remote sensing

MDPI

Article

Seasonal Dependence of Aerosol Data Assimilation and Forecasting Using Satellite and Ground-Based Observations

Seunghee Lee [1,†], Ganghan Kim [1,†], Myong-In Lee [1,*], Yonghan Choi [2], Chang-Keun Song [1] and Hyeon-Kook Kim [1]

[1] Department of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, Korea; seungheelee@unist.ac.kr (S.L.); ghkim2000@unist.ac.kr (G.K.); cksong@unist.ac.kr (C.-K.S.); hyeonkook@unist.ac.kr (H.-K.K.)
[2] Korea Polar Research Institute, Incheon 21990, Korea; yhdchoi@kopri.re.kr
* Correspondence: milee@unist.ac.kr; Tel.: +82-52-217-2813
† These authors contributed equally to this work.

Abstract: This study examines the performance of a data assimilation and forecasting system that simultaneously assimilates satellite aerosol optical depth (AOD) and ground-based PM_{10} and $PM_{2.5}$ observations into the Weather Research and Forecasting model coupled with Chemistry (WRF-Chem). The data assimilation case for the surface PM_{10} and $PM_{2.5}$ concentrations exhibits a higher consistency with the observed data by showing more correlation coefficients than the no-assimilation case. The data assimilation also shows beneficial impacts on the PM_{10} and $PM_{2.5}$ forecasts for South Korea for up to 24 h from the updated initial condition. This study also finds deficiencies in data assimilation and forecasts, as the model shows a pronounced seasonal dependence of forecasting accuracy, on which the seasonal changes in regional atmospheric circulation patterns have a significant impact. In spring, the forecast accuracy decreases due to large uncertainties in natural dust transport from the continent by north-westerlies, while the model performs reasonably well in terms of anthropogenic emission and transport in winter. When the south-westerlies prevail in summer, the forecast accuracy increases with the overall reduction in ambient concentration. The forecasts also show significant accuracy degradation as the lead time increases because of systematic model biases. A simple statistical correction that adjusts the mean and variance of the forecast outputs to resemble those in the observed distribution can maintain the forecast skill at a practically useful level for lead times of more than a day. For a categorical forecast, the skill score of the data assimilation run increased by up to 37% compared to that of the case with no assimilation, and the skill score was further improved by 10% through bias correction.

Keywords: data assimilation; air quality modeling; aerosol forecast; seasonal dependence

Citation: Lee, S.; Kim, G.; Lee, M.-I.; Choi, Y.; Song, C.-K.; Kim, H.-K. Seasonal Dependence of Aerosol Data Assimilation and Forecasting Using Satellite and Ground-Based Observations. *Remote Sens.* **2022**, *14*, 2123. https://doi.org/10.3390/rs14092123

Academic Editor: Oleg Dubovik

Received: 18 March 2022
Accepted: 27 April 2022
Published: 28 April 2022

1. Introduction

East Asia is a region with high aerosol concentrations due to emissions from various natural sources, such as frequent Asian dust events from the Taklamakan and Gobi Deserts [1,2] and Siberian wildfires [3,4]. Moreover, the region has experienced rapid economic growth, which has increased anthropogenic emissions from agricultural, industrial, commercial, and residential sources. This has resulted in a rapid change in environmental conditions and air pollution [5–7]. Despite the regulation policies introduced to reduce air pollutants, particularly in China [8], air pollution is still an important international agenda that needs to be resolved [9].

While it would take considerable time to mitigate anthropogenic emissions and improve air quality, taking appropriate actions and decisions in advance according to air quality forecasts could help reduce health risks and negative impacts on the socioeconomy. There is a growing demand for accurate and reliable predictions of air quality that can be disseminated to the general public. Many operational centers and research institutes have

developed comprehensive atmospheric chemical transport models to produce air quality forecasts. However, the forecast skill of these models is still low and not satisfactory for providing early warnings [6,10,11], requiring substantial advances in systems by improving atmospheric chemical transport models, updating emission inventories, and utilizing more observational data from ground-based and satellite observations [11–15]. In particular, data assimilation improves the analysis quality and initial conditions set for the forecast models as it allows the incorporation of more observations in Northeast Asia from satellites such as the Moderate Resolution Imaging Spectroradiometer (MODIS) [16], Geostationary Ocean Color Imager (GOCI) [17] and Geostationary Environment Monitoring Spectrometer (GEMS) [18]. The appropriate application of data assimilation techniques can provide benefits when using multiple observational data sources [19]. Data assimilation overcomes the spatial and temporal limitations of current observation networks by considering both uncertainties in the observational data and background model forecasts [13,15,20–29]. Kim et al. [28] used the three-dimensional variational (3D-VAR) method for assimilating GOCI AOD and ground-based PM data during the Korea–US Air Quality (KORUS-AQ) study period (1 May to 10 June 2016). They indicated the effects of data assimilation in improving air quality forecast in Korea and examined the forecast skills sensitive to single- and multiple-observation input data. Their results highlighted that using multiple observations can overcome the spatial and temporal observation vacancy and provide sustainable air quality forecasts. Xia et al. [21] also used 3D-VAR for the Chinese Pungyun-3A satellite and MODIS AOD during the Asian Dust event in the spring of 2011. They indicated that the effects of data assimilation in improving air quality forecasts in China were the highest when AODs from both satellites were used simultaneously. Jung et al. [22] tested Optimal Interpolation (OI) in assimilating the GOCI AOD to the two-way coupled Weather Research and Forecasting (WRF) model for meteorology and the Community Multiscale Air Quality (CMAQ) chemistry and transport model that enables the interactive radiative effects by aerosols on weather and air quality. In this test in the East Asia domain during May and June 2016, the shortwave radiation on land tended to decrease because of the direct radiative effects of aerosols, hence reducing the surface air temperature, atmospheric boundary layer height, and wind speed. These modifications contributed to an increase in the surface concentrations of major gaseous and particulate pollutants. Not only variational methods but also ensemble-based methods can be used to improve the initial conditions and forecasts [13,26,27]. For example, Peng et al. [13] optimized the initial chemical conditions and emission data using the ensemble Kalman filter (EnKF), thereby improving $PM_{2.5}$ forecasts. Choi et al. [26] examined the dependence of model performance on data assimilation methods when assimilating the MODIS AOD by comparing the performances of the variational and ensemble-based methods with a hybrid scheme that combined both approaches. Additionally, recently, there was a study that improved the performance of satellite data assimilation by combining data assimilation and machine learning [29].

Although previous studies have demonstrated the beneficial impacts of data assimilation when using ground-based and satellite aerosol observations, most studies had a relatively short testing period. Considering that the data assimilation and forecast performance depend on the region and testing period, the performance evaluation should be based on a rigorous test with a sufficiently long testing period for a specific system. Particularly in East Asia, where summer and winter monsoons produce distinct variations in regional atmospheric circulation and aerosol transport patterns, the forecast performance is likely to vary significantly depending on the season. Additionally, natural and anthropogenic emissions exhibit substantial seasonality. For example, regional air quality is affected by Asian Dust episodes in spring [1,2], whereas coal-fired heating increases anthropogenic emissions in winter [30].

To develop a reliable air quality forecast model for operational purposes, this study first constructs a data assimilation and forecast system using the WRF model coupled with chemistry (WRF-Chem) and 3D-VAR data assimilation method. Satellite-derived and ground-based observations were assimilated, including AOD observations from MODIS

and GOCI, and surface PM_{10} and $PM_{2.5}$ concentrations. The performance of the system was evaluated for one representative month for each season to determine the accuracy of the data analysis and forecast for surface PM_{10} and $PM_{2.5}$. Specifically, a comparison of forecasts with and without data assimilation measured the extent to which data assimilation quantitatively improves forecasting. The forecasts tended to degrade with an increase in the lead time owing to systematic biases in the forecasting model. Therefore, this study further conducted simple bias correction on the air quality forecast outputs and compared the results with the original forecast. This study consists of the following sections: Section 2 describes the experimental settings, models, and data used for the data assimilation and forecast validation. Section 3 presents the performance of data assimilation and forecast accuracy. Section 4 provides a discussion and Section 5 presents conclusions.

2. Model, Data, and Methods

2.1. Model

The air quality model used in this study was WRF-Chem version 3.9.1 [31], developed by the National Center for Atmospheric Research (NCAR), an online model with coupled interactions between atmospheric chemistry and meteorology. The model considers the direct and indirect effects of atmospheric chemicals and aerosols on atmospheric radiative transfer and cloud microphysical processes. The model uses Goddard Chemistry Aerosol Radiation and Transport (GOCART) [32] as an aerosol scheme, including 15 types of dust, organic carbon (OC), black carbon (BC), sulfate, and sea salt. Atmospheric chemistry was formulated using the Model for Ozone and Related Chemical Tracers (MOZART) [33]. Parameterized physics included the Lin scheme [34] for microphysics, the Rapid Radiative Transfer Model (RRTM) scheme [35] for longwave radiation, the Goddard scheme [36] for shortwave radiation, the Grell-3D scheme [37] for convection, the Yonsei University (YSU) scheme [38] for atmospheric boundary layer turbulence, and Noah Land-Surface Model [39] for land surface physics.

Figure 1 shows the model domain in East Asia (D01) at a horizontal resolution of 27 km with a nested sub-domain (D02) with a resolution of 9 km above the Korean Peninsula. The model was configured with 30 vertical layers, with a topmost layer at 50 hPa. The initial and lateral boundary conditions for meteorology were specified using the Met Office Unified Model (UM) [40] global forecasts produced by the the Korea Meteorological Administration. The initial and lateral boundary conditions for chemicals were specified using MOZART-4 [41] global chemistry reanalysis data produced by the Goddard Earth Observation System 5 (GEOS-5) model coupled with MOZART. The model uses the Emissions Database for Global Atmospheric Research–Hemispheric Transport of Air Pollution (EDGAR-HTAP) [42] for the anthropogenic emissions of gases and aerosols. The model also includes biological emissions from the Model of Emissions of Gases and Aerosols from Nature (MEGAN) [43].

2.2. Data Assimilation

The data assimilation method was the 3D-VAR algorithm implemented by the Gridpoint Statistical Interpolation (GSI) version 3.5 [44,45]. Details of applying the 3D-VAR to the GOCART aerosol scheme have previously been described by Liu et al. [20], and the essential parts are briefly described here.

The 3D-VAR algorithm finds the best estimate "analysis" by considering the Euclidian distance from observations at irregularly spaced points and a gridded background field. The analysis vector (x) can be determined by minimizing the scalar cost function J(x), given by

$$\mathrm{J(x)} = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}[H(x) - y]^T R^{-1}[H(x) - y] \quad (1)$$

In Equation (1), x_b denotes the model background state, y is the vector of observations, and B and R are the background and observation error covariance matrices, respectively, representing the relative weights for the analysis. H is an observation operator that converts

the model values to observations and interpolates the model values to observation positions in space. The AOD observation operator was calculated based on the Community Radiative Transfer Model (CRTM) [46], and the observation operators of PM_{10} and $PM_{2.5}$ have been described by Pagowski et al. [47].

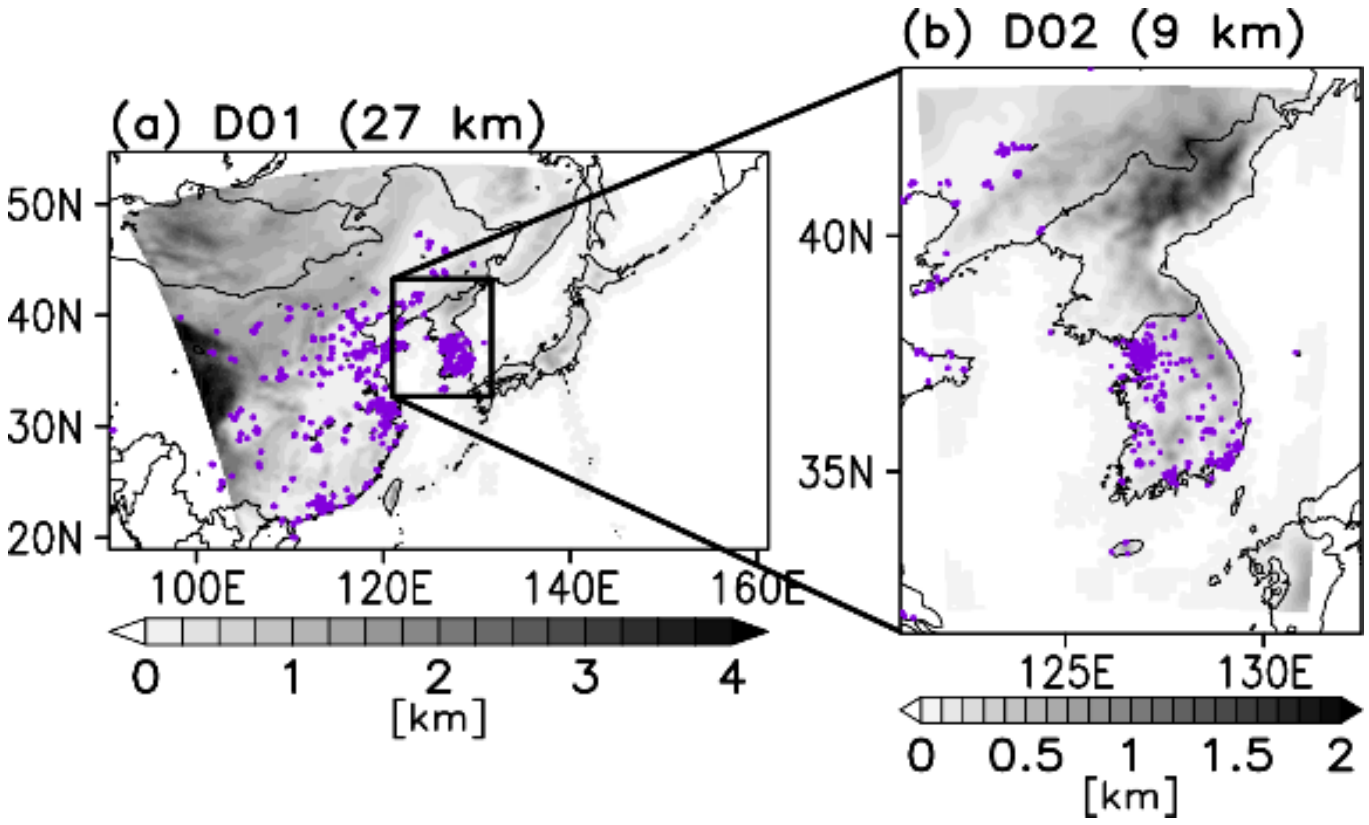

Figure 1. Simulation domain for WRF-Chem with surface elevation (shaded, unit: kilometer). (**a**) shows domain 1 (D01) and (**b**) shows domain 2 (D02). Purple dots indicate the locations of ground-based PM observation.

Proper specification of the model background error, on which the performance of data assimilation depends, is one of the most critical parts of the data assimilation technique. The National Meteorological Center (NMC) method [48] was used to obtain the background error covariance B, which extracts the systematic error by calculating the difference between two model background forecasts starting at different initial times. In this study, the background error was calculated using the difference between the 24 and 12 h forecasts. The forecast differences were collected at 12 h intervals to calculate the background error covariance statistics for each season in January, April, July, and October 2017.

2.3. Observation Data and Errors

AODs from both satellites, MODIS and GOCI, and the ground-based PM observations from the China Monitoring Network (CMN) and the National Ambient Monitoring Information System (NAMIS) in Korea were integrated as inputs for 3D-VAR data assimilation. Detailed information on the input data is presented in Table 1. The table also shows the observation errors specified for the MODIS AOD as a function of the AOD value (τ), which follows Levy et al. [49] for land (ε_L) and Remer et al. [50] for ocean (ε_O) errors. The errors for GOCI follow the error equations specified by Choi et al. [51].

Table 1. Input observation data and the observation error equations. ε_O and ε_L are the errors over ocean and land, respectively, and τ is the AOD. ε_m is the measurement error for ground-based PM observation, ε_r is the representative error, ε_t is the total error, Π_o is the PM concentration, γ = 0.5 is the adjustable parameter, Δx is the model grid size, and L = 3 km is the radius of influence of observation according to Elbern et al. [52].

Data	Period	Resolution (km)	Variables	Error Equation
MODIS (Lv 2)	2000–Present 5 min (Aqua & Terra)	10 × 10	550 nm AOD	$\varepsilon_O = 0.05\tau + 0.03$ $\varepsilon_L = 0.15\tau + 0.05$
GOCI	2011–Present 1 h (00–07 UTC)	6 × 6	550 nm AOD	$\varepsilon_O = 0.185\tau + 0.037$ $\varepsilon_L = 0.137\tau + 0.073$
NAMIS	2014–Present 1 h	Station data	PM_{10}, $PM_{2.5}$	$\varepsilon_m = 1.5 + 0.0075\,\Pi_o$ $\varepsilon_r = \gamma\,\varepsilon_m\sqrt{\frac{\Delta x}{L}}$ $\varepsilon_t = \sqrt{\varepsilon_m^2 + \varepsilon_r^2}$
CMN	2014–Present 1 h	Station data	PM_{10}, $PM_{2.5}$	

Ground-based PM observations were available from 1514 stations in China and 361 stations in South Korea, as indicated in Figure 1. The PM observation error consisted of measurement and representative errors. The measurement error was defined according to Schwartz et al. [53] and the representative error was obtained from Elbern et al. [52] and Pagowski et al. [47].

In addition, the quality of the observed data was controlled prior to data assimilation. For satellite AODs, only data with a confidence flag of 3 (the best quality) were used. Data thinning was also applied, where high-resolution satellite data were coarsely sampled to match the model resolution [16]. One AOD observation was randomly sampled for every 27 km × 27 km grid for Domain 1 and every 9 km × 9 km grid for Domain 2. In the case of PM_{10} ($PM_{2.5}$), unrealistic observation values were discarded when they were higher than 300 (150) μg m^{-3} or the innovation (H (x) − y) values were higher than 300 (150) μg m^{-3}.

2.4. Experiments

Two sets of experiments were performed in this study (Figure 2). The first run involved an unconstrained WRF-Chem simulation that provided 24 h model forecasts for a month with meteorological fields reinitialized by atmospheric reanalysis data at 00 UTC every day (hereafter NoDA, no data assimilation). In NoDA, the chemical and aerosol fields were restarted from previous forecasts. The second run comprised continuous data assimilation and forecast runs with a 6 h cycle (hereafter DA, data assimilation). The aerosol fields were assimilated via satellite and ground-based observations from MODIS, GOCI, CMN, and NAMIS. The meteorological fields were reinitialized at 00 UTC every day, as for NoDA, but 3D-VAR data assimilation was performed at 00, 06, 12, and 18 UTC for the aerosol fields to provide the initial conditions for the next 6 h. The chemistry fields were restarted from previous forecasts, as for the NoDA. For satellite data assimilation, the geostationary GOCI data in the visible radiation bands are only available hourly from 00 to 07 UTC. The sun-synchronous MODIS data are available twice a day at 03–04 and 17–18 UTC from Aqua, and 02–03 and 13–14 UTC from Terra. This study used a time window of ±3 h, and all available satellite observations during this period were collated for aerosol data assimilation for the given period. This helped to increase the number of satellite observation samples that are often masked by clouds. Time weighting was not applied to the observations sampled at different times within the window. The improvements in forecasting by 3D-VAR aerosol data assimilation were measured by first conducting 24 h model forecasts starting from the assimilated aerosol states and then comparing the forecasts with NoDA runs. Because of computational costs, only January, April, July, and October 2017 were selected for testing

seasonal changes in the forecasting performance, and for each month, the 24 h forecasts were conducted at 00, 06, 12, and 18 UTC every day.

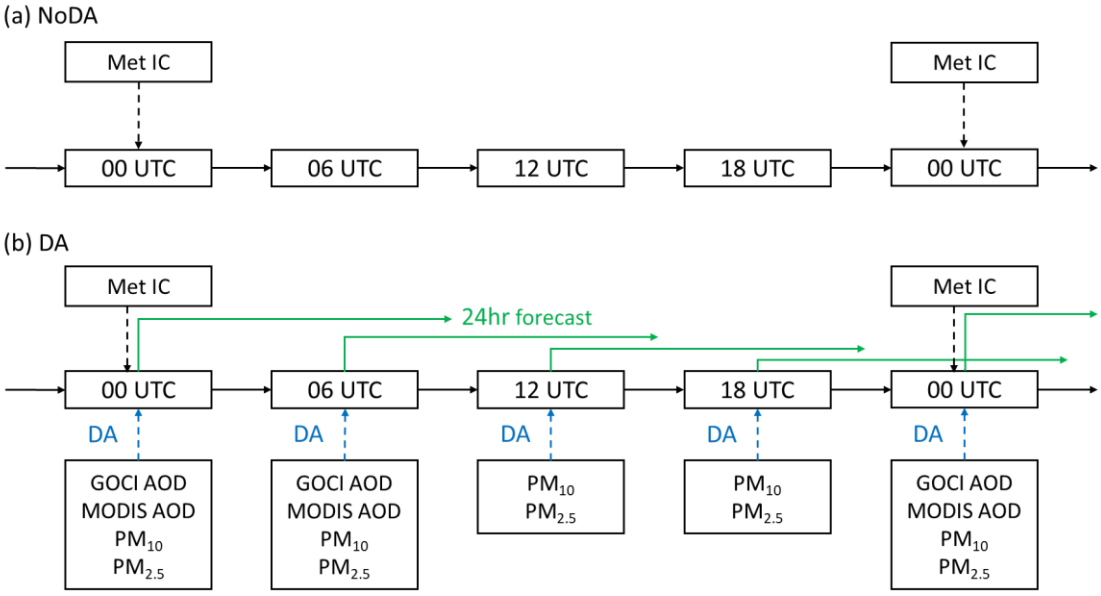

Figure 2. Schematic diagram of (**a**) NoDA and (**b**) DA experiments. The dashed line in black indicates meteorology initialization at 00 UTC every day, and the blue dashed line indicates the aerosol data assimilation with a 6 h cycle. The solid line represents the WRF-Chem model integration.

2.5. Verification

The accuracy of the air quality forecasts for surface PM concentrations was evaluated using the temporal correlation coefficient and root-mean-square error (RMSE) compared with the observed values. As discussed later in the text, the WRF-Chem forecast model exhibits a systematic bias as the forecast lead time increases. A simple statistical correction was applied to remove biases in the forecast values using the following equation:

$$X' = \left(X_f - \overline{X_f}\right)\frac{\sigma_o}{\sigma_f} + \overline{X_o} \quad (2)$$

where X' is the corrected model forecast, X_f is the original model forecast, $\overline{X_f}$ is the time average of X_f, and $\overline{X_o}$ is the average of observations. σ_f and σ_o are the standard deviations of the forecasts and observations, respectively. Equation (2) adjusts the mean and variance of the forecast values to those of the observed values, assuming a normal distribution. These forecast outputs are denoted by BiasC (i.e., bias-corrected) in Section 3. Note that this bias-correction method can also be applied to real-time forecasts if the long-term historical observations and the hindcasts are available.

In addition, a contingency table was constructed to record the hit and false-alarm rates of the categorical air quality forecasts. The categories for clean, normal, polluted, and extremely polluted PM_{10} ($PM_{2.5}$) concentrations were 0–30 (0–15), 31–80 (16–35), 81–150 (36–75), and 151–(76–) μg m^{-3}, respectively, as officially defined by the National Institute of Meteorology (NIER) in Korea. Table 2 presents a contingency table.

Table 2. Contingency table for the categorical forecasts.

		Observation			
		Clean	**Normal**	**Polluted**	**Extremely Polluted**
Forecast	Clean	A1	A2	A3	A4
	Normal	B1	B2	B3	B4
	Polluted	C1	C2	C3	C4
	Extremely Polluted	D1	D2	D3	D4

Based on Table 2, the total hit rate (THR), polluted case hit rate (PHR), and false-alarm rate (FAR) were calculated using the following equations:

$$THR = \frac{A1 + B2 + C3 + D4}{n} \times 100 \quad (3)$$

$$PHR = \frac{C3 + D4}{III + IV} \times 100 \quad (4)$$

$$FAR = \frac{II}{II + IV} \times 100 \quad (5)$$

where n indicates the sum of all values in the contingency table, and II, III, and IV are defined as

$$II = C1 + C2 + D1 + D2 \quad (6)$$

$$III = A3 + A4 + B3 + B4 \quad (7)$$

$$IV = C3 + C4 + D3 + D4 \quad (8)$$

This study also presented the categorical forecast skill using the Heidke Skill Score (HSS) [54], defined as:

$$HSS = \frac{\sum_{i=1}^{I} p(y_i, o_i) - \sum_{i=1}^{I} p(y_i)p(o_i)}{1 - \sum_{i=1}^{I} p(y_i)p(o_i)} \quad (9)$$

where i indicates the index of the category, I is the total number of forecast categories, $p(y_i, o_i)$ is the joint probability distribution of forecasts and observations, and $p(y_i)$ and $p(o_i)$ are the marginal distributions of forecasts and observations, respectively. $\sum_{i=1}^{I} p(y_i, o_i)$ represents the proportion of correct forecasts (i.e., the forecast accuracy of actual forecasts), and $\sum_{i=1}^{I} p(y_i)p(o_i)$ represents the correct proportion for the random reference forecast. Therefore, HSS indicates the extent to which the forecast has improved compared to a random reference forecast. HSS values range from negative infinity to 1. A value of 1 indicates a perfect forecast, 0 indicates that the forecast has the same performance as the random forecast, and the negative value suggests a worse performance than the random forecast. A HSS higher than zero indicates that the forecast is practically useful.

3. Results

3.1. Seasonal Dependence of Model Background Error

This study specifies the model background error covariance differently for each season to reflect the seasonal dependence of the forecast errors in the model. Examining the model background error for each aerosol species and season identifies which aerosol species are dominant and how they are transported depending on the season.

Figure 3 compares the vertical distributions of the standard deviation of the model background error, averaged for each month and species: hydrophobic BC (BC1), hydrophilic BC (BC2), hydrophobic OC (OC1), hydrophilic OC (OC2), sea salt (SEAS), dust (DUST), sulfate (SULF) and unspeciated $PM_{2.5}$ (P25). The errors were generally larger in the lower atmosphere. As most aerosols originate from the surface, a higher concentration contributes

more to model errors and forecast uncertainty. Note that the background error depends significantly on the season, owing to differences in the dominant emission types and atmospheric circulation. On average, the errors tended to increase in winter for OC, BC, and P25, owing to increased fossil fuel combustion and biomass burning in this season. A stronger surface wind in winter due to the enhanced East Asian jet stream also contributes to active aerosol transport and therefore induces more forecast errors for these species. Hydrophilic aerosols (BC2 and OC2) show larger errors than hydrophobic aerosols (BC1 and OC1), particularly at levels above 800 hPa. In the GOCART aerosol scheme, the emitted carbonaceous aerosols are initially hydrophobic, and then they become hydrophilic through the aging process with a typical e-folding time scale of approximately 2.5 days [55]. This process causes a sharp decrease in errors in BC1 and OC1 from the surface, and the errors are maximized right above the atmospheric boundary layer in BC2 and OC2. Therefore, hydrophilic aerosols are more prone to large-scale transport by atmospheric winds above the boundary layer. The hydrophilic black carbon (BC2) undergoes further transformation into hydrophilic organic carbon (OC2). This is to compensate for the secondary organic aerosol formation which is not adequately represented in the GOCART scheme. Therefore, increased OC2 shows the greatest error among carbonaceous species. This process is dominant in winter, with other seasons exhibiting comparable error magnitudes.

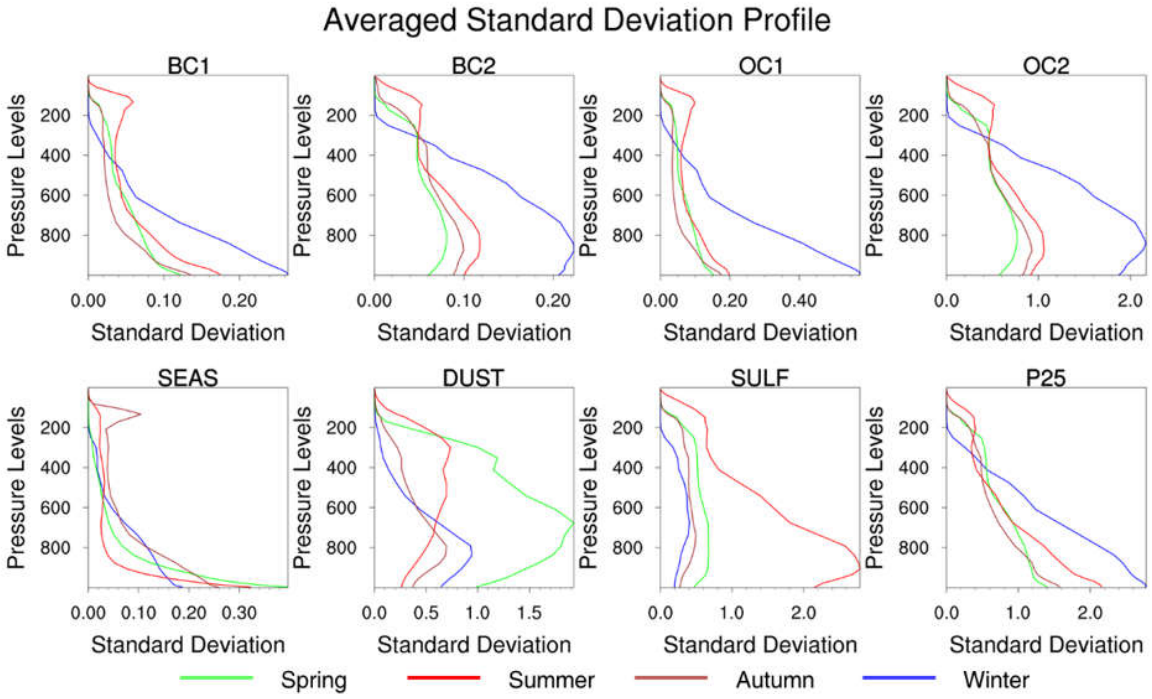

Figure 3. Vertical distributions of the standard deviation of the model background error for 8 GOCART aerosol species averaged over Domain 2 in winter (blue), spring (green), summer (red), and autumn (brown). Values in four different-sized bins are averaged for dust and sea salt.

The natural dust error is the greatest in spring owing to the more frequent Asian Dust transport from the inland continental area. The mid-troposphere near the 600 hPa level shows the largest error rather than the ground level, implying that this species is transported to Domain 2 by background winds rather than originating from the surface. When the prevailing wind at low elevations changes to being southerly in summer, the error for dust becomes small, particularly in the lower atmosphere. When the low-level wind blows from the ocean toward the continent in summer, there are substantial errors in the upper troposphere, where westerlies prevail in all seasons. In contrast, the forecast error did not show a clear seasonal variation for natural sea salt.

The background error for sulfate was the greatest in summer and lowest in winter. Sulfate aerosols are mostly produced by anthropogenic sources, such as power plants and

industries. This study used the EDGAR-HTAP emission, which assumes the emission of only sulfur dioxide (SO_2) gas, with no emission of aerosols. Therefore, all sulfate aerosols in the model forecast formed secondarily. In our investigation, even though the SO_2 emissions by EDGAR-HTPA peaked in winter, the simulated sulfate aerosol concentration peaked in summer. The transformation of SO_2 into sulfate aerosols is affected by many factors, including temperature and other meteorological conditions. The sulfate concentration maxima in the warm season are also supported by observational data from Seoul [56], as abundant sunlight and high relative humidity are favorable meteorological conditions for secondary aerosol formation [57].

3.2. Performance of Data Assimilation

Next, this study examined the overall performance of data assimilation for PM_{10} and $PM_{2.5}$ at the surface. Figure 4 shows the average surface PM concentration for Domain 1. The 6-hourly PM analysis fields were averaged for January, April, July, and October 2017 to represent the four seasons. The values over the desert area are masked out intentionally in the figure because of the excessive dust concentration in the model, and there were no reliable ground-based stations to verify the model simulation. The spatial patterns for PM_{10} and $PM_{2.5}$ were quite similar in NoDA (cf. Figure 4b,f). Both underestimated the concentration in Manchuria and South Korea and overestimated it over the Sichuan Basin and inland China. The only exception was the Beijing–Tianjin–Hebei (BTH) area, which is one of the most polluted regions in China, where NoDA underestimated PM_{10} but overestimated $PM_{2.5}$. DA showed a significant improvement compared to NoDA. First, the overall overestimation bias in central and southern China and the underestimation bias in northern China and South Korea in the PM_{10} concentration decreased through data assimilation.

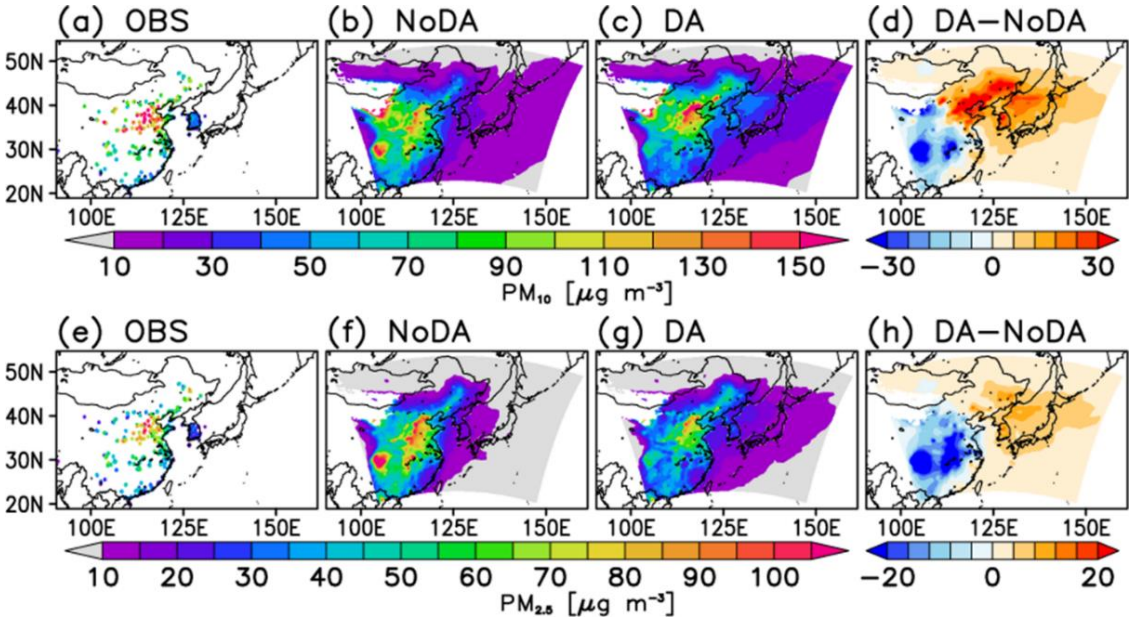

Figure 4. Seasonal average of surface PM concentration analysis at the lowest model level for (**a**,**e**) observation, (**b**,**f**) NoDA, and (**c**,**g**) DA experiments on Domain 1. (**d**,**h**) Differences in DA and NoDA experiments. The top row shows PM_{10}, and the bottom row shows $PM_{2.5}$. The desert region is masked out.

Figure 5 compares the spatial distribution of the time-averaged PM concentration in Domain 2. The comparison focused on South Korea to examine the data assimilation impacts at a high resolution of 9 km. The figure shows observations only in South Korea owing to poor data availability. The surface PM concentration showed a local maximum in the northwest around the Seoul metropolitan area and its vicinity. As the ground-

based observation stations are clustered mainly in the cities, the PM concentration also showed local maxima around cities, with many data gaps over rural areas. As in East Asia (Domain 1), NoDA significantly underestimated the PM_{10} and $PM_{2.5}$ concentrations in South Korea. Although the concentration generally increased from west to east, like in the observations, the forecast failed to clearly delineate local contributions by major cities and industrial regions. This seems to be partly due to the substantial uncertainty in the local emissions specified by the EDGAR-HTAP. DA showed a spatial distribution that was in good agreement with the observations, although it still had difficulty in reproducing high values induced by local pollution sources such as in the southeast.

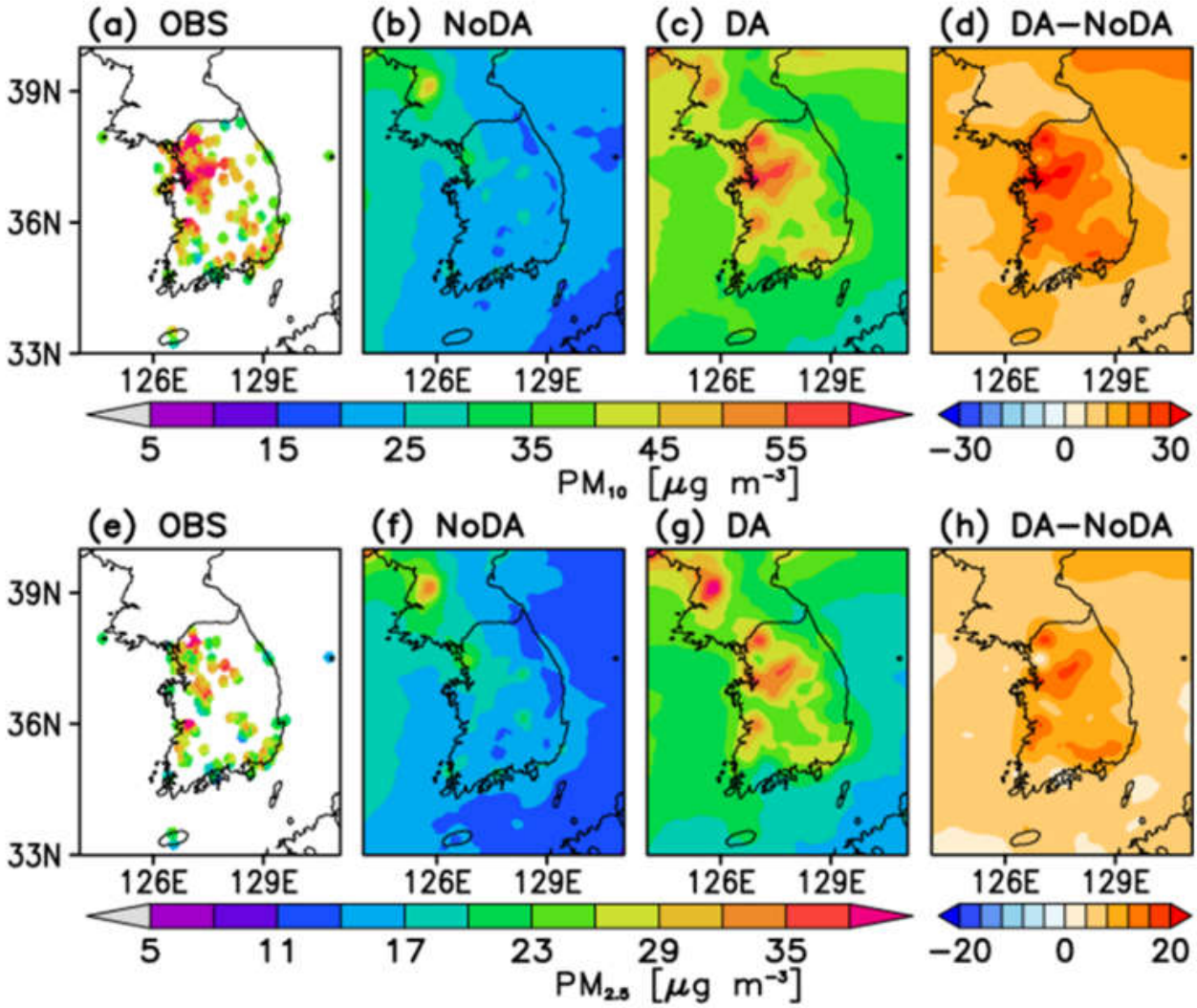

Figure 5. Same in Figure 4 except for Domain 2.

The impact of data assimilation was quantitatively assessed based on scatter plots between observations and model analysis (Figure 6). NoDA showed poor consistency with the observations for both PM_{10} and $PM_{2.5}$, with underestimation biases. DA showed a better alignment with the observed values, with a correlation coefficient as high as 0.86 for PM_{10} and 0.88 for $PM_{2.5}$. The RMSE was also significantly reduced, indicating that aerosol analysis was considerably improved through data assimilation.

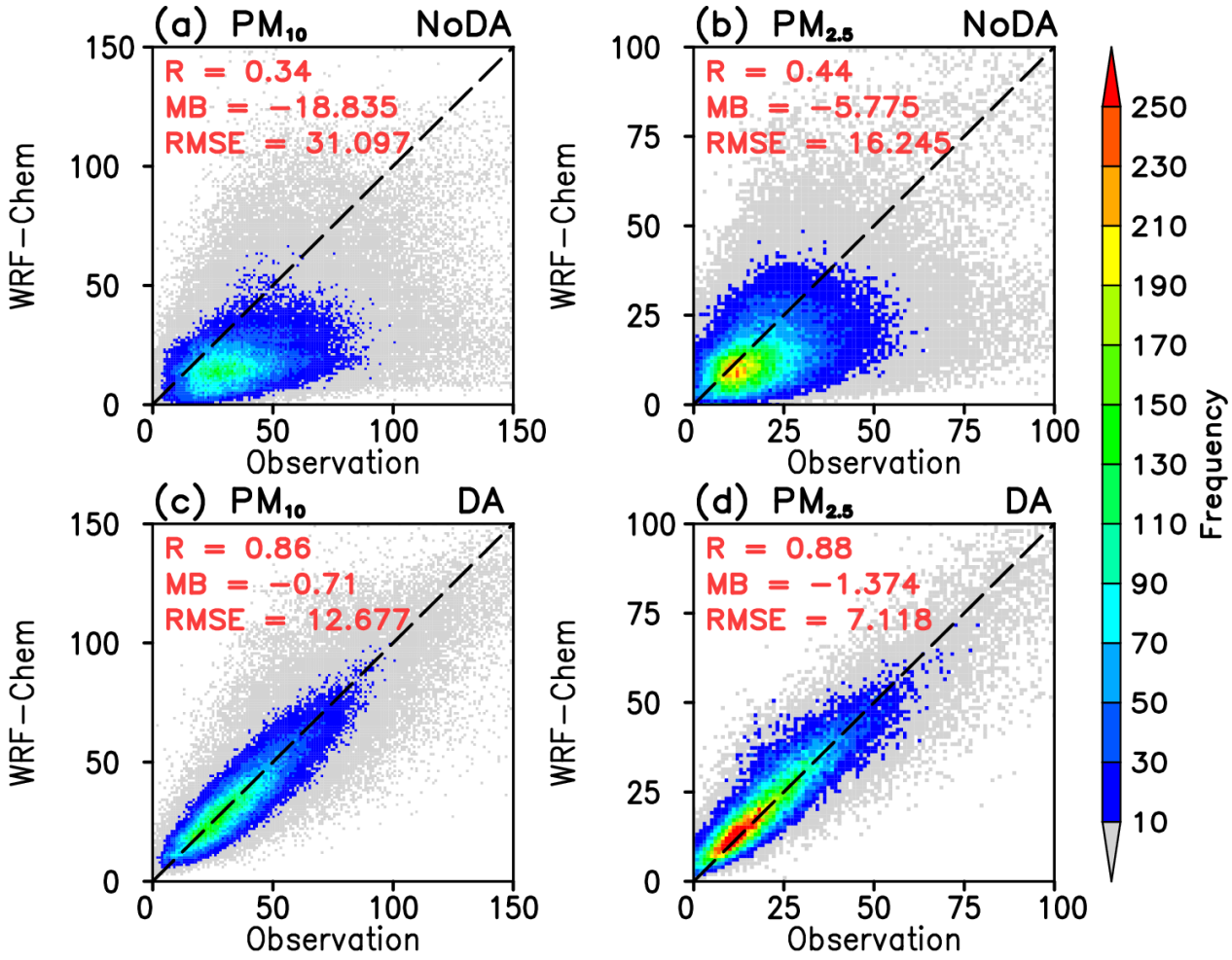

Figure 6. Scatter plots of PM_{10} and $PM_{2.5}$ values from observations and the model analysis results from (**a**,**b**) NoDA and (**c**,**d**) DA experiments over South Korea. The shaded values indicate the frequency. The correlation (R), mean bias (MB), and root-mean-square error (RMSE) values are also provided in each panel.

The surface PM_{10} distributions and low-level (850 hPa) winds during each season are compared in Figure 7. The PM_{10} concentration was the highest in January and April in the BTH region and lowest in July and October in the BTH region and South Korea. In January, north-westerly winds prevail over East Asia. In the meantime, the PM_{10} concentration in the BTH region showed a peak, presumably due to increased anthropogenic emissions. The concentration also increased in South Korea, indicating the impacts of large-scale transport by the north-westerlies as well as an increase in local emissions. Even though the concentration tends to decrease in April over the BTH region, likely because of lower emissions, the PM_{10} concentration increased on the downstream side over the Korean Peninsula, East Sea, East China Sea, and the Japanese archipelago. This is possibly because of the eastward and southward transport of Asian dust from the Gobi and Taklamakan deserts in spring by the background winds. In summer and autumn, the south-westerlies prevail over South Korea because of the high pressure of the North Pacific Ocean; the wind blows from the relatively clean South Sea in Korea, and the effect of the long-distance transport of fine dust is low. The seasonal variation of $PM_{2.5}$ (not shown) is similar to that of PM_{10}.

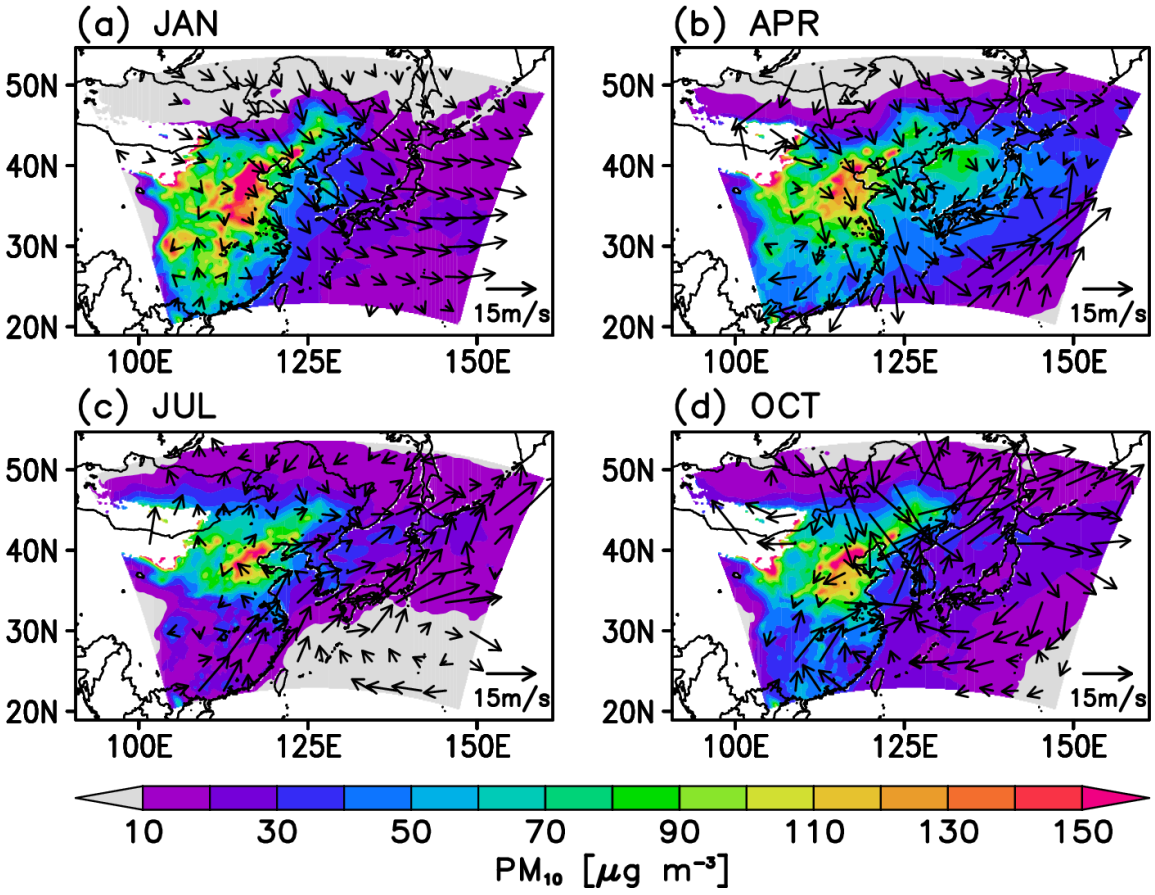

Figure 7. Monthly mean distributions of PM_{10} (shaded) and 850 hPa wind (vector) from the DA experiments for (**a**) January, (**b**) April, (**c**) July, and (**d**) October.

Figure 8 compares the time average and standard deviation of PM concentration in each season in South Korea. The observed PM_{10} values showed pronounced seasonal variations, with high values in April and January and low values in July and October. Characterized by strong seasonality in low-level circulation due to the East Asian monsoon, the background aerosol concentration reached high values in spring and winter, when northerlies or westerlies prevailed and transported air pollutants from continental areas. The background concentration decreased when the low-level winds turned into southerlies and easterlies during summer and fall. The seasonal evolution of $PM_{2.5}$ was not very different from that of PM_{10}, although the latter showed a peak in the Asian Dust season in spring [58,59]. This also suggests that anthropogenic influences are more important for $PM_{2.5}$ than for PM_{10}. Previous studies have suggested the importance of removal processes that contribute to the seasonality of aerosol concentrations, such as the high PM concentration in winter due to dry and stagnant weather conditions and the low concentration in summer due to aerosol washout by precipitation [60]. The contribution of the dry and wet deposition of aerosols seems to partially support their importance, based on our results. As measured by the standard deviation, the daily variation showed a larger amplitude in spring and winter, indicating large day-to-day variations owing to transporting wind and meteorological conditions rather than stagnant conditions. The East Asian summer monsoon is also characterized by intensive rainfall days that vary substantially daily. This suggests that seasonal variations in aerosol concentrations are affected not only by natural and anthropogenic emissions but also by meteorological conditions.

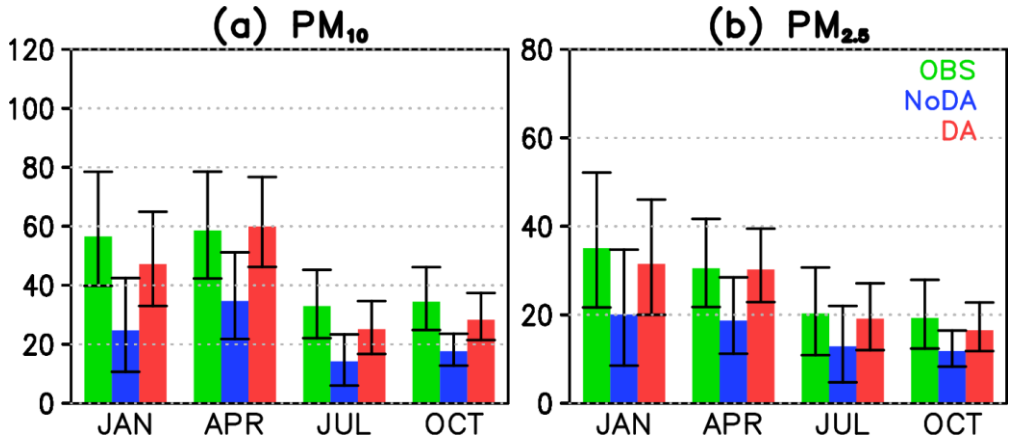

Figure 8. Time-averaged (**a**) PM_{10} and (**b**) $PM_{2.5}$ concentrations in South Korea in each season for the observations (green), NoDA (blue), and DA (red). The one standard deviation of daily values is indicated by each bar.

In general, NoDA also reproduced the observed seasonal variations in PM_{10}, with high values in spring and winter and low values in summer and autumn. One deficiency is a relatively weak concentration in January, when the observed and simulated values showed the greatest difference. This suggests that the model underestimates local anthropogenic sources and transport from the west. NoDA also reproduces the seasonal variations in $PM_{2.5}$, but with an overall underestimation in all seasons, as for PM_{10}. DA showed significantly improved results in terms of the monthly mean values and seasonal variations. The larger standard deviation of PM concentrations in winter and spring was also well-represented in both the NoDA and DA.

3.3. Forecast Verification

Next, we examined the impacts of aerosol data assimilation on the forecasts for the initial 24 h. Figure 9 compares the correlations and normalized RMSEs (nRMSEs) for the surface PM forecasts from NoDA and DA for all seasons. The correlation coefficient was also compared with that of the persistence forecast. In the persistence forecast, the forecast values were assumed to be identical to the initial observed values for the entire forecasting time. Starting from 1.0, the persistence forecast accuracy gradually decreased as the forecast lead time increased. The curve of the correlation coefficient for the persistence forecast dropped below 0.4 after 24 h for both PM_{10} and $PM_{2.5}$ forecasts. Any prediction above this persistent curve can be regarded as practically meaningful. However, the nRMSE curve tends to increase over time owing to increasing forecast errors, as in the persistence forecast.

The NoDA forecast for PM_{10} (Figure 9a) displayed poor accuracy with a correlation coefficient of approximately 0.4 for the initial 24 h. This is even lower than that of the persistence forecast, suggesting that the NoDA forecast is not useful. The NoDA forecast for $PM_{2.5}$ (Figure 9b) showed similar behavior, although the overall correlation coefficient was slightly higher than that of the PM_{10} forecasts for all forecasting lead times. The nRMSEs for both PM_{10} and $PM_{2.5}$ forecasts (Figure 9c,d) were not changed significantly depending on the forecast lead time. When the forecast was initialized with data assimilation, both PM_{10} and $PM_{2.5}$ forecasts improved significantly. They displayed a correlation coefficient of 0.9 or more at the time of data assimilation, and continued to maintain a correlation coefficient above 0.5 until 24 h. In particular, the $PM_{2.5}$ forecasts were better performed than the PM_{10} forecasts, with a correlation coefficient higher than 0.6 until 24 h. Both correlation coefficient curves from the DA experiments suggest that data assimilation is beneficial for PM forecasts. This was statistically evident, at least for the initial 6 h. However, the forecasting performance showed some deficiencies, such that the correlation coefficients decreased faster than the persistence curve for the initial few hours with higher nRMSE values, particularly in the PM_{10} forecasts.

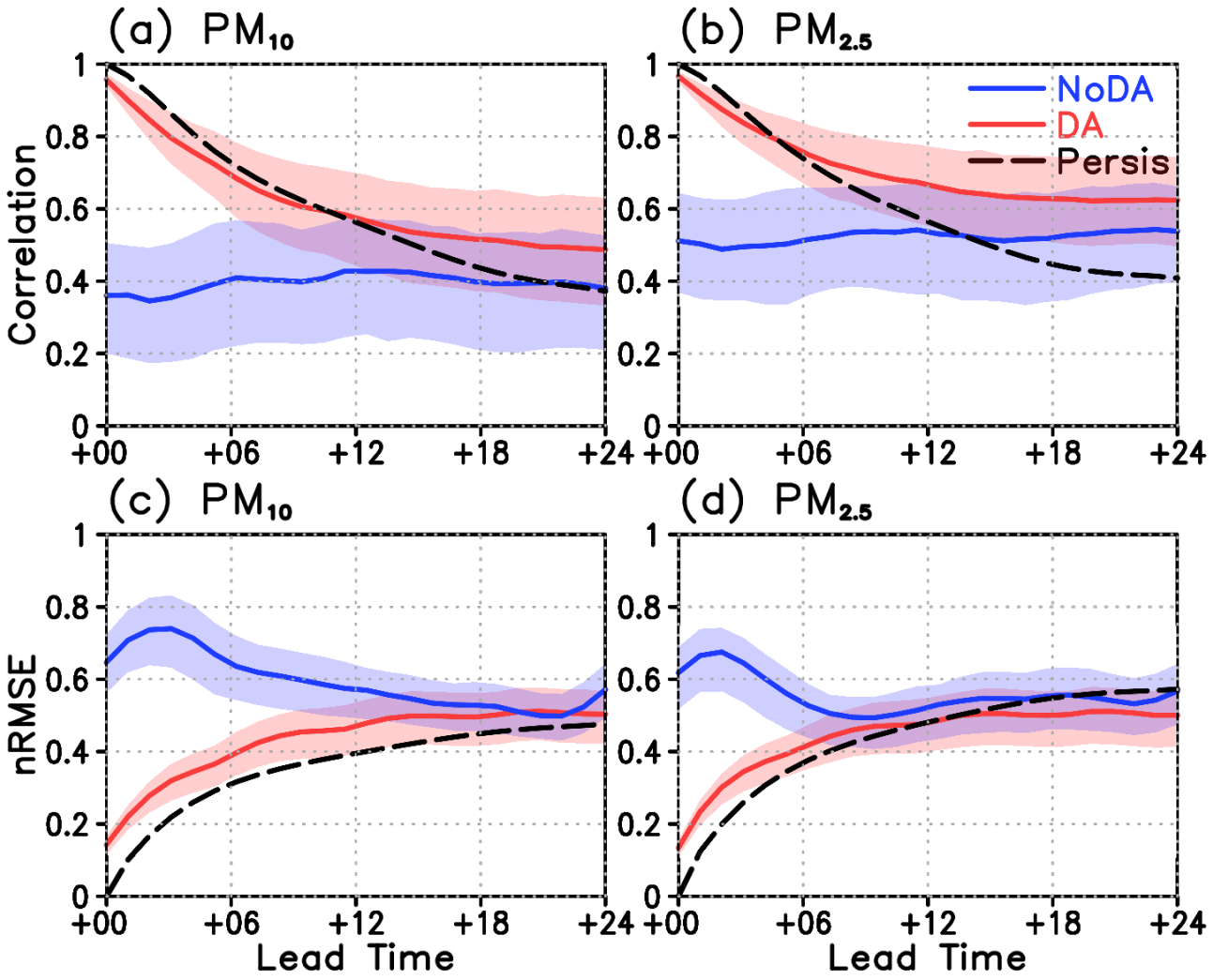

Figure 9. (**a**,**b**) Correlation coefficient and (**c**,**d**) normalized RMSE (nRMSE) curves for initial 24 h for the PM_{10} and $PM_{2.5}$ forecasts with NoDA (blue) and DA (red). The persistence curves (black dotted) are included as a baseline prediction for comparing the accuracy with numerical forecasts. RMSE is normalized by the observed mean values for PM_{10} and $PM_{2.5}$. The color shading indicates the variability of the correlation coefficient and nRMSE shown as one standard deviation value calculated using 100,000 bootstrap iterations.

The dependence of the forecast accuracy on the season is shown in Figure 10. The superior accuracy for the $PM_{2.5}$ forecasts is evident in all seasons compared with those of the PM_{10} forecasts. The positive impacts of DA on aerosol forecasts were also statistically clear for at least the initial 6 h, regardless of the season. The forecast accuracy also exhibited pronounced seasonal variations in terms of the correlation and normalized RMSEs. The forecast accuracy worsened in spring (April) and improved in summer (July) for both PM_{10} and $PM_{2.5}$ forecasts. As discussed in Section 3.1, the poor forecast accuracy in spring seems to be mainly due to the misrepresentation of natural dust processes by the model, which exaggerates dust transport in this region (not shown). The model showed great difficulty in data assimilation for both PM_{10} and $PM_{2.5}$, as indicated by the relatively low correlation even at the time of initialization. The rest of the season demonstrated higher accuracy. For example, the DA forecasts in winter (January) outperformed the NoDA and persistence forecasts throughout the initial 24 h. Considering the higher $PM_{2.5}$ concentrations in winter than in spring (Figure 8), this result indicates that the problem of simulating natural dust affects the accuracy of background PM_{10} concentration the most.

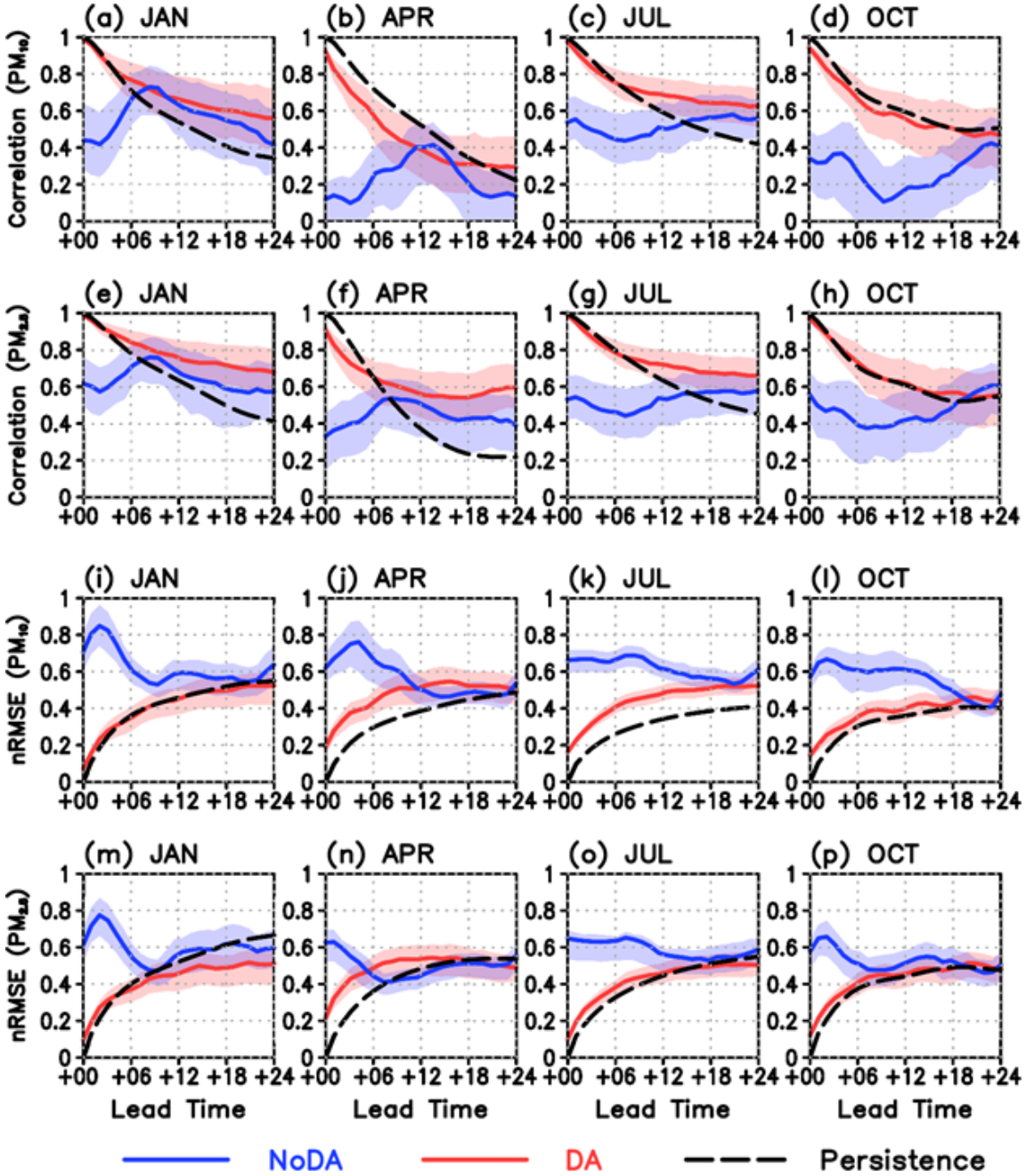

Figure 10. (**a**–**h**) Correlation coefficient and (**i**–**p**) normalized RMSE (nRMSE) curves for the NoDA (blue) and DA (red) forecasts and persistence curves (black dotted) in each month. Other parameters are the same as in Figure 9.

3.4. Bias Correction

As shown above, the impact of data assimilation lasts mostly for the initial hours, after which the forecast error grows and degrades the forecast accuracy owing to the systematic model bias. Systematic bias is caused by various factors, such as deficiencies in chemistry and aerosol parameterization, emissions, and simulated meteorological conditions, and is often characterized by the differences in the time mean and variance between the observed and forecast values [61]. While improving the forecast accuracy by improving the model itself is challenging, the accuracy can be substantially improved by statistically correcting the forecast values. A straightforward method is to replace the mean and standard deviation of the forecast values with the observed statistics based on Equation (2).

Figure 11 highlights the systematic bias in the current WRF-Chem forecasts, where the NoDA forecasts significantly underestimate both PM_{10} and $PM_{2.5}$ concentrations. Although the forecasts initialized by DA improved the histogram for $PM_{2.5}$, systematic bias remained in PM_{10} forecasts. Simple bias correction in the mean and standard deviation (BiasC) provides a better agreement in PM_{10} and $PM_{2.5}$ histograms, enabling forecasts for the observed high values.

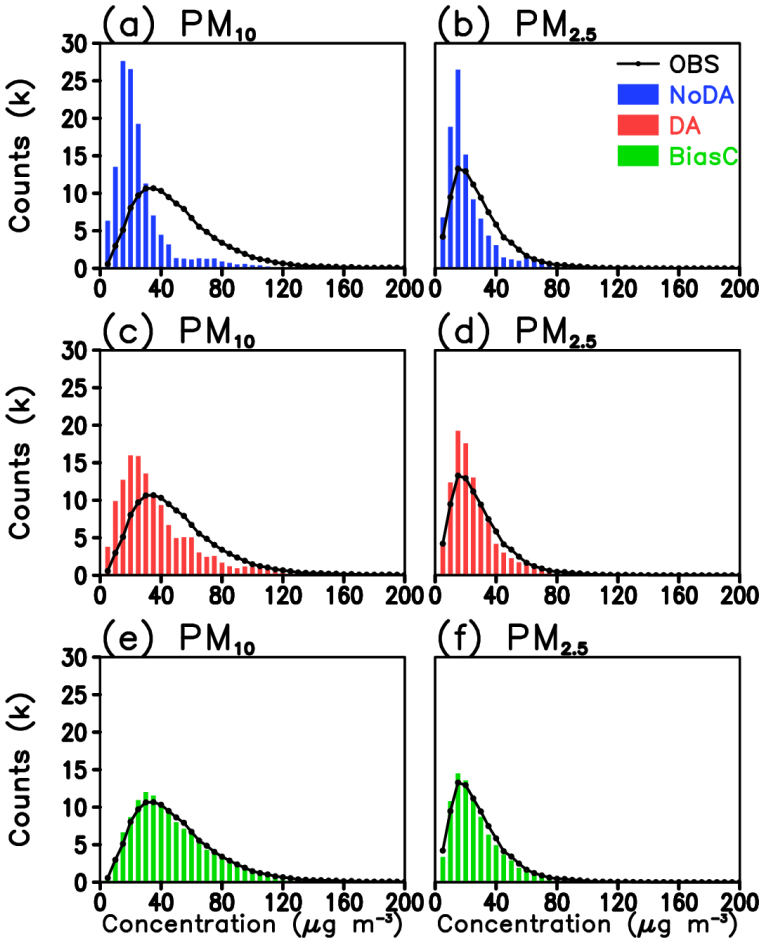

Figure 11. Histograms of the PM_{10} (left) and $PM_{2.5}$ (right) concentrations from the forecasts by (**a**,**b**) NoDA, (**c**,**d**) DA, and (**e**,**f**) the bias correction (BiasC). All forecasts are the values after 24 h from the start. The solid lines with closed dots indicate the observed values.

Finally, we compared the categorical forecasting accuracy for each forecasting method (Figure 12). The hit rate (HR) for the NoDA forecasts for PM_{10} and $PM_{2.5}$ remained below 50%, and even lower than 10% for the polluted cases, indicating a low forecast accuracy. Regardless of the forecast lead time, the false-alarm rates (FAR) for PM_{10} and $PM_{2.5}$ were higher than 80% and 60%, respectively. In addition, no systematic changes in HR or FAR accuracy of the NoDA forecasts were observed with an increase in forecast lead time, implying that the forecast skill is almost insignificant. On the other hand, both DA and BiasC forecasts exhibited systematic changes with a decrease in HR and an increase in FAR over time, as the influence of initial conditions on the forecast diminished. Overall, DA forecasts showed significantly improved forecast performance compared with NoDA because of the improvement in initial conditions due to data assimilation. However, this indicated a poor HR of less than 50% for the polluted cases for PM_{10}, although the $PM_{2.5}$ forecasts performed slightly better. The BiasC forecasts improved the accuracy in all cases compared with the DA forecasts. The bias correction particularly improved the $PM_{2.5}$ forecasts for the polluted cases, with a HR higher than 50% for the 6 h forecast.

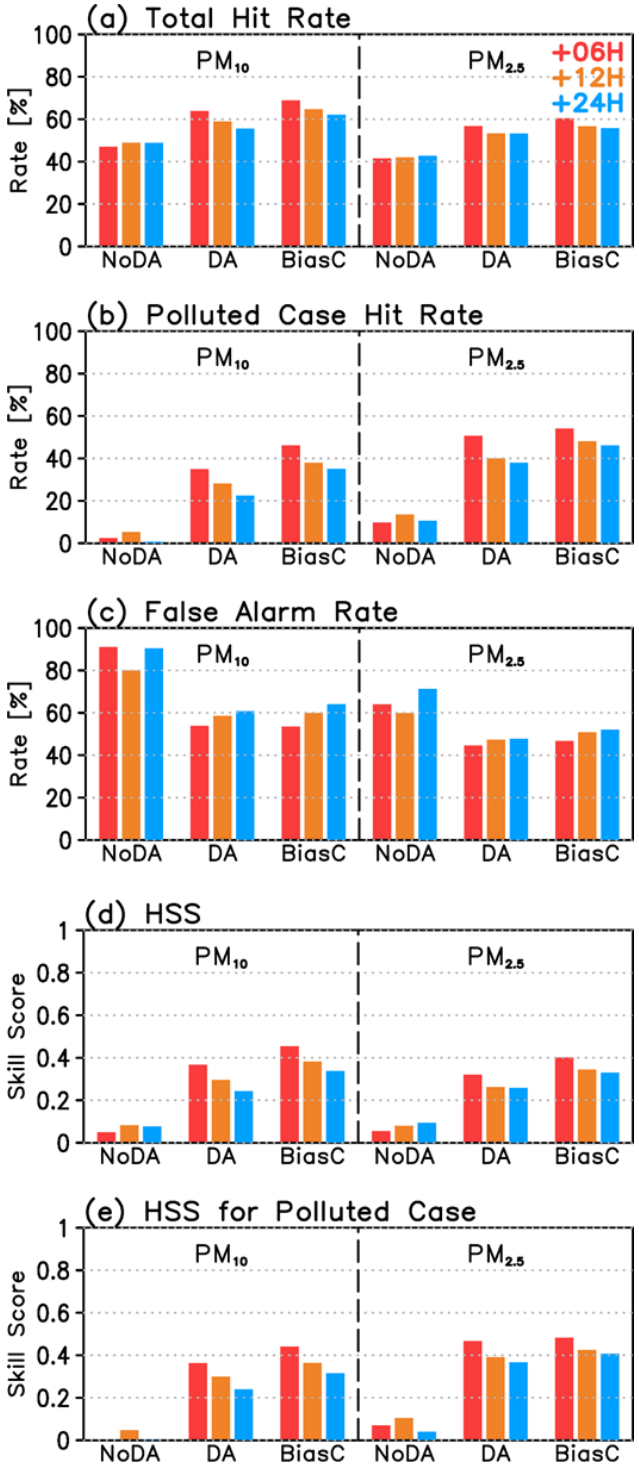

Figure 12. Categorical forecast accuracy for NoDA, DA, and BiasC for PM_{10} (left) and $PM_{2.5}$ (right) as measured by (**a**) the hit rate (HR), (**b**) HR for the polluted cases, and (**c**) false-alarm rate, (**d**) Heidke Skill Score (HSS), and (**e**) HSS for the polluted cases. The color bars indicate the accuracy after 6 h (red), 12 h (orange), and 24 h (blue) after initialization. The polluted cases are those when the PM_{10} ($PM_{2.5}$) concentration is above 80 (35) μg m^{-3}.

Note that BiasC increases HR and FAR. As shown in Figure 12, the DA forecasts generally underestimated the observed high concentration values. False alarms can also occur when high concentration forecasts increase because of bias correction. In this regard, the comparison of HSS can be more objective while comparing the forecast performance, as

it considers both the accuracy and false alarms. The HSSs were close to zero for all forecast lead times in the NoDA forecasts, which indicates poor forecasting accuracy. When data assimilation was applied to initialize the forecast model, the HSS for forecasts increased to 0.37 for 6 h forecasts and remained above 0.2 until the 24 h forecasts for PM_{10} (Figure 12d). The HSS for $PM_{2.5}$ was slightly higher than that for PM_{10}. The DA forecasts maintained comparable HSS for the polluted cases, with a greater improvement in $PM_{2.5}$. The bias correction of the DA forecasts contributed to an increase in HSS for all forecast lead times and for both PM_{10} and $PM_{2.5}$. In particular, the HSS from the BiasC forecasts maintained a HSS above 0.4 up to 24 h in the polluted $PM_{2.5}$ cases. The results suggest that even a simple bias correction of the forecast outputs can substantially increase the prediction skill, in addition to the proper initialization of the numerical forecast model by the DA.

4. Discussion

This study clearly shows that the assimilation of satellite and ground-based data for use in numerical air quality forecasting models can significantly improve surface PM forecasts. However, its beneficial impact lasts for a few hours from the start of the forecast, and then tends to gradually decrease due to forecast biases. Specifically, the 24 h forecasts by the current model present a lower concentration than those observed, thereby limiting the accuracy beyond a day, particularly for polluted cases. This study attempted a simple bias correction method to help reduce the underestimation bias, but the hit rate for the polluted cases remained less than 50% after 24 h of forecast. This problem represents the fundamental difficulty in state-of-the-art air quality forecasting systems. In our study, the systematic biases of the model forecasts also varied according to seasons, and the developed model showed a pronounced seasonal variation in the quality of aerosol analysis and forecast accuracy. Therefore, systematic model biases should be carefully understood by rigorously testing the model under various conditions and examining the baseline representation of the formation, transport, and deposition of major aerosol species.

A few deficiencies were identified for this specific model in our examination of the seasonal dependence of the forecast performance. The model forecasts were especially poor in spring, when natural dust transport prevails in East Asia. The current natural dust scheme implemented in GOCART is based on Ginoux et al. [62], and parameterizes the dust emission flux with surface wind and topographical complexity. Considering its significant contribution in spring, the parameterization uncertainty should be carefully examined using comprehensive observational and independent reanalysis data. In particular, the sensitivity of dust emissions to meteorological forcing should be investigated. The current study specified the forecasted meteorological fields using the Korea Meteorological Administration unified model, which is not likely to be accurate because of weather forecast errors. In our preliminary test, the model showed reduced dust emissions and weaker long-range transport when we specified atmospheric reanalysis data instead of unified model forecasts.

Another deficiency of the GOCART scheme is that secondary aerosol formation is not considered despite its importance in Northeast Asia. In particular, secondary nitrate aerosol formation, known for its importance in major cities in Korea [63], was not represented in the scheme. Sensitivity experiments concerning aerosol schemes, particularly in the context of secondary aerosol formation, are underway, using other schemes that consider secondary aerosols in more detail, such as MADE/VBS [64].

Finally, there is a substantial uncertainty in the current anthropogenic emissions, which should be specified in the model. The current system uses EDGAR-HTAP data based on the inventory archived in 2010. This is obviously outdated, considering the strong countermeasures against anthropogenic emissions in recent years in Korea and China. However, when this study tested other emission data with an updated emission inventory in Northeast Asia in 2015, the results did not change significantly. Bottom-up emission data may contain substantial uncertainties, the reliability of which should be quantitatively discussed using inverse modeling techniques suggested in recent studies [65].

5. Conclusions

In this study, an aerosol data assimilation system was developed which assimilates satellite AODs from MODIS and GOCI and ground-based PM_{10} and $PM_{2.5}$ observations in Northeast Asia. The WRF-Chem model with interactive meteorology and chemistry simulates the entire domain with a horizontal resolution of 27 km, covering China, Korea, and Japan, with a nested domain over South Korea at a finer resolution of 9 km. The data assimilation used the 3D-VAR method applied every 6 h to enable aerosol analysis, and the impacts of aerosol data assimilation on the forecast skill for up to 24 h were evaluated for all seasons by conducting data assimilation and forecasting experiments in January, April, July, and October 2017.

Without aerosol data assimilation, the baseline forecasts (NoDA) showed a large discrepancy from the observed PM_{10} and $PM_{2.5}$ variations in time and space. They also provided poor initial conditions for the forecasting model, which substantially degraded the forecasting accuracy. Data assimilation with satellite and ground-based observations significantly improved surface PM_{10} and $PM_{2.5}$ analyses with a higher correlation with the observed data than that in the case without data assimilation. It also showed beneficial impacts on the PM_{10} and $PM_{2.5}$ forecasts over South Korea up to 24 h from the initial state. This study also highlighted that the analysis quality improvement induced by data assimilation varied significantly depending on the season, resulting in seasonal forecast accuracy changes. As the dominant aerosol species and prevailing atmospheric circulation differ across seasons, data assimilation and forecast quality showed distinctive seasonal variations. Specifically, the forecast accuracy was significantly degraded in spring, presumably because of the poor initialization of natural dust. The model performed better in winter when anthropogenic sources were dominant.

As the DA forecast progressed over time, its accuracy tended to decrease due to the systematic model bias. Current WRF-Chem forecasts with DA significantly underestimate both PM_{10} and $PM_{2.5}$ concentrations. Simple bias correction (BiasC) in the mean and standard deviation of the DA forecasts provided better agreement with the observed distributions of PM_{10} and $PM_{2.5}$ concentrations. The bias-corrected forecasts improved the accuracy in the total and polluted cases compared to the DA forecasts. They particularly improved $PM_{2.5}$ forecasts for polluted cases with a HR higher than 50% for the 6 h forecast. The bias correction of the DA forecasts also improved the HSS for all forecast lead times. In particular, the HSS after bias correction maintained a value above 0.4 for up to 24 h in the polluted cases for $PM_{2.5}$. The HSS increased by 37% through data assimilation and further improved by 10% through bias correction.

Author Contributions: Conceptualization, M.-I.L.; Formal analysis, S.L. and G.K.; Funding acquisition, M.-I.L. and C.-K.S.; Investigation, S.L. and G.K.; Methodology, S.L., G.K. and Y.C.; Project administration, M.-I.L.; Resources, C.-K.S. and H.-K.K.; Software, S.L., G.K. and Y.C.; Supervision, M.-I.L.; Validation, S.L. and G.K.; Visualization, S.L.; Writing—original draft, S.L. and G.K.; Writing—review and editing, M.-I.L., Y.C., C.-K.S. and H.-K.K. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Basic Science Research Program of the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (NRF-2021R1A2C1008210). This research was also supported by the FRIEND (Fine Particle Research Initiative in East Asia Considering National Differences) Project through the National Research Foundation of Korea (NRF), funded by the Ministry of Science and ICT (grant no. 2020M3G1A1114615).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: WRF-Chem model version 3.9.1 is available on https://www2.mmm.ucar.edu/wrf/users/download/get_sources.html#WRF-Chem (accessed on 26 April 2022). The ground-based PM_{10} and $PM_{2.5}$ observational data were downloaded from the Beijing Municipal Environmental Protection Monitoring Center (http://beijingair.sinaapp.com/ (accessed on 26 April

2022) for China and the National Institute of Environmental Research AirKorea website (https://www.airkorea.or.kr/web/last_amb_hour_data?pMENU_NO=123 (accessed on 26 April 2022); only available in Korean) for South Korea. The Geostationary Ocean Color Imager (GOCI) AOD Yonsei aerosol retrieval (YAER) version 2 data are available upon request to Choi et al. [51]. The other data presented in this study are available upon reasonable request from the corresponding author.

Acknowledgments: The model simulations were performed by using the supercomputing resource of the Korea Meteorological Administration (National Center for Meteorological Supercomputer). The authors thank Myungje Choi and Jhoon Kim from Yonsei University for providing GOCI AOD data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shao, Y.; Dong, C.H. A review on East Asian dust storm climate, modelling and monitoring. *Glob. Planet. Chang.* **2006**, *52*, 1–22. [CrossRef]
2. Nakajima, T.; Yoon, S.-C.; Ramanathan, V.; Shi, G.-Y.; Takemura, T.; Higurashi, A.; Takamura, T.; Aoki, K.; Sohn, B.-J.; Kim, S.-W.; et al. Overview of the Atmospheric Brown Cloud East Asian Regional Experiment 2005 and a study of the aerosol direct radiative forcing in east Asia. *J. Geophys. Res.* **2007**, *112*, 24–91. [CrossRef]
3. Lee, S.; Lee, M.-I.; Song, C.-K.; Kim, K.-M.; da Silva, A.M. Interannual variation of the East Asia Jet Stream and its impact on the horizontal distribution of aerosol in boreal spring. *Atmos. Environ.* **2020**, *223*, 117296. [CrossRef]
4. Zhu, C.; Kanaya, Y.; Yoshikawa-Inoue, H.; Irino, T.; Seki, O.; Tohjima, Y. Sources of atmospheric black carbon and related carbonaceous components at Rishiri Island, Japan: The roles of Siberian wildfires and of crop residue burning in China. *Environ. Pollut.* **2019**, *247*, 55–63. [CrossRef]
5. Ohara, T.; Akimoto, H.; Kurokawa, J.; Horii, N.; Yamaji, K.; Yan, X.; Hayasaka, T. An Asian emission inventory of anthropogenic emission sources for the period 1980–2020. *Atmos. Chem. Phys.* **2007**, *7*, 4419–4444. [CrossRef]
6. Park, R.J.; Kim, S.-W. Air quality modeling in East Asia: Present issues and future directions. *Asia-Pac. J. Atmos. Sci.* **2014**, *50*, 105–120. [CrossRef]
7. Chatani, S.; Amann, M.; Goel, A.; Hao, J.; Klimont, Z.; Kumar, A.; Mishra, A.; Sharma, S.; Wang, S.X.; Wang, X.Y.; et al. Photochemical roles of rapid economic growth and potential abatement strategies on tropospheric ozone over South and East Asia in 2030. *Atmos. Chem. Phys.* **2014**, *14*, 9279–9293. [CrossRef]
8. Moon, K.-J.; Cheo, H.; Jeon, K.; Yang, X.; Meng, F.; Kim, D.; Park, H.-J.; Kim, J. Review on the Current Status and Policy on $PM_{2.5}$ in China. *J. Korean Soc. Atmos. Environ.* **2018**, *34*, 373–392. [CrossRef]
9. Lee, K.; Chandra, I.; Seto, T.; Inomata, Y.; Hayashi, M.; Takami, A.; Yoshino, A.; Otani, Y. Aerial Observation of Atmospheric Nanoparticles on Fukue Island, Japan. *Aerosol Air Qual. Res.* **2018**, *19*, 981–994. [CrossRef]
10. Engel-Cox, J.; Kim Oanh, N.T.; van Donkelaar, A.; Martin, R.V.; Zell, E. Toward the next generation of air quality monitoring: Particulate Matter. *Atmos. Environ.* **2013**, *80*, 584–590. [CrossRef]
11. Zhong, M.; Saikawa, E.; Liu, Y.; Naik, V.; Horowitz, L.W.; Takigawa, M.; Zhao, Y.; Lin, N.H.; Stone, E.A. Air quality modeling with WRF-Chem v3.5 in East Asia: Sensitivity to emissions and evaluation of simulated air quality. *Geosci. Model Dev.* **2016**, *9*, 1201–1218. [CrossRef]
12. Jin, J.; Segers, A.; Heemink, A.; Yoshida, M.; Han, W.; Lin, H.X. Dust Emission Inversion Using Himawari-8 AODs Over East Asia: An Extreme Dust Event in May 2017. *J. Adv. Model. Earth Syst.* **2019**, *11*, 446–467. [CrossRef]
13. Peng, Z.; Liu, Z.; Chen, D.; Ban, J. Improving $PM_{2.5}$ forecast over China by the joint adjustment of initial conditions and source emissions with an ensemble Kalman filter. *Atmos. Chem. Phys.* **2017**, *17*, 4837–4855. [CrossRef]
14. Ma, C.; Wang, T.; Jiang, Z.; Wu, H.; Zhao, M.; Zhuang, B.; Li, S.; Xie, M.; Li, M.; Liu, J.; et al. Importance of Bias Correction in Data Assimilation of Multiple Observations Over Eastern China Using WRF-Chem/DART. *J. Geophys. Res. Atmos.* **2020**, *125*, 1–22. [CrossRef]
15. Xia, X.; Min, J.; Wang, Y.; Shen, F.; Yang, C.; Sun, Z. Assimilating Himawari-8 AHI aerosol observations with a rapid-update data assimilation system. *Atmos. Environ.* **2019**, *215*, 116866. [CrossRef]
16. Liu, Z.Q.; Rabier, F. The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study. *Q. J. R. Meteorol. Soc.* **2002**, *128*, 1367–1386. [CrossRef]
17. Ryu, J.-H.; Han, H.-J.; Cho, S.; Park, Y.-J.; Ahn, Y.-H. Overview of geostationary ocean color imager (GOCI) and GOCI data processing system (GDPS). *Ocean Sci. J.* **2012**, *47*, 223–233. [CrossRef]
18. Kim, J. GEMS (geostationary environment monitoring spectrometer) onboard the GeoKOMPSAT to monitor air quality in high temporal and spatial resolution over asia-pacific region. In Proceedings of the EGU General Assembly Conference Abstracts, Vienna, Austria, 22–27 April 2012; Volume 14, p. 4051.
19. Kalnay, E. *Atmospheric Modeling, Data Assimilation and Predictability*; Cambridge University Press: Cambridge, UK, 2003; ISBN 0521791790.
20. Liu, Z.; Liu, Q.; Lin, H.C.; Schwartz, C.S.; Lee, Y.H.; Wang, T. Three-dimensional variational assimilation of MODIS aerosol optical depth: Implementation and application to a dust storm over East Asia. *J. Geophys. Res. Atmos.* **2011**, *116*, 1–19. [CrossRef]
21. Xia, X.; Min, J.; Shen, F.; Wang, Y.; Yang, C.; Xia, C.; Min, J.Z.; Shen, F.F.; Wang, Y.B.; Yang, C. Aerosol Data Assimilation Using Data from Fengyun-3A and MODIS: Application to a Dust Storm over East Asia in 2011. *Adv. Atmos. Sci.* **2019**, *36*, 1–14. [CrossRef]

22. Jung, J.; Souri, A.H.; Wong, D.C.; Lee, S.; Jeon, W.; Kim, J.; Choi, Y. The Impact of the Direct Effect of Aerosols on Meteorology and Air Quality Using Aerosol Optical Depth Assimilation during the KORUS-AQ Campaign. *J. Geophys. Res. Atmos.* **2019**, *124*, 8303–8319. [CrossRef]
23. Saide, P.E.; Kim, J.; Song, C.H.; Choi, M.; Cheng, Y.; Carmichael, G.R. Assimilation of next generation geostationary aerosol optical depth retrievals to improve air quality simulations. *Geophys. Res. Lett.* **2014**, *41*, 9188–9196. [CrossRef]
24. Feng, S.; Jiang, F.; Jiang, Z.; Wang, H.; Cai, Z.; Zhang, L. Impact of 3DVAR assimilation of surface $PM_{2.5}$ observations on $PM_{2.5}$ forecasts over China during wintertime. *Atmos. Environ.* **2018**, *187*, 34–49. [CrossRef]
25. Cheng, Y.; Dai, T.; Goto, D.; Schutgens, N.A.J.; Shi, G.; Nakajima, T. Investigating the assimilation of CALIPSO global aerosol vertical observations using a four-dimensional ensemble Kalman filter. *Atmos. Chem. Phys.* **2019**, *19*, 13445–13467. [CrossRef]
26. Choi, Y.; Chen, S.; Huang, C.; Earl, K.; Chen, C.; Schwartz, C.S.; Matsui, T. Evaluating the Impact of Assimilating Aerosol Optical Depth Observations on Dust Forecasts over North Africa and the East Atlantic Using Different Data Assimilation Methods. *J. Adv. Model. Earth Syst.* **2020**, *12*, 1–30. [CrossRef] [PubMed]
27. Schwartz, C.S.; Liu, Z.; Lin, H.-C.; Cetola, J.D. Assimilating aerosol observations with a "hybrid" variational-ensemble data assimilation system. *J. Geophys. Res. Atmos.* **2014**, *119*, 4043–4069. [CrossRef]
28. Kim, G.; Lee, S.; Im, J.; Song, C.-K.; Kim, J.; Lee, M. Aerosol data assimilation and forecast using Geostationary Ocean Color Imager aerosol optical depth and in-situ observations during the KORUS-AQ observing period. *GISci. Remote Sens.* **2021**, *58*, 1175–1194. [CrossRef]
29. Lee, S.; Park, S.; Lee, M.I.; Kim, G.; Im, J.; Song, C.K. Air Quality Forecasts Improved by Combining Data Assimilation and Machine Learning With Satellite AOD. *Geophys. Res. Lett.* **2022**, *49*, 1–10. [CrossRef]
30. Fan, M.; He, G.; Zhou, M. The winter choke: Coal-Fired heating, air pollution, and mortality in China. *J. Health Econ.* **2020**, *71*, 102316. [CrossRef]
31. Grell, G.A.; Peckham, S.E.; Schmitz, R.; McKeen, S.A.; Frost, G.; Skamarock, W.C.; Eder, B. Fully coupled "online" chemistry within the WRF model. *Atmos. Environ.* **2005**, *39*, 6957–6975. [CrossRef]
32. Chin, M.; Ginoux, P.; Kinne, S.; Torres, O.; Holben, B.N.; Duncan, B.N.; Martin, R.V.; Logan, J.A.; Higurashi, A.; Nakajima, T. Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and Sun photometer measurements. *J. Atmos. Sci.* **2002**, *59*, 461–483. [CrossRef]
33. Brasseur, G.P.; Hauglustaine, D.A.; Walters, S.; Rasch, P.J.; Müller, J.; Granier, C.; Tie, X.X. MOZART, a global chemical transport model for ozone and related chemical tracers: 1. Model description. *J. Geophys. Res. Atmos.* **1998**, *103*, 28265–28289. [CrossRef]
34. Lin, Y.-L.; Farley, R.D.; Orville, H.D. Bulk parameterization of the snow field in a cloud model. *J. Clim. Appl. Meteorol.* **1983**, *22*, 1065–1092. [CrossRef]
35. Mlawer, E.J.; Taubman, S.J.; Brown, P.D.; Iacono, M.J.; Clough, S.A. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res. Atmos.* **1997**, *102*, 16663–16682. [CrossRef]
36. Chou, M.-D.; Suarez, M.J. *An Efficient Thermal Infrared Radiation Parameterization for Use in General Circulation Models*; NASA Technical; NASA: Greenbelt, MD, USA, 1994; pp. 1–85.
37. Grell, G.A.; Dévényi, D. A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys. Res. Lett.* **2002**, *29*, 34–38. [CrossRef]
38. Hong, S.-Y.; Noh, Y.; Dudhia, J. A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Weather Rev.* **2006**, *134*, 2318–2341. [CrossRef]
39. Chen, F.; Mitchell, K.; Schaake, J.; Xue, Y.; Pan, H.-L.; Koren, V.; Duan, Q.Y.; Ek, M.; Betts, A. Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *J. Geophys. Res. Atmos.* **1996**, *101*, 7251–7268. [CrossRef]
40. Bellouin, N.; Collins, W.J.; Culverwell, I.D.; Halloran, P.R.; Hardiman, S.C.; Hinton, T.J.; Jones, C.D.; McDonald, R.E.; McLaren, A.J.; O'Connor, F.M.; et al. The HadGEM2 family of met office unified model climate configurations. *Geosci. Model Dev.* **2011**, *4*, 723–757.
41. Emmons, L.K.; Walters, S.; Hess, P.G.; Lamarque, J.-F.; Pfister, G.G.; Fillmore, D.; Granier, C.; Guenther, A.; Kinnison, D.; Laepple, T. Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4). *Geosci. Model Dev.* **2010**, *3*, 43–67. [CrossRef]
42. Janssens-Maenhout, G.; Dentener, F.; Van Aardenne, J.; Monni, S.; Pagliari, V.; Orlandini, L.; Klimont, Z.; Kurokawa, J.; Akimoto, H.; Ohara, T. *EDGAR-HTAP: A Harmonized Gridded Air Pollution Emission Dataset Based on National Inventories*; JRC Publications Repository; Publications Office of the European Union: Luxembourg, 2012; pp. 299–2012.
43. Guenther, A.; Karl, T.; Harley, P.; Wiedinmyer, C.; Palmer, P.I.; Geron, C. Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature). *Atmos. Chem. Phys.* **2006**, *6*, 3181–3210. [CrossRef]
44. Kleist, D.T.; Parrish, D.F.; Derber, J.C.; Treadon, R.; Wu, W.-S.; Lord, S. Introduction of the GSI into the NCEP global data assimilation system. *Weather Forecast.* **2009**, *24*, 1691–1705. [CrossRef]
45. Hu, M.; Shao, H.; Stark, D.; Newman, K.; Zhou, C.; Ge, G.; Zhang, X. *Grid-Point Statistical Interpolation (GSI) User's Guide Version 3.6*; Developmental Testbed Center Rep.: Boulder, CO, USA, 2017; p. 149.
46. Han, Y. *JCSDA Community Radiative Transfer Model (CRTM): Version 1*; National Oceanic and Atmospheric Administration: Washington, DC, USA, 2006.

47. Pagowski, M.; Grell, G.A.; McKeen, S.A.; Peckham, S.E.; Devenyi, D. Three-dimensional variational data assimilation of ozone and fine particulate matter observations: Some results using the Weather Research and Forecasting—Chemistry model and Grid-point Statistical Interpolation. *Q. J. R. Meteorol. Soc.* **2010**, *136*, 2013–2024. [CrossRef]
48. Parrish, D.F.; Derber, J.C. The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Weather Rev.* **1992**, *120*, 1747–1763. [CrossRef]
49. Levy, R.C.; Mattoo, S.; Munchak, L.A.; Remer, L.A.; Sayer, A.M.; Patadia, F.; Hsu, N.C. The Collection 6 MODIS aerosol products over land and ocean. *Atmos. Meas. Tech.* **2013**, *6*, 2989–3034. [CrossRef]
50. Remer, L.A.; Kaufman, Y.J.; Tanré, D.; Mattoo, S.; Chu, D.A.; Martins, J.V.; Li, R.-R.; Ichoku, C.; Levy, R.C.; Kleidman, R.G. The MODIS aerosol algorithm, products, and validation. *J. Atmos. Sci.* **2005**, *62*, 947–973. [CrossRef]
51. Choi, M.; Kim, J.; Lee, J.; Kim, M.; Park, Y.J.; Holben, B.; Eck, T.F.; Li, Z.; Song, C.H. GOCI Yonsei aerosol retrieval version 2 products: An improved algorithm and error analysis with uncertainty estimation from 5-year validation over East Asia. *Atmos. Meas. Tech.* **2018**, *11*, 385–408. [CrossRef]
52. Elbern, H.; Strunk, A.; Schmidt, H.; Talagrand, O. Emission rate and chemical state estimation by 4-dimensional variational inversion. *Atmos. Chem. Phys.* **2007**, *7*, 3749–3769. [CrossRef]
53. Schwartz, C.S.; Liu, Z.; Lin, H.C.; McKeen, S.A. Simultaneous three-dimensional variational assimilation of surface fine particulate matter and MODIS aerosol optical depth. *J. Geophys. Res. Atmos.* **2012**, *117*, 1–22. [CrossRef]
54. Heidke, P. Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdienst. *Geogr. Ann.* **1926**, *8*, 301–349.
55. Kumar, R.; Barth, M.C.; Pfister, G.G.; Nair, V.S.; Ghude, S.D.; Ojha, N. What controls the seasonal cycle of black carbon aerosols in India? *J. Geophys. Res. Atmos.* **2015**, *120*, 7788–7812. [CrossRef]
56. Kim, Y.; Seo, J.; Kim, J.Y.; Lee, J.Y.; Kim, H.; Kim, B.M. Characterization of $PM_{2.5}$ and identification of transported secondary and biomass burning contribution in Seoul, Korea. *Environ. Sci. Pollut. Res.* **2018**, *25*, 4330–4343. [CrossRef]
57. Kim, S.U.; Kim, K.Y. Physical and chemical mechanisms of the daily-to-seasonal variation of PM_{10} in Korea. *Sci. Total Environ.* **2020**, *712*, 136429. [CrossRef]
58. Jung, M.-I.; Son, S.-W.; Kim, H.C.; Kim, S.-W.; Park, R.J.; Chen, D. Contrasting synoptic weather patterns between non-dust high particulate matter events and Asian dust events in Seoul, South Korea. *Atmos. Environ.* **2019**, *214*, 116864. [CrossRef]
59. Kim, H.C.; Kim, S.; Son, S.W.; Lee, P.; Jin, C.S.; Kim, E.; Kim, B.U.; Ngan, F.; Bae, C.; Song, C.K. Synoptic perspectives on pollutant transport patterns observed by satellites over East Asia: Case studies with a conceptual model. *Atmos. Chem. Phys. Discuss* **2016**, *673*, 1–30. [CrossRef]
60. Lee, H.S.; Kang, C.-M.; Kang, B.-W.; Kim, H.-K. Seasonal variations of acidic air pollutants in Seoul, South Korea. *Atmos. Environ.* **1999**, *33*, 3143–3152. [CrossRef]
61. Carrassi, A.; Bocquet, M.; Bertino, L.; Evensen, G. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdiscip. Rev. Clim. Chang.* **2018**, *9*, 1–50. [CrossRef]
62. Ginoux, P.; Chin, M.; Tegen, I.; Prospero, J.M.; Holben, B.; Dubovik, O.; Lin, S.-J. Sources and distributions of dust aerosols simulated with the GOCART model. *J. Geophys. Res. Atmos.* **2001**, *106*, 20255–20273. [CrossRef]
63. Jo, H.Y.; Lee, H.J.; Jo, Y.J.; Heo, G.; Lee, M.; Kim, J.A.; Park, M.S.; Lee, T.; Kim, S.W.; Lee, Y.H.; et al. A case study of heavy $PM_{2.5}$ secondary formation by N_2O_5 nocturnal chemistry in Seoul, Korea in January 2018: Model performance and error analysis. *Atmos. Res.* **2022**, *266*, 105951. [CrossRef]
64. Lee, H.J.; Jo, H.Y.; Song, C.K.; Jo, Y.J.; Park, S.Y.; Kim, C.H. Sensitivity of simulated $PM_{2.5}$ concentrations over Northeast Asia to different secondary organic aerosol modules during the KORUS-AQ campaign. *Atmosphere* **2020**, *11*, 1004. [CrossRef]
65. Ma, C.; Wang, T.; Mizzi, A.P.; Anderson, J.L.; Zhuang, B.; Xie, M.; Wu, R. Multiconstituent Data Assimilation with WRF-Chem/DART: Potential for Adjusting Anthropogenic Emissions and Improving Air Quality Forecasts Over Eastern China. *J. Geophys. Res. Atmos.* **2019**, *124*, 7393–7412. [CrossRef]