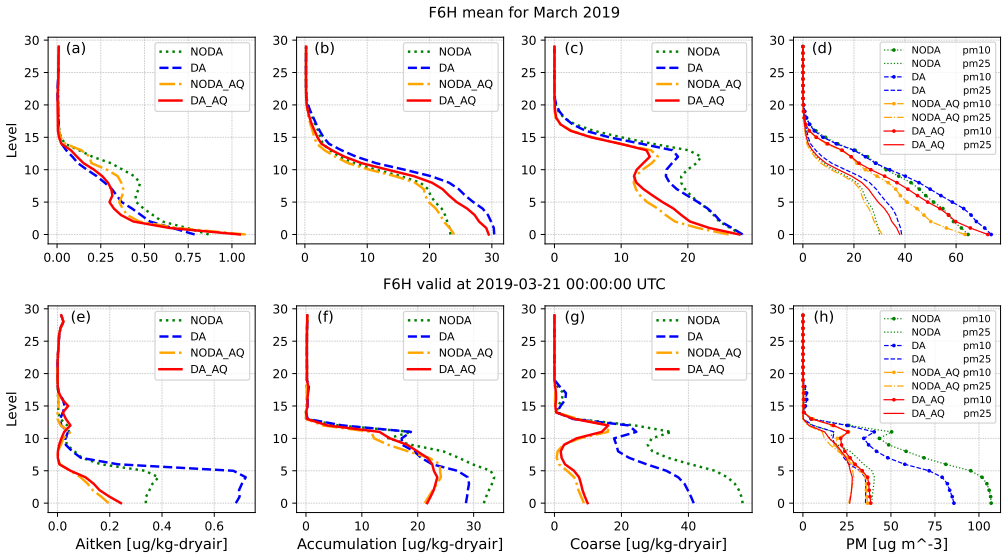


Figure 13. Vertical profile of aerosol species in each mode from (a)-(c) and (e)-(g). $PM_{2.5}$ and PM_{10} are displayed in (d) and (h). Upper panels show the monthly mean for March 2019 while lower panels are 6 hour forecasts from the analysis at 18 UTC 20 March 2019.

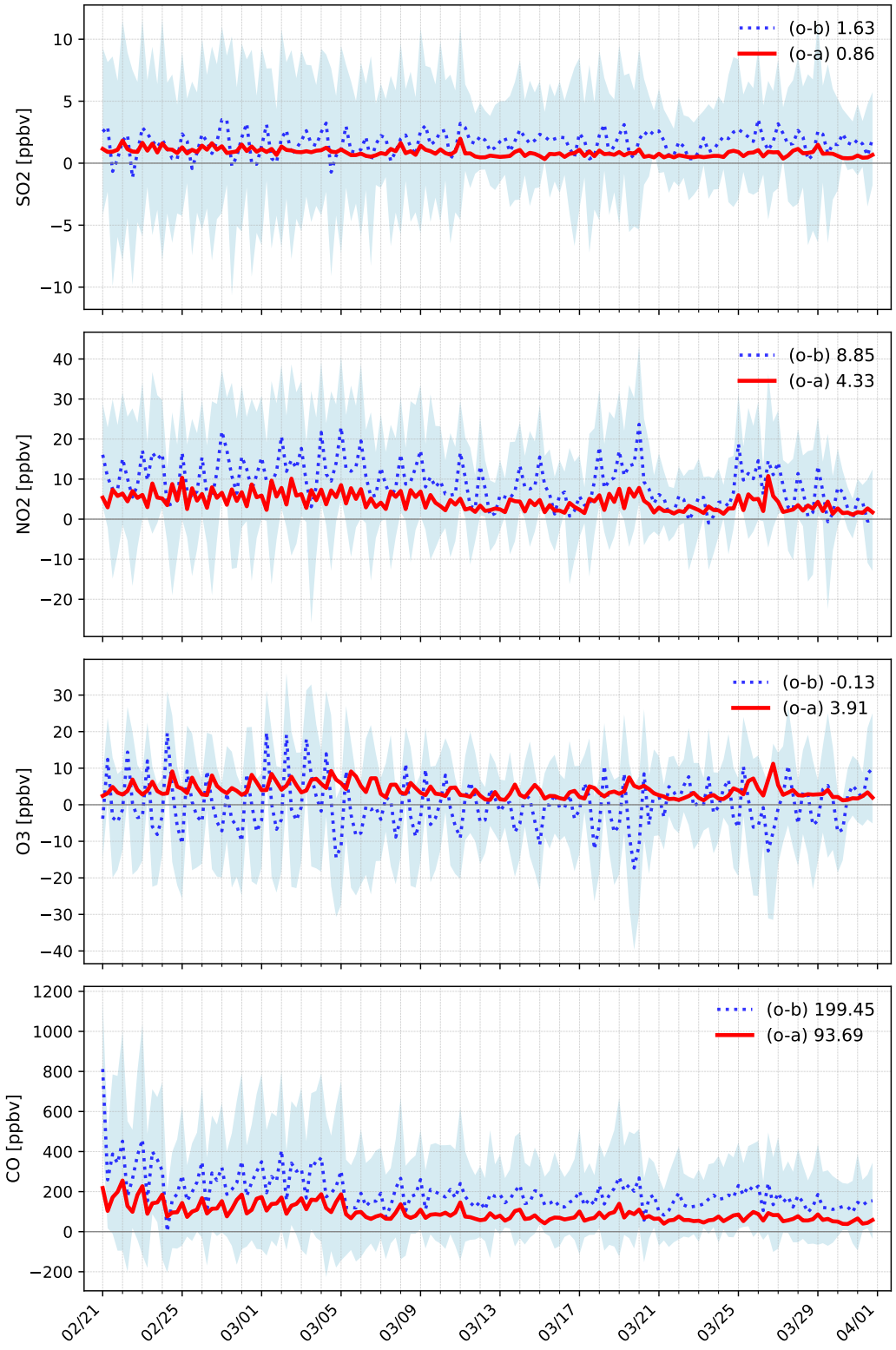


Figure S1. Same as Fig.5, but for four gas species.

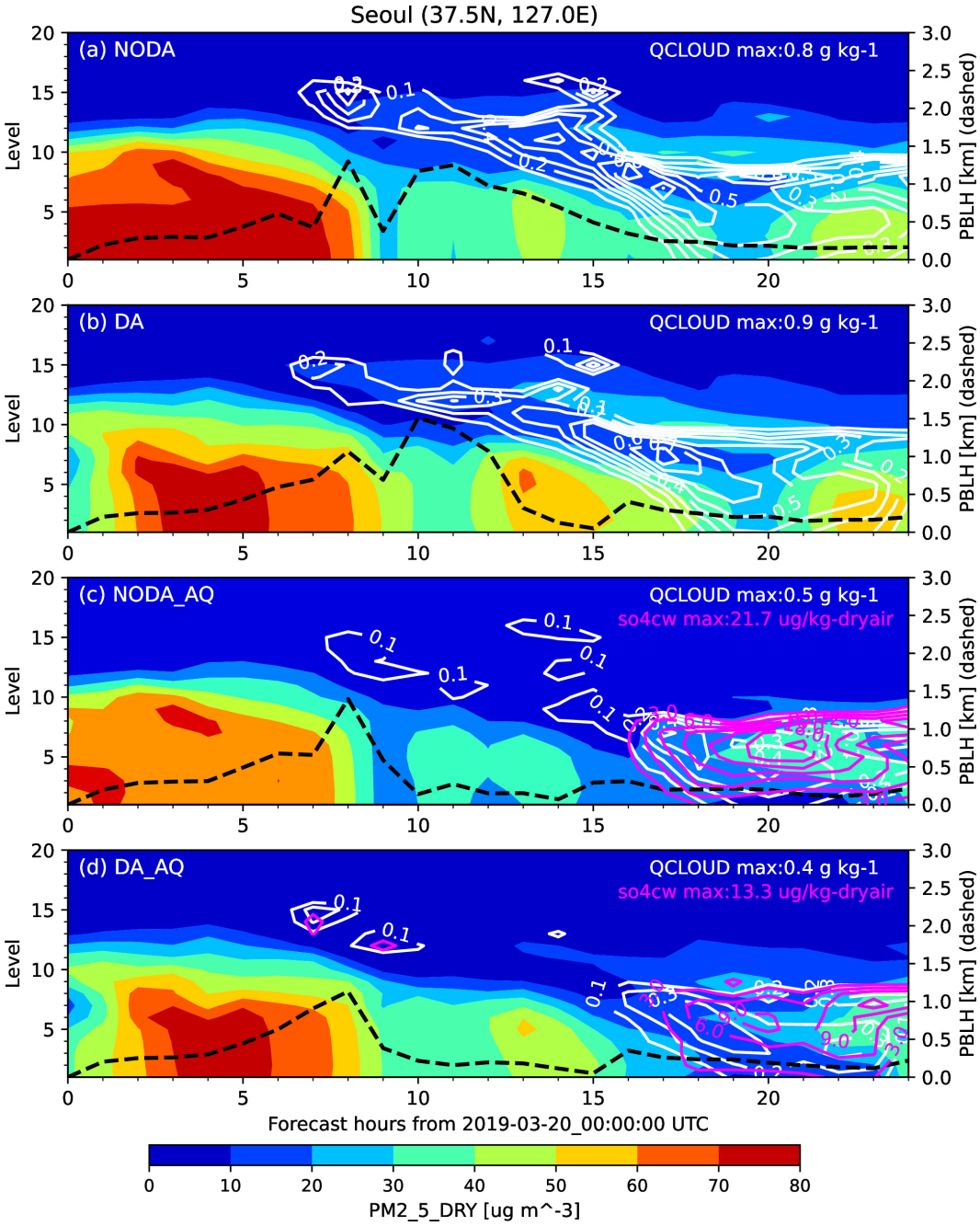


Figure S2. Same as Fig.11, but for $PM_{2.5}$ (filled). The sum of cloud-borne sulfate aerosols in the accumulation and Aitken modes (e.g., so4cwj+so4cwi) is contoured in pink, superimposed with cloud water mixing ratios (QCLOUD ([g kg-1]); white solid lines). The PBL height is also plotted in black dashed lines with the y-axis on the right.

Acknowledgments

This research is based upon work supported by the National Center for Atmospheric Research which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1755088. The code development was also supported in part by a grant from the National Institute of Environment Research (NIER), funded by the Ministry of Environment (MOE) of the Republic of Korea (NIER-2022-01-02-076). We acknowledge the use of the WRF-Chem preprocessor tool named "megan_bio_emiss" provided by the Atmospheric Chemistry Observations and Modeling Lab (ACOM) of NCAR. Big thanks go to Dr. Duseong Jo at ACOM/NCAR for his internal review and fruitful discussion. The first author is also grateful for the Python code for visualizing VIIRS aerosol data provided by the Aerosols and Atmospheric Composition Science Team at the NOAA/NESDIS Center for Satellite Applications and Research (STAR). All the simulations, data processing and analysis were conducted on NCAR's supercomputers (cheyenne and casper).

References

Abdul-Razzak, H., & Ghan, S. J. (2002). A parameterization of aerosol activation 3. sectional representation. *Journal of Geophysical Research: Atmospheres*, *107*(D3). doi: https://doi.org/10.1029/2001JD000483

Ackermann, I. J., Hass, H., Memmesheimer, M., Ebel, A., Binkowski, F. S., & Shankar, U. (1998). Modal aerosol dynamics model for europe: development and first applications. *Atmospheric Environment*, *32*(17), 2981–2999. Retrieved from `https://doi.org/10.1016/S1352-2310(98)00006-5`

Ahmadov, R., McKeen, S. A., Robinson, A. L., Bahreini, R., Middlebrook, A. M., de Gouw, J. A., ... Trainer, M. (2012). A volatility basis set model for summertime secondary organic aerosols over the eastern united states in 2006. *Journal of Geophysical Research: Atmospheres*, *117*(D06301). doi: 10.1029/2011JD016831

Aquila, V., Hendricks, J., Lauer, A., Riemer, N., Vogel, H., Baumgardner, D., ... DallÁmico, M. (2011). Made-in: a new aerosol microphysics submodel for global simulation of insoluble particles and their mixing state. *Geoscientific Model Development*, *4*(2), 325–355. doi: 10.5194/gmd-4-325-2011

Baklanov, A., Brunner, D., Carmichael, G., Flemming, J., Freitas, S., Gauss, M., ... Vogel, B. (2017). Key issues for seamless integrated chemistry–meteorology modeling. *Bulletin of the American Meteorological Society*, *98*(11), 2285-2292. doi: 10.1175/BAMS-D-15-00166.1

Baklanov, A., Schlünzen, K., Suppan, P., Baldasano, J., Brunner, D., Aksoyoglu, S., ... Zhang, Y. (2014). Online coupled regional meteorology chemistry models in europe: current status and prospects. *Atmospheric Chemistry and Physics*, *14*(1), 317–398. doi: 10.5194/acp-14-317-2014

Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Žabkar, R., Carmichael, G. R., ... Seigneur, C. (2015). Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models. *Atmospheric Chemistry and Physics*, *15*(10), 5325–5358. doi: 10.5194/acp-15-5325-2015

Chapman, E. G., Gustafson Jr, W., Easter, R. C., Barnard, J. C., Ghan, S. J., Pekour, M. S., & Fast, J. D. (2009). Coupling aerosol-cloud-radiative processes in the wrf-chem model: Investigating the radiative impact of elevated point sources. *Atmospheric Chemistry and Physics*, *9*(3), 945–964.

Chen, D., Liu, Z., Ban, J., Zhao, P., & Chen, M. (2019). Retrospective analysis of 2015-2017 wintertime pm_2.5 in china: response to emission regulations and the role of meteorology. *Atmospheric Chemistry and Physics*, *19*(11), 7409–7427. doi: 10.5194/acp-19-7409-2019

Courtier, P., Thépaut, J.-N., & Hollingsworth, A. (1994). A strategy for operational implementation of 4d-var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, *120*(519), 1367-1387. doi: 10.1002/qj.49712051912

Dee, D. P. (2005). Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 3323—3343. doi: doi:10.1256/qj.05.137

Dee, D. P., & Da Silva, A. M. (1998). Data assimilation in the presence of forecast bias. *Quarterly Journal of the Royal Meteorological Society*, *124*(545), 269–295.

Descombes, G., Auligné, T., Vandenberghe, F., Barker, D. M., & Barré, J. (2015). Generalized background error covariance matrix model (gen_be v2.0). *Geosci. Model Dev.*, *8*, 669–696. doi: 10.5194/gmd-8-669-2015

Eck, T. F., Holben, B., Kim, J., Beyersdorf, A., Choi, M., Lee, S., ... Park, S. (2020). Influence of cloud, fog, and high relative humidity during pollution transport events in south korea: Aerosol properties and pm2.5 variability. *Atmospheric Environment*, *232*, 117530. doi: https://doi.org/10.1016/j.atmosenv.2020.117530

Eck, T. F., Holben, B. N., Reid, J. S., Xian, P., Giles, D. M., Sinyuk, A., ... Xia, X. (2018). Observations of the interaction and transport of fine mode aerosols with cloud and/or fog in northeast asia from aerosol robotic network and satellite remote sensing. *Journal of Geophysical Research: Atmospheres*, *123*(10), 5560-5587. doi: https://doi.org/10.1029/2018JD028313

Ervens, B. (2015). Modeling the processing of aerosol and trace gases in clouds and fogs. *Chemical Reviews*, *115*(10), 4157-4198. Retrieved from `https://doi.org/10.1021/cr5005887` doi: 10.1021/cr5005887

Fast, J. D., Gustafson Jr., W. I., Easter, R. C., Zaveri, R. A., Barnard, J. C., Chapman, E. G., ... Peckham, S. E. (2006). Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of houston using a fully coupled meteorology-chemistry-aerosol model. *Journal of Geophysical Research: Atmospheres*, *111*(D21305). doi: 10.1029/2005JD006721

Friedl, M., McIver, D., Hodges, J., Zhang, X., Muchoney, D., Strahler, A., ... Schaaf, C. (2002). Global land cover mapping from modis: algorithms and early results. *Remote Sensing of Environment*, *83*(1), 287 - 302. Retrieved from `https://doi.org/10.1016/S0034-4257(02)00078-0`

Ghan, S., Laulainen, N., Easter, R., Wagener, R., Nemesure, S., Chapman, E., ... Leung, R. (2001). Evaluation of aerosol direct radiative forcing in mirage. *J. Geophys. Res.*, *106*(D6), 5295–5316. doi: 10.1029/2000JD900502

Grell, G. A., & Baklanov, A. (2011). Integrated modeling for forecasting weather and air quality: A call for fully coupled approaches. *Atmospheric Environment*, *45*(38), 6845 - 6851. Retrieved from `https://doi.org/10.1016/j.atmosenv.2011.01.017`

Grell, G. A., & Dévényi, D. (2002). A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophysical Research Letters*, *29*(14), 38-1-38-4. doi: https://doi.org/10.1029/2002GL015311

Grell, G. A., Peckham, S., Schmitz, R., McKeen, S., Frost, G., Skamarock, W. C., & Eder, B. (2005). Fully coupled "online" chemistry within the WRF model. *Atmos. Environ.*, *39*(37), 6957–6975. Retrieved from `http://dx.doi.org/10.1016/j.atmosenv.2005.04.027`

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., & Geron, C. (2006). Estimates of global terrestrial isoprene emissions using megan (model of emissions of gases and aerosols from nature. *Atmospheric Chemistry and Physics*, *6*(11), 3181-3210. Retrieved from `https://hal.archives-ouvertes.fr/hal-00295995`

Ha, S. (2022). Implementation of aerosol data assimilation in wrfda (v4.0.3) for wrf-chem (v3.9.1) using the racm/made-vbs scheme. *Geoscientific Model Development*, *15*(4), 1769–1788. doi: 10.5194/gmd-15-1769-2022

Ha, S. (2023a, July). *Emissions data.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.8133021` doi: 10.5281/zenodo.8133021

Ha, S. (2023b). *Surface chemical observations.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.8132952` doi: 10.5281/zenodo.8132952

Ha, S. (2023c, July). *Wrf updates for chem_opt=109.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.8133100` doi: 10.5281/zenodo.8133100

Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, *134*, 2318–2341. doi: 10.1175/MWR3199.1

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

Kim, J., Jeong, U., Ahn, M.-H., Kim, J. H., Park, R. J., Lee, H., ... Choi, Y. (2020). New era of air quality monitoring from space: Geostationary environ-

ment monitoring spectrometer (gems). *Bulletin of the American Meteorological Society*, *101*(1), E1–E22. doi: https://doi.org/10.1175/BAMS-D-18-0013.1

Kumar, R., Delle Monache, L., Bresch, J., Saide, P. E., Tang, Y., Liu, Z., ... Djalalova, I. (2019). Toward improving short-term predictions of fine particulate matter over the united states via assimilation of satellite aerosol optical depth retrievals. *Journal of Geophysical Research: Atmospheres*. doi: 10.1029/2018JD029009

Lin, Y.-L., Farley, R. D., & Orville, H. D. (1983). Bulk parameterization of the snow field in a cloud model. *Journal of Climate and Applied Meteorology*, *22*(6), 1065–1092. doi: 10.1175/1520-0450(1983)022⟨1065:BPOTSF⟩2.0.CO;2

Lorenc, A. (1986). Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, *112*, 1177–1194.

Madronich, S. (1987). Photodissociation in the atmosphere: 1. actinic flux and the effect of ground reflections and clouds. *J. Geophys. Res.*, *92*, 9740–9752.

Ménard, R., Gauthier, P., Rochon, Y., Robichaud, A., de Grandpré, J., Yang, Y., ... Chabrillat, S. (2019). Coupled stratospheric chemistry-meteorology data assimilation. part ii: Weak and strong coupling. *Atmosphere*, *10*(12). Retrieved from `https://www.mdpi.com/2073-4433/10/12/798` doi: 10.3390/atmos10120798

Morrison, H., Thompson, G., & Tatarskii, V. (2009). Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Monthly Weather Review*, *137*(3), 991–1007. doi: 10.1175/2008MWR2556.1

Ntelekos, A., Smith, J., Donner, L., Fast, J., Gustafson, W., Chapman, E., & Krajewski, W. (2009). The effects of aerosols on intense convective precipitation in the northeastern united states. *Quarterly Journal of the Royal Meteorological Society*, *135*(643), 1367–1391. doi: 10.1002/qj.476

Parrish, D. F., & Derber, J. C. (1992). The national meteorological center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, *120*(8), 1747–1763.

Pfister, G. G., Avise, J., Wiedinmyer, C., Edwards, D. P., Emmons, L. K., Diskin, G. D., ... Wisthaler, A. (2011). Co source contribution analysis for california during arctas-carb. *Atmospheric Chemistry and Physics*, *11*(15), 7515–7532. doi: 10.5194/acp-11-7515-2011

Pfister, G. G., Eastham, S. D., Arellano, A. F., Aumont, B., Barsanti, K. C., Barth, M. C., ... Brasseur, G. P. (2020). The multi-scale infrastructure for chemistry and aerosols (musica). *Bulletin of the American Meteorological Society*, *101*(10), E1743 - E1760. doi: 10.1175/BAMS-D-19-0331.1

Rosenfeld, D., Lohmann, U., Raga, G. B., O'Dowd, C. D., Kulmala, M., Fuzzi, S., ... Andreae, M. O. (2008). Flood or drought: How do aerosols affect precipitation? *science*, *321*(5894), 1309–1313.

Ryu, Y.-H., Min, S.-K., & Knote, C. (2022a). Contrasting roles of clouds as a sink and source of aerosols: A quantitative assessment using wrf-chem over east asia. *Atmospheric Environment*, *277*, 119073. doi: https://doi.org/10.1016/j.atmosenv.2022.119073

Ryu, Y.-H., Min, S.-K., & Knote, C. (2022b). Role of upwind precipitation in transboundary pollution and secondary aerosol formation: A case study during the korus-aq field campaign. *Journal of Applied Meteorology and Climatology*, *61*(2), 159 - 174. doi: https://doi.org/10.1175/JAMC-D-21-0162.1

Saide, P. E., Spak, S. N., Carmichael, G. R., Mena-Carrasco, M. A., Yang, Q., Howell, S., ... Springston, S. R. (2012). Evaluating wrf-chem aerosol indirect effects in southeast pacific marine stratocumulus during vocals-rex. *Atmospheric Chemistry and Physics*, *12*, 3045—3064. doi: 10.5194/acp-12-3045-2012

Saide, P. E., Spak, S. N., Pierce, R. B., Otkin, J. A., Schaack, T. K., Heidinger, A. K., ... Carmichael, G. R. (2015). Central american biomass burning smoke can increase tornado severity in the u.s. *Geophysical Research Letters*, *42*(3), 956-965. doi: 10.1002/2014GL062826

Seinfeld, J. H., & Pandis, S. N. (2006). *Atmospheric chemistry and physics: From air pollution to climate change* (Second ed.). New York: John Wiley & Sons. Inc.

Stevens, B., & Feingold, G. (2009). Untangling aerosol effects on clouds and precipitation in a buffered system. *Nature*, *461*, 607–613. doi: https://doi.org/10.1038/nature08281

Stockwell, W. R., Kirchner, F., Kuhn, M., & Seefeld, S. (1997). A new mechanism for regional atmospheric chemistry modeling. *Journal of Geophysical Research: Atmospheres*, *102*(D22), 25847-25879. doi: 10.1029/97JD00849

Sun, W., Liu, Z., Chen, D., Zhao, P., & Chen, M. (2020). Development and application of the wrfda-chem three-dimensional variational (3dvar) system: aiming to improve air quality forecasting and diagnose model deficiencies. *Atmos. Chem. Phys*, *20*, 9311–9329. Retrieved from `https://doi.org/10.5194/acp-20-9311-2020`

Tao, W.-K., Chen, J.-P., Li, Z., Wang, C., & Zhang, C. (2012). Impact of aerosols on convective clouds and precipitation. *Reviews of Geophysics*, *50*(2). doi: https://doi.org/10.1029/2011RG000369

Tuccella, P., Curci, G., Grell, G. A., Visconti, G., Crumeyrolle, S., Schwarzenboeck, A., & Mensah, A. A. (2015). A new chemistry option in wrf-chem v. 3.4 for the simulation of direct and indirect aerosol effects using vbs: evaluation against impact-eucaari data. *Geoscientific Model Development*, *8*(9), 2749–2776.

Yang, Q., Easter, R. C., Campuzano-Jost, P., Jimenez, J. L., Fast, J. D., Ghan, S. J., ... Wisthaler, A. (2015). Aerosol transport and wet scavenging in deep convective clouds: A case study and model evaluation using a multiple passive tracer analysis approach. *Journal of Geophysical Research: Atmospheres*, *120*(16), 8448-8468. doi: https://doi.org/10.1002/2015JD023647

Yang, Q., Jr., W. I. G., Fast, J. D., Wang, H., Easter, R. C., Morrison, H., ... Mena-Carrasco, M. A. (2011). Assessing regional scale predictions of aerosols, marine stratocumulus, and their interactions during vocals-rex using wrf-chem. *Atmospheric Chemistry and Physics*, *11*(23), 11951–11975. doi: 10.5194/acp-11-11951-2011

Figure 1.

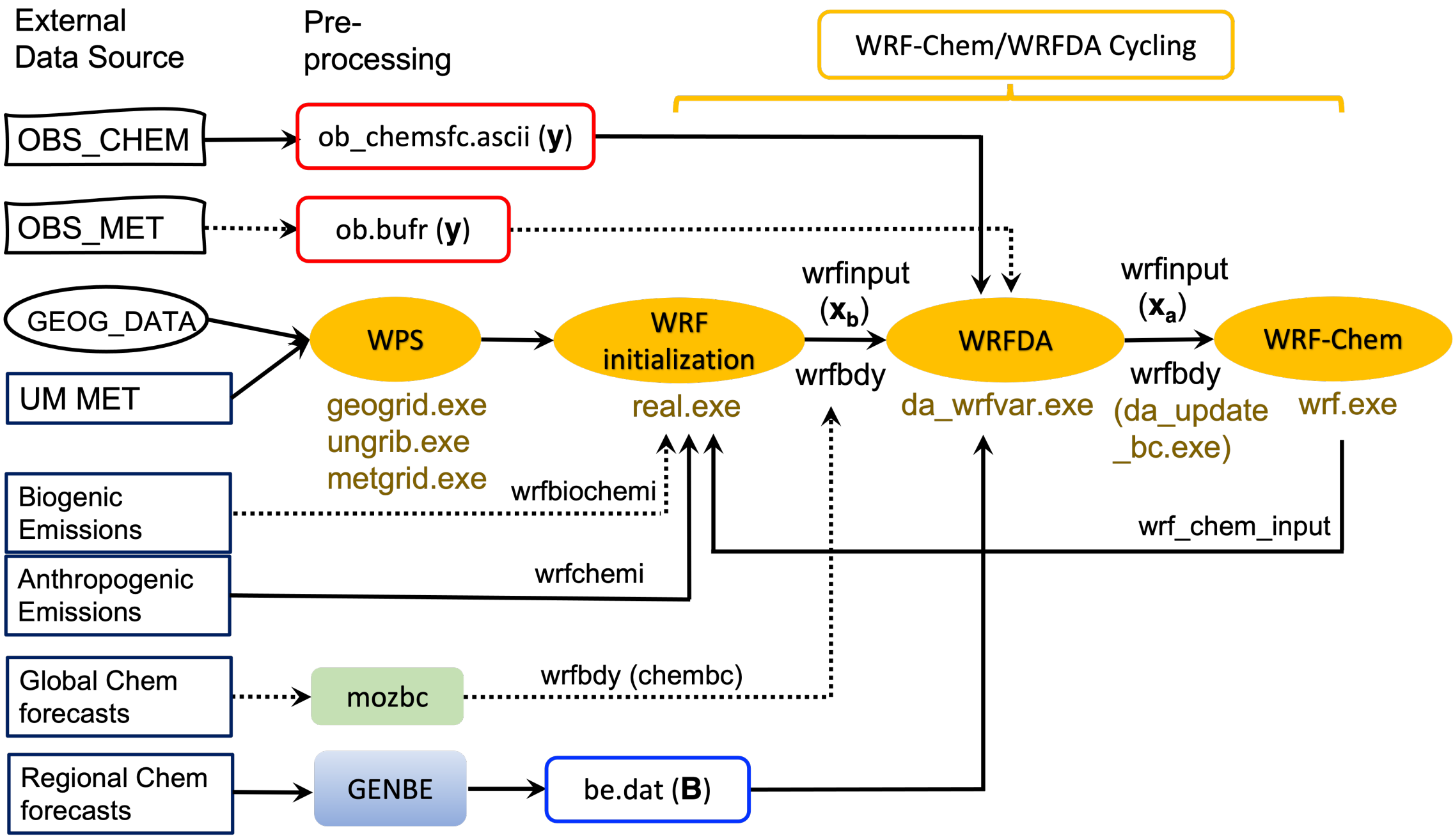

External Data Source
Pre-processing
WRF-Chem/WRFDA Cycling
OBS_CHEM
ob_chemsfc.ascii (y)
OBS_MET
ob.bufr (y)
GEOG_DATA
UM MET
WPS
geogrid.exe
ungrib.exe
metgrid.exe
WRF initialization
real.exe
wrfinput (xb)
wrfbdy
WRFDA
da_wrfvar.exe
wrfinput (xa)
wrfbdy
(da_update_bc.exe)
WRF-Chem
wrf.exe
Biogenic Emissions
wrfbiochemi
Anthropogenic Emissions
wrfchemi
wrf_chem_input
Global Chem forecasts
mozbc
wrfbdy (chembc)
Regional Chem forecasts
GENBE
be.dat (B)

Figure 2.

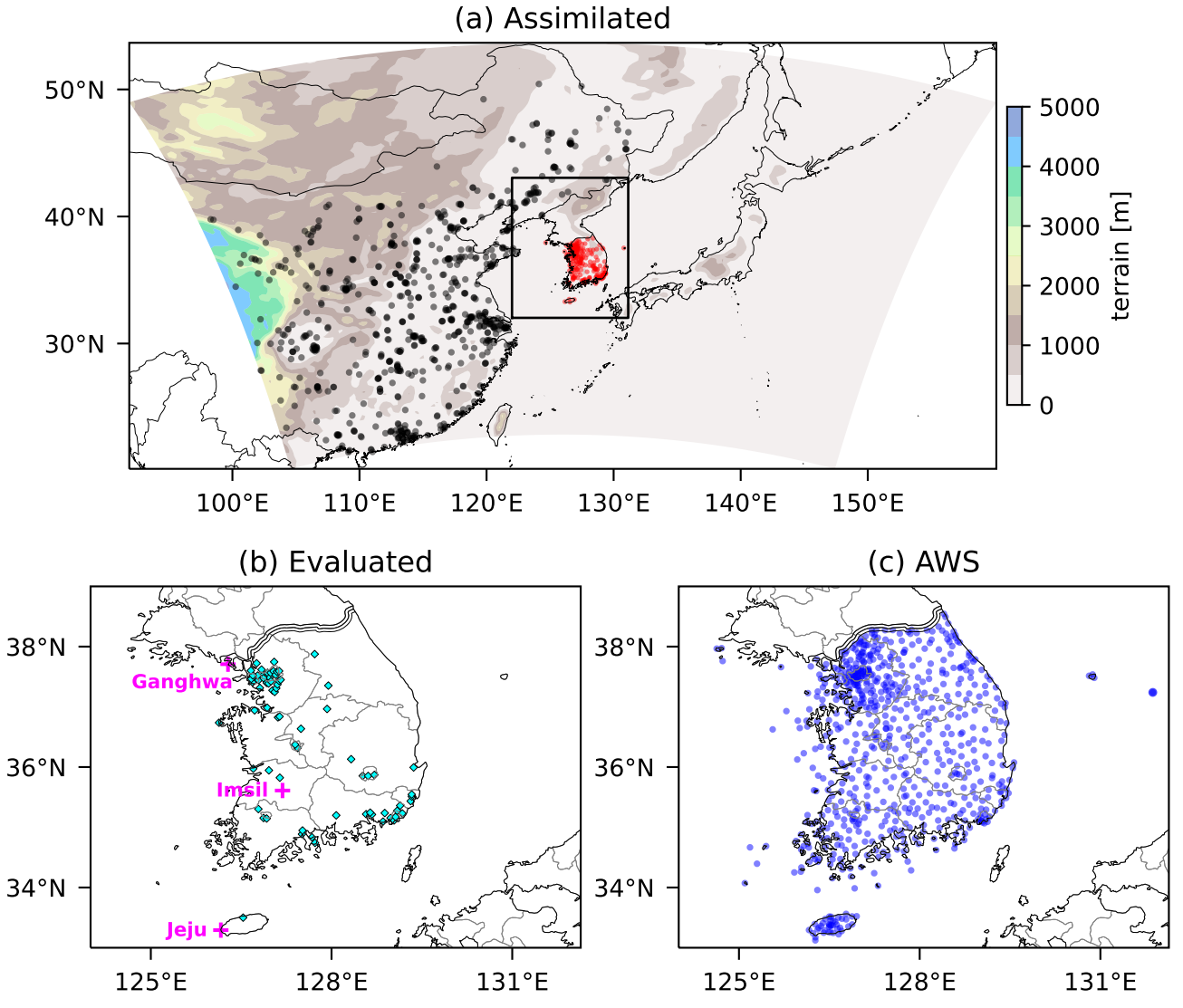

(a) Assimilated
50°N
40°N
30°N
100°E
110°E
120°E
130°E
140°E
150°E
5000
4000
3000
2000
1000
0
terrain [m]
(b) Evaluated
38°N
36°N
34°N
Ganghwa
Imsil
Jeju
125°E
128°E
131°E
(c) AWS
38°N
36°N
34°N
125°E
128°E
131°E

Figure 3.

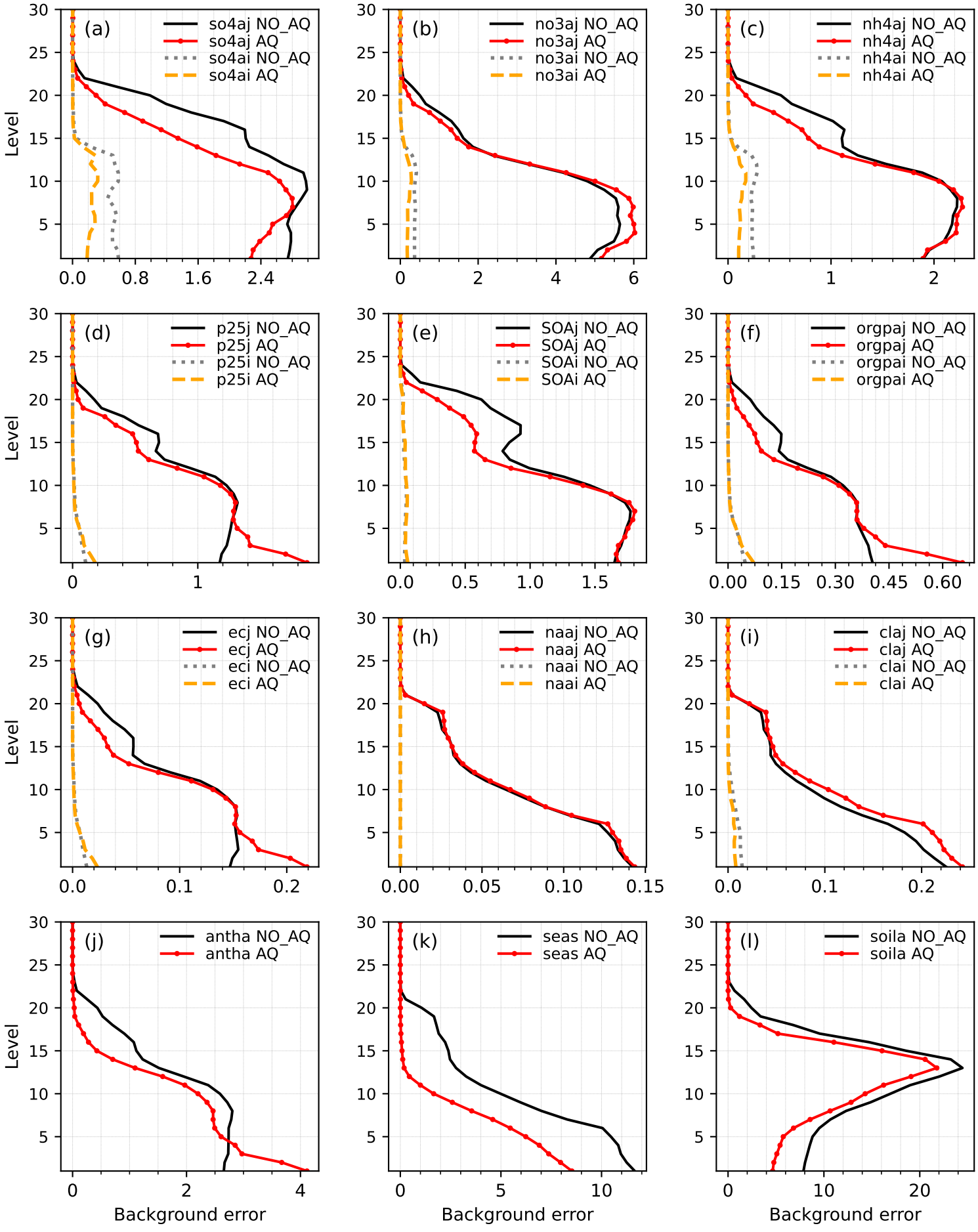

(a)
so4aj NO_AQ
so4aj AQ
so4ai NO_AQ
so4ai AQ
(b)
no3aj NO_AQ
no3aj AQ
no3ai NO_AQ
no3ai AQ
(c)
nh4aj NO_AQ
nh4aj AQ
nh4ai NO_AQ
nh4ai AQ
(d)
p25j NO_AQ
p25j AQ
p25i NO_AQ
p25i AQ
(e)
SOAj NO_AQ
SOAj AQ
SOAi NO_AQ
SOAi AQ
(f)
orgpaj NO_AQ
orgpaj AQ
orgpai NO_AQ
orgpai AQ
(g)
ecj NO_AQ
ecj AQ
eci NO_AQ
eci AQ
(h)
naaj NO_AQ
naaj AQ
naai NO_AQ
naai AQ
(i)
claj NO_AQ
claj AQ
clai NO_AQ
clai AQ
(j)
antha NO_AQ
antha AQ
(k)
seas NO_AQ
seas AQ
(l)
soila NO_AQ
soila AQ
Level
Background error

Figure 4.

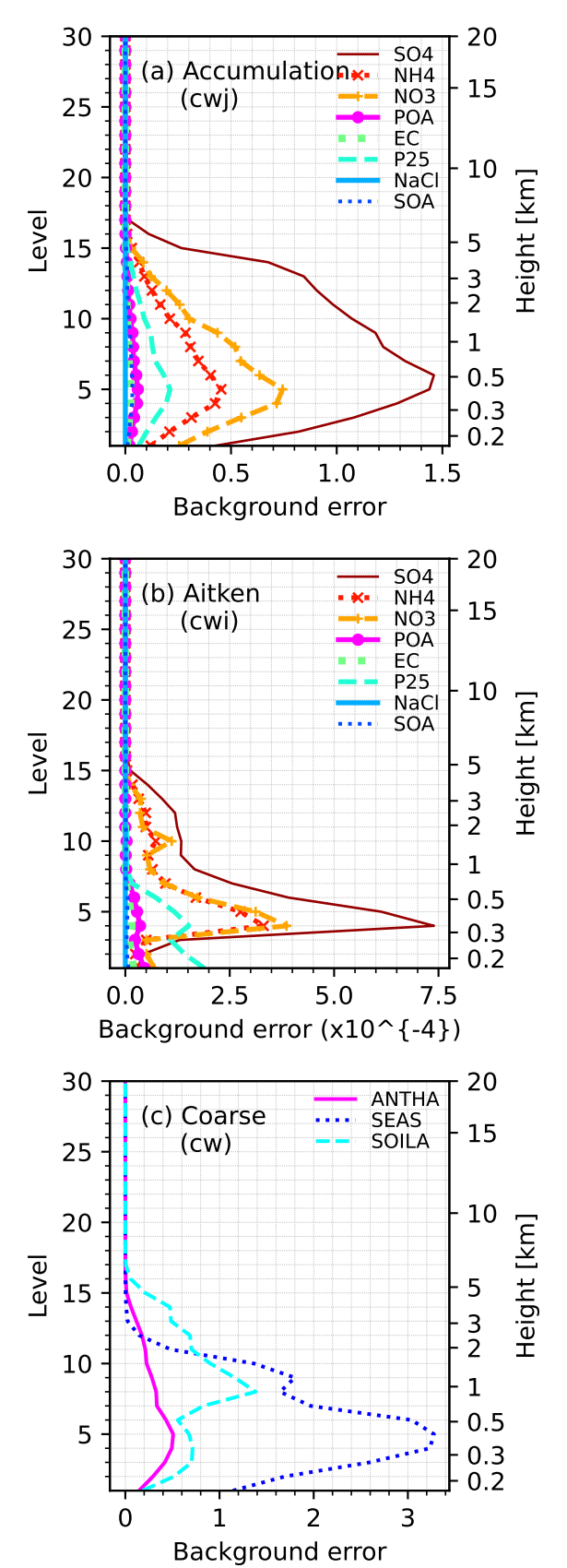

(a) Accumulation
(cwj)
SO4
NH4
NO3
POA
EC
P25
NaCl
SOA
Level
Height [km]
Background error
(b) Aitken
(cwi)
Background error (x10^{-4})
(c) Coarse
(cw)
ANTHA
SEAS
SOILA

Figure 5.

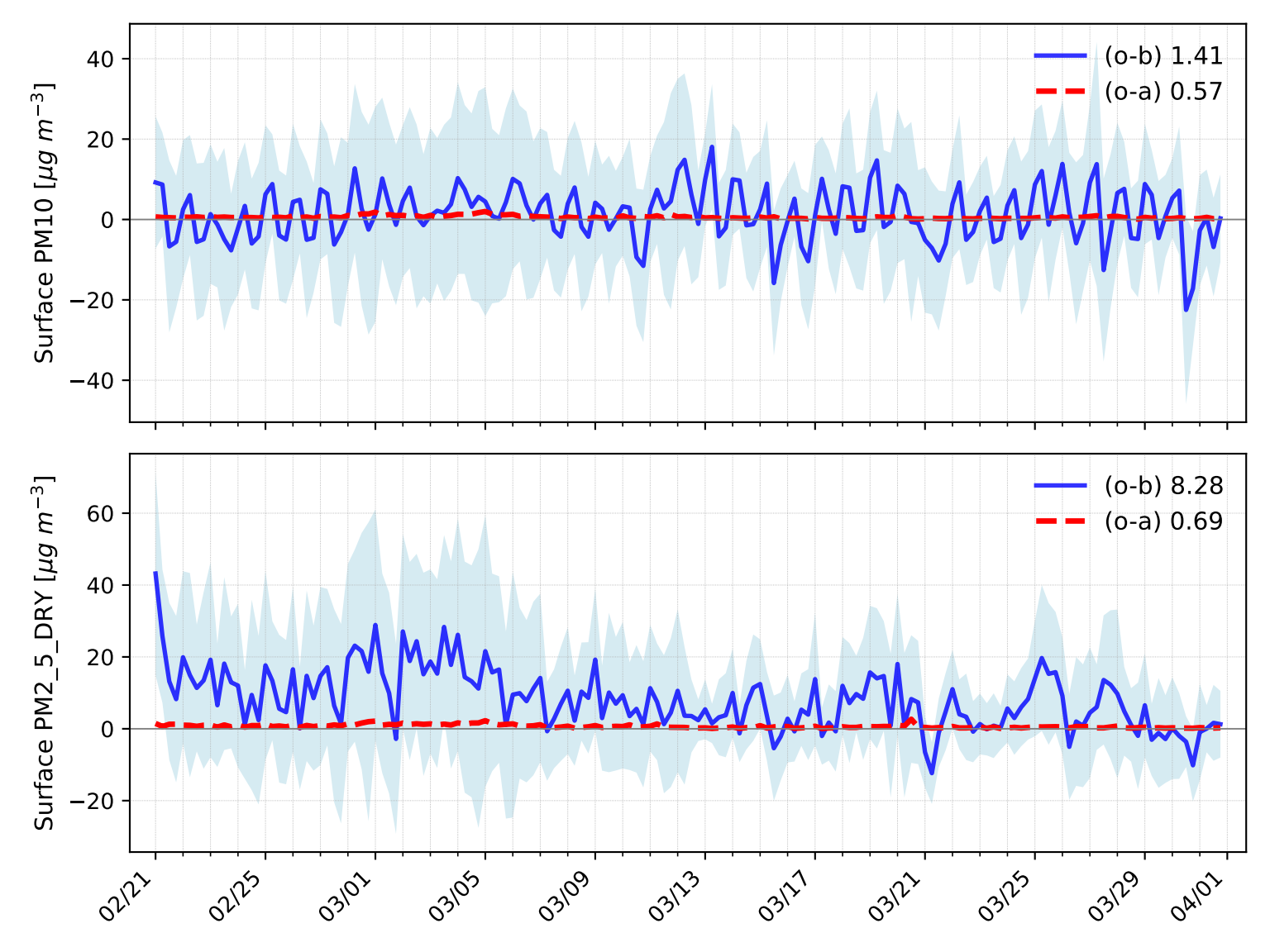

(o-b) 1.41
(o-a) 0.57
Surface PM10 [$\mu g\ m^{-3}$]
40
20
0
−20
−40
(o-b) 8.28
(o-a) 0.69
Surface PM2_5_DRY [$\mu g\ m^{-3}$]
60
40
20
0
−20
02/21
02/25
03/01
03/05
03/09
03/13
03/17
03/21
03/25
03/29
04/01

Figure 6.

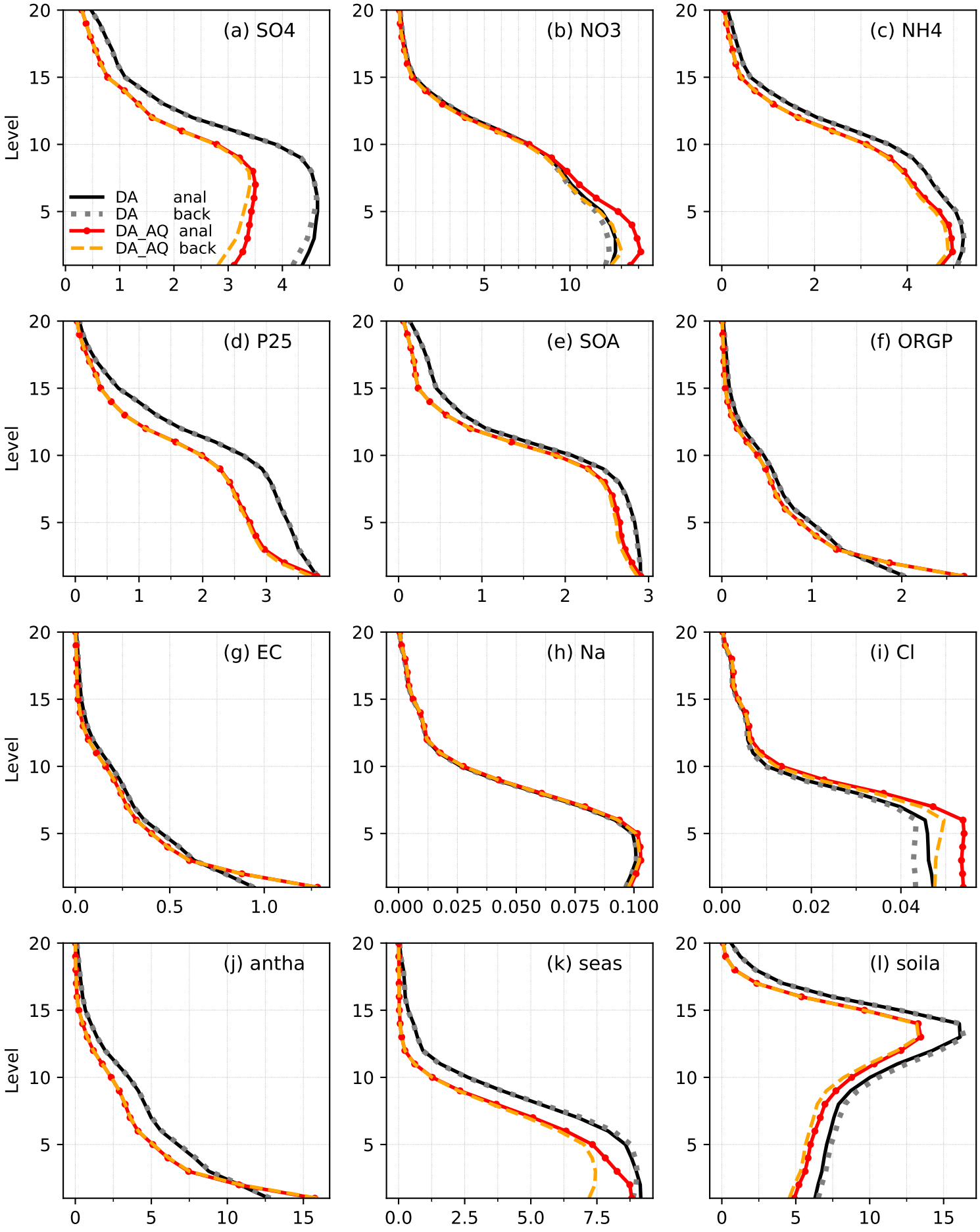

(a) SO4
(b) NO3
(c) NH4
(d) P25
(e) SOA
(f) ORGP
(g) EC
(h) Na
(i) Cl
(j) antha
(k) seas
(l) soila
Level
DA anal
DA back
DA_AQ anal
DA_AQ back

Figure 7.

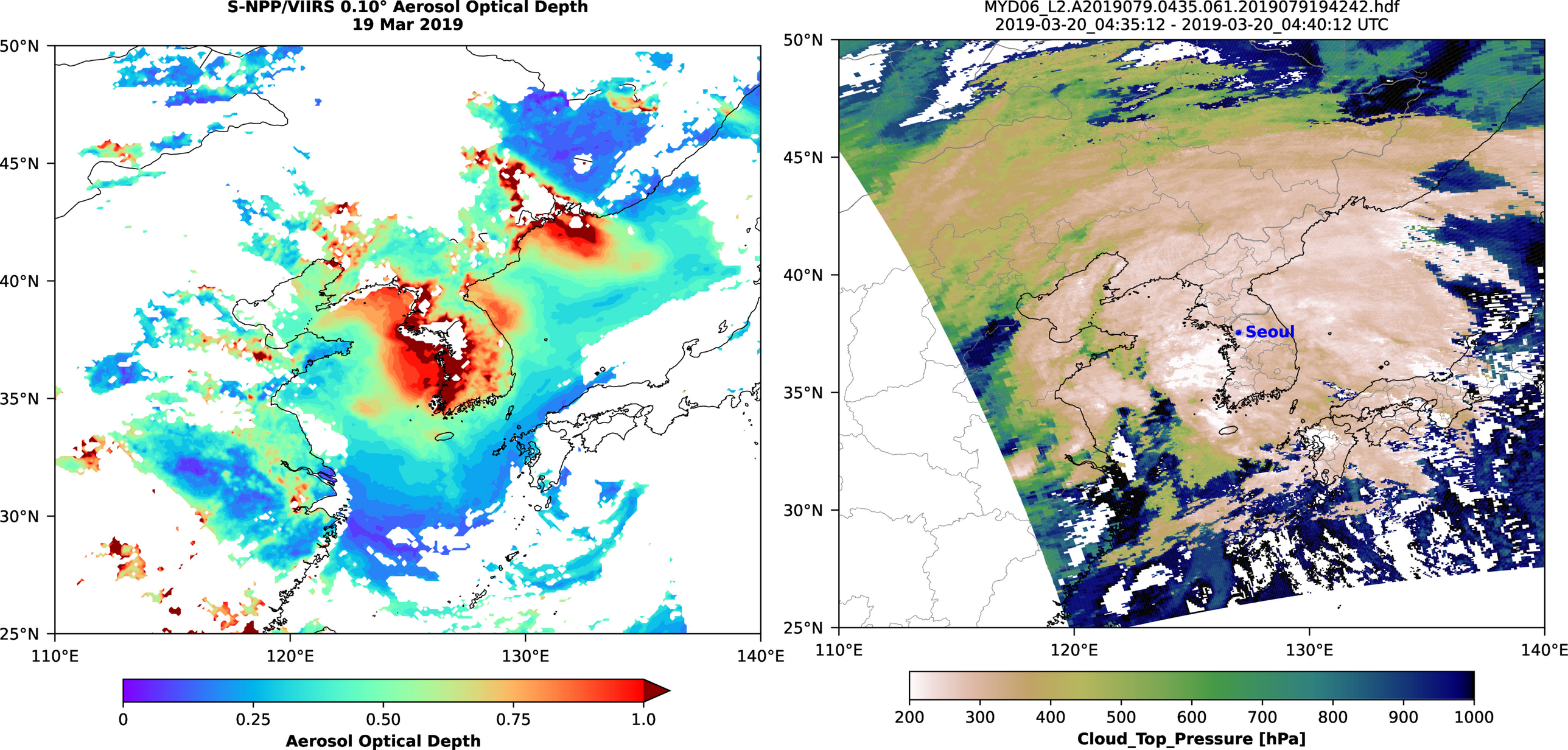

S-NPP/VIIRS 0.10° Aerosol Optical Depth
19 Mar 2019
50°N
45°N
40°N
35°N
30°N
25°N
110°E
120°E
130°E
140°E
0
0.25
0.50
0.75
1.0
Aerosol Optical Depth
MYD06_L2.A2019079.0435.061.2019079194242.hdf
2019-03-20_04:35:12 - 2019-03-20_04:40:12 UTC
Seoul
200
300
400
500
600
700
800
900
1000
Cloud_Top_Pressure [hPa]

Figure 8.

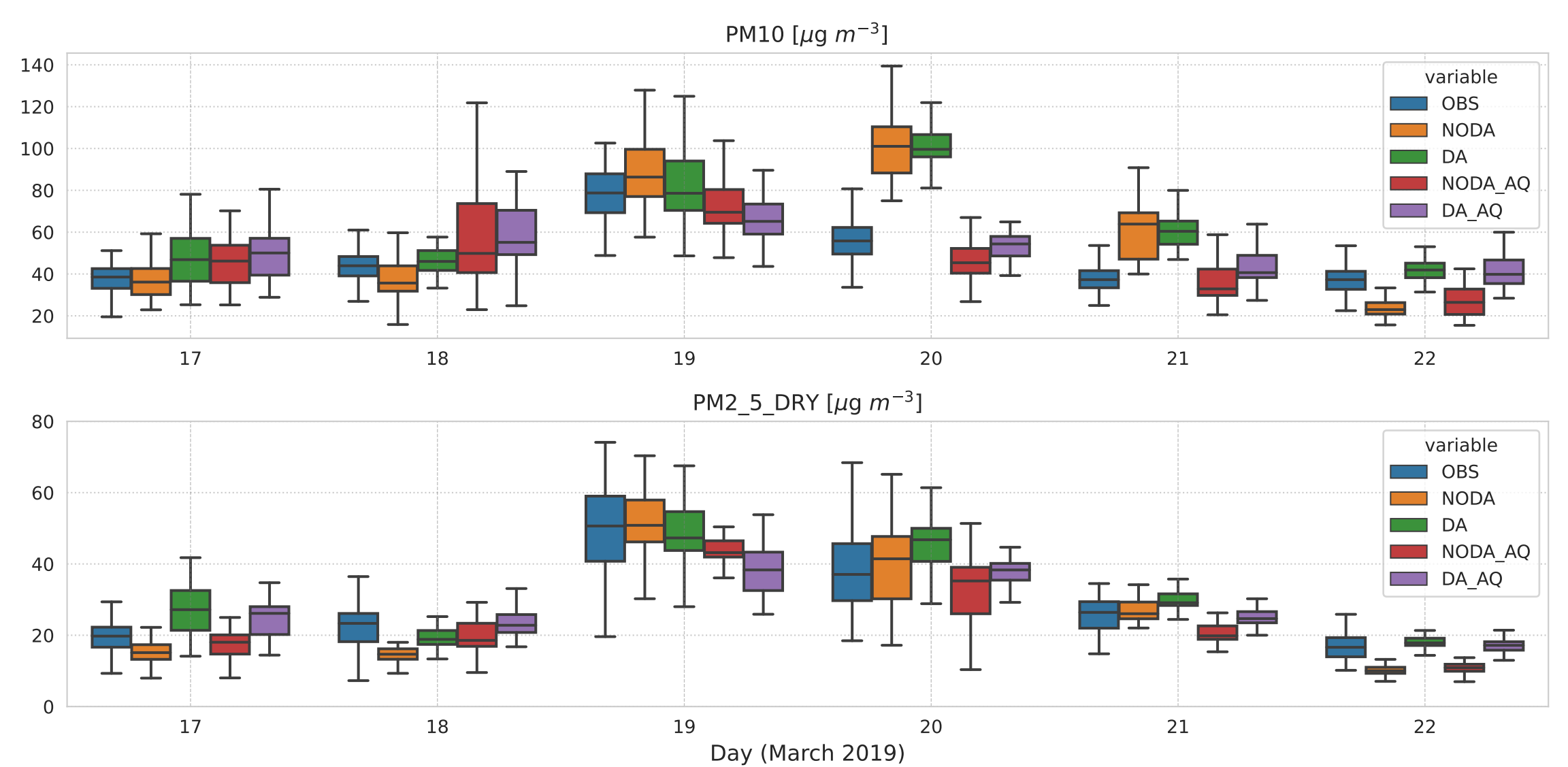

PM10 [$\mu g\ m^{-3}$]
140
120
100
80
60
40
20
17
18
19
20
21
22
variable
OBS
NODA
DA
NODA_AQ
DA_AQ
PM2_5_DRY [$\mu g\ m^{-3}$]
80
60
40
20
0
17
18
19
20
21
22
Day (March 2019)
variable
OBS
NODA
DA
NODA_AQ
DA_AQ

Figure 9.

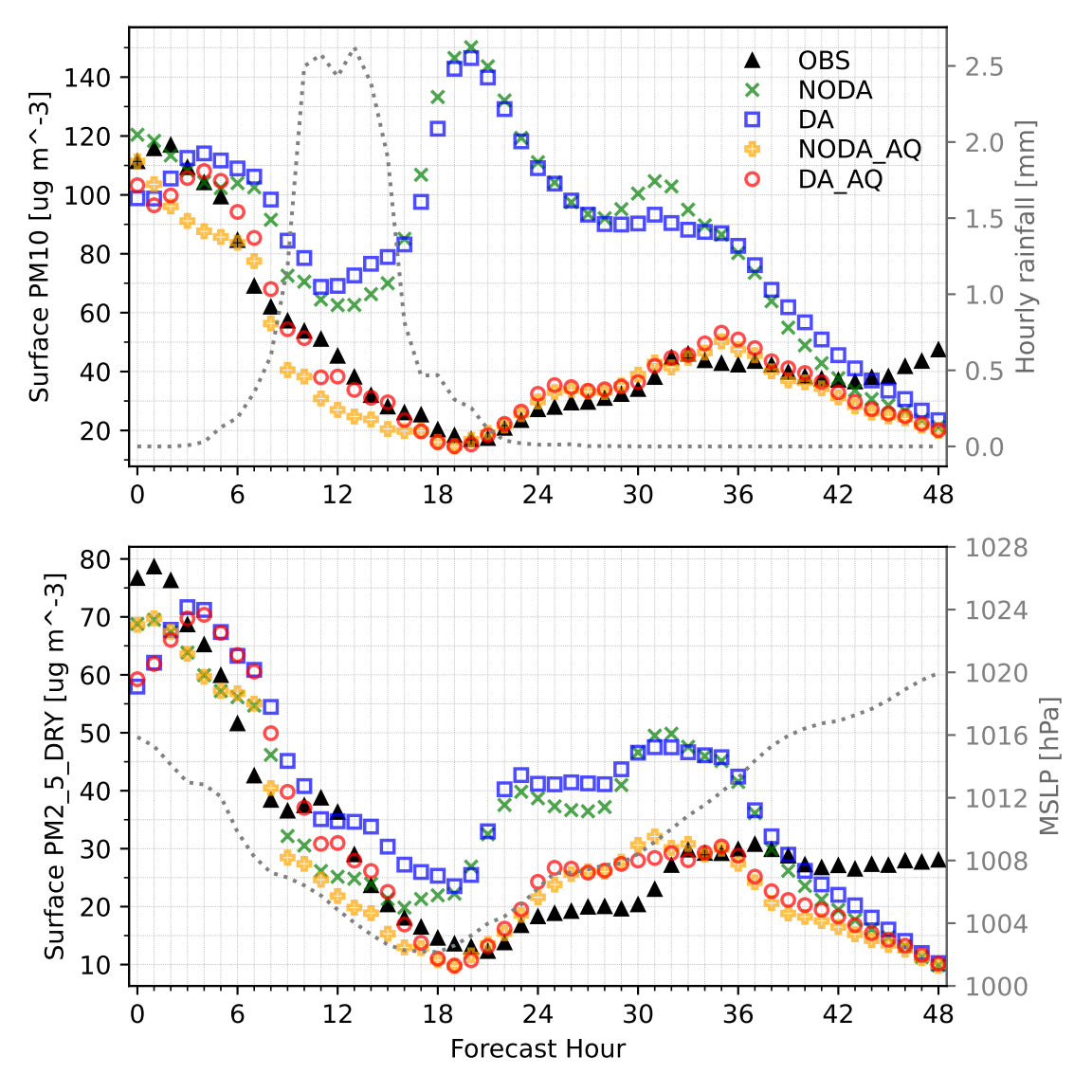

OBS
NODA
DA
NODA_AQ
DA_AQ
Surface PM10 [ug m^-3]
Hourly rainfall [mm]
Surface PM2_5_DRY [ug m^-3]
MSLP [hPa]
Forecast Hour

Figure 10.

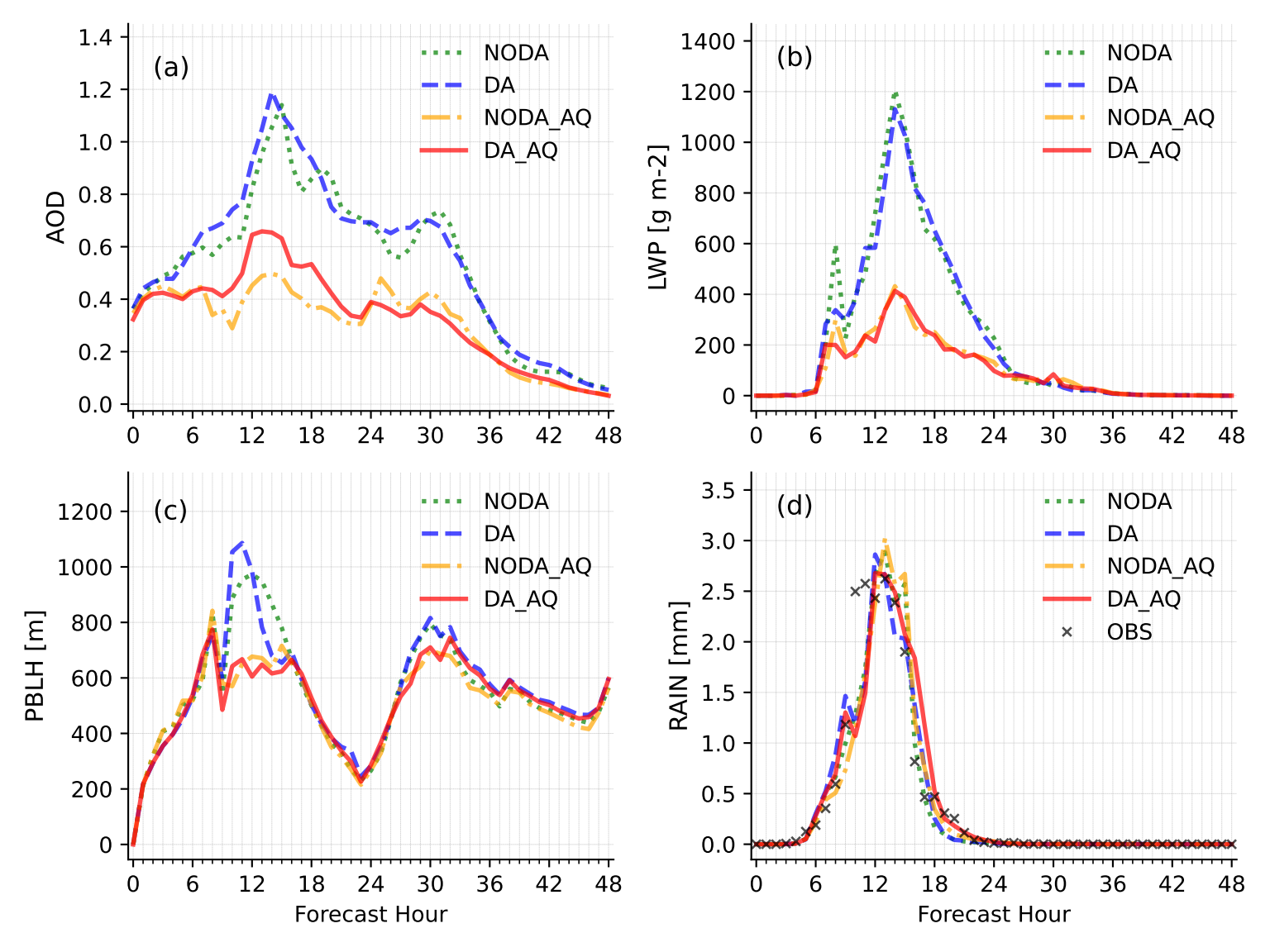

(a)
AOD
NODA
DA
NODA_AQ
DA_AQ
(b)
LWP [g m-2]
(c)
PBLH [m]
Forecast Hour
(d)
RAIN [mm]
OBS

Figure 11.

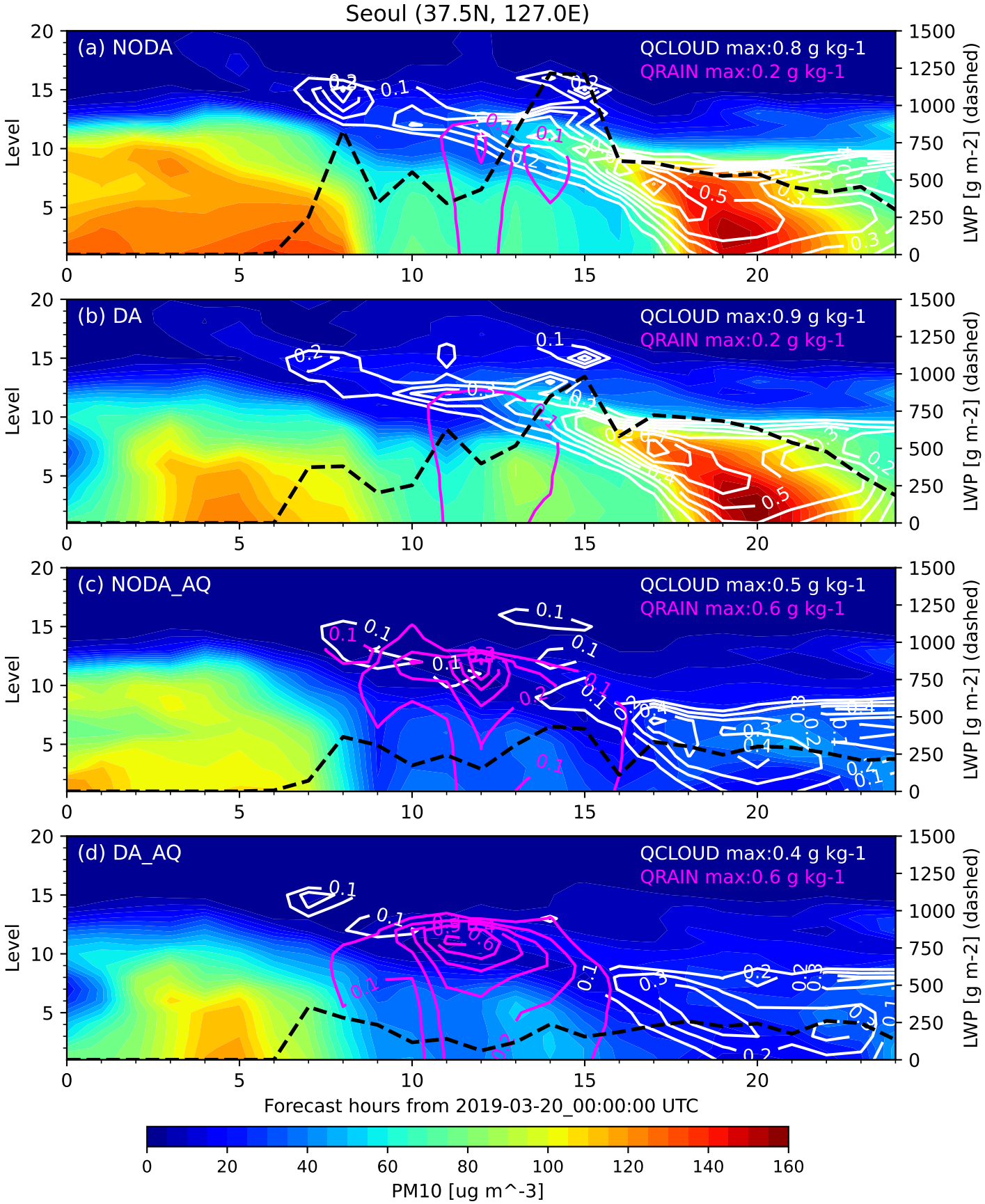

Seoul (37.5N, 127.0E)
(a) NODA
QCLOUD max:0.8 g kg-1
QRAIN max:0.2 g kg-1
(b) DA
QCLOUD max:0.9 g kg-1
QRAIN max:0.2 g kg-1
(c) NODA_AQ
QCLOUD max:0.5 g kg-1
QRAIN max:0.6 g kg-1
(d) DA_AQ
QCLOUD max:0.4 g kg-1
QRAIN max:0.6 g kg-1
Level
LWP [g m-2] (dashed)
Forecast hours from 2019-03-20_00:00:00 UTC
PM10 [ug m^-3]

Figure 12.

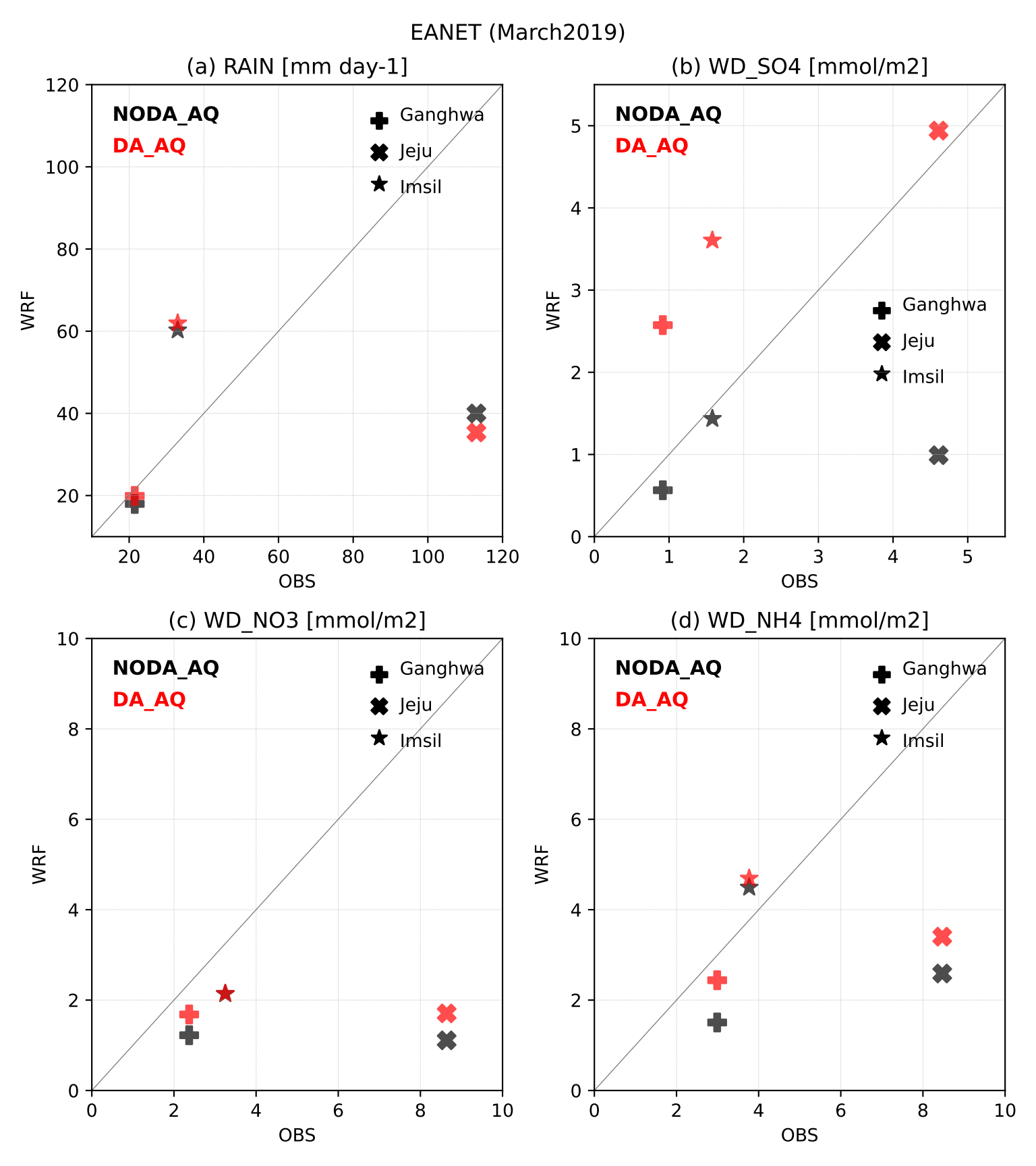
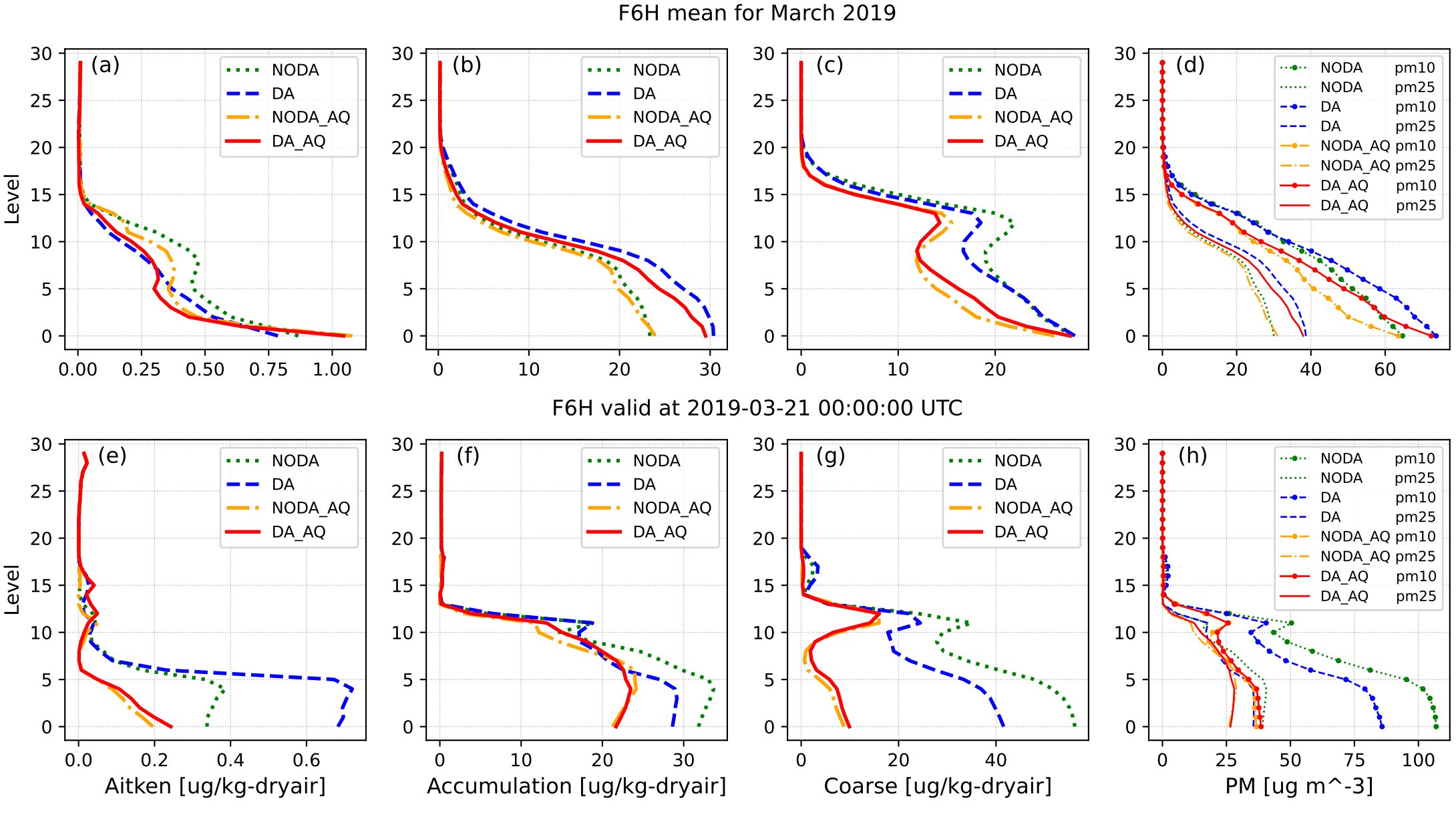

EANET (March2019)
(a) RAIN [mm day-1]
(b) WD_SO4 [mmol/m2]
(c) WD_NO3 [mmol/m2]
(d) WD_NH4 [mmol/m2]
NODA_AQ
DA_AQ
Ganghwa
Jeju
Imsil
WRF
OBS

Figure 13.


F6H mean for March 2019
F6H valid at 2019-03-21 00:00:00 UTC
Level
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
NODA
DA
NODA_AQ
DA_AQ
pm10
pm25
Aitken [ug/kg-dryair]
Accumulation [ug/kg-dryair]
Coarse [ug/kg-dryair]
PM [ug m^-3]