

Response to Reviewer 1:

#1. The authors have carried out inverse modeling for ozone initial conditions in the capital region of Korea, using 4D-Var data assimilation. In this process, they have expanded the capability of the CMAQ optimization code to account for correlation between the prior control variables. The optimal initial conditions are also used in a forecast to show improved model agreement with observations not used in the 4D-Var process. The manuscript presentation is clear enough to show that these beneficial advances have been made.

➤ Thank you very much for the comprehensive review of our manuscript.

#2. However, the grammatical errors throughout make it difficult to assess the scientific validity. This paper needs to be edited by a fluent English speaker before a full review can be made.

➤ Before this manuscript was published in ACPD, it has been entirely revised by native English speaker. We, however, found some wrongly edited part, and the corrected manuscript is attached in supplement. Please reconsider our revised manuscript.

#3. As a further critique, the introduction reads as a disjointed literature review of previous data assimilation work. The authors need to relate the content to the specific motivations of the current paper, connecting the dots for the reader as they go.

➤ We totally agree with this comment. Most of the previous studies have focused on a phenomena of meteorologically synoptic scale using satellite based observation as well as ground data. The transport of air pollution forced by a local circulation such as land-sea breeze is poorly studied. The limitation of earlier studies and the relation between them and this study are additionally explained in the Introduction (page 4, line 8) to help the readers understanding of the motivation of our study.

1 **Variational data assimilation for the optimized ozone initial state and the short-time**
2 **forecasting**

3

4 **S.-Y. Park¹, D.-H. Kim¹, S.-H. Lee², and H. W. Lee³**

5 [1]{Institute of Environmental Studies, Pusan National University, Busan, Korea}

6 [2]{Department of Earth Science Education, Pusan National University, Busan, Korea}

7 [3]{Division of Earth Environmental System, Pusan National University, Busan, Korea}

8 Correspondence to: H. W. Lee (hwlee@pusan.ac.kr)

9

10 **Abstract**

11 In this study, we apply the four-dimensional variational (4D-Var) data assimilation to
12 optimize initial ozone state and to improve the predictability of air quality. The numerical
13 modeling systems used for simulations of atmospheric condition and chemical formation are
14 the Weather Research and Forecasting (WRF) model and the Community Multiscale Air
15 Quality (CMAQ) model . The study area covers the capital region of South Korea, where the
16 surface measurement sites are relatively evenly distributed.

17 The 4D-Var code previously developed for the CMAQ model is modified to consider
18 background error in matrix form, and various numerical tests are conducted. The results are
19 evaluated with an idealized covariance function for the appropriateness of the modified codes.

20 The background error is then constructed using the NMC method with long-term modeling
21 results, and the characteristics of the spatial correlation scale related to local circulation is
22 analyzed. The background error is applied in the 4D-Var research, and a surface observational
23 assimilation is conducted to optimize the initial concentration of ozone. The statistical results
24 for the 12-hour assimilation periods and the 120 observatory sites show a 49.4% decrease in
25 the root mean squared error (RMSE), and a 59.9% increase in the index of agreement (IOA).

26 The temporal variation of spatial distribution of the analysis increments indicates that the
27 optimized initial state of ozone concentration is transported to inland areas by the clockwise-
28 rotating local circulation during the assimilation windows.

1 To investigate the predictability of ozone concentration after the assimilation window, a
2 short-time forecasting is carried out. The ratios of the RMSE with assimilation versus that
3 without assimilation are 8% and 13% for the +24 and +12 hours, respectively. Such a
4 significant improvement in the forecast accuracy is obtained solely by using the optimized
5 initial state. The potential improvement in ozone prediction for both the daytime and
6 nighttime with application of data assimilation is also presented.

7 **1 Introduction**

8 Data assimilation provides a consistent represent of the physical state such as the atmosphere
9 by blending imperfect model predictions and noisy observations. As a technique that applies
10 observational information to numerical models with the aim of increasing model
11 predictability, data assimilation is actively used in Numerical Weather Prediction (NWP) and
12 Ocean modeling studies (Daley, 1991; Courtier et al., 1998; Rabier et al., 2000, Kalnay, 2002;
13 Navon, 2009; Evensen, 2007). With more chemical observations available in recent years,
14 including the satellite data, data assimilation is expected to make more contributions to
15 weather forecasting and further improve the predictability of air quality. When the data
16 assimilation technique is used in an air quality model, it not only improves the initial
17 concentration distribution of pollutants, but also optimizes the emissions. In addition to the
18 boundary inflow concentration (Carmichael et al., 2008), emission is also one crucial factor in
19 the numerical prediction of various air pollutants. Several data assimilation techniques have
20 been developed. The four-dimensional variational (4D-Var) data assimilation requires an
21 adjoint model for use in non-linear numerical models. This represents an applied area in the
22 use of adjoint sensitivity (Elbern and Schmidt, 2001; Penenko et al., 2002; Sandu et al., 2005;
23 Hakami et al., 2007).

24 Research using the adjoint model in air quality models started in the mid-1990s. The adjoint
25 models used in and before the year 2000 are well described in the review paper of Wang et al.
26 (2001). Sandu and Chai (2011) and Carmichael et al. (2008) presented subsequent research,
27 and described many areas in which the adjoint method has been applied. More recently, more
28 comprehensive reviews including coupled chemistry meteorology models were well
29 addressed by Boucquet et al. (2015).

30 Elbern et al. (1997) were the first to assimilate tropospheric air quality data into the European
31 air pollution dispersion model. They argued that back then the existing air quality data
32 assimilation was limited solely to stratospheric ozone data from satellite observations, which

1 is far less than enough for better air quality prediction. In their study, they performed data
2 assimilation using both data generated by the model and various information from
3 observations. The results indicated that when using the model-generated data, the
4 predictability is improved not only for the chemical species directly related with those used in
5 the data assimilation, but also for those not used in the data assimilation. In their following
6 research, Elbern and Schmidt (2001) applied 4D-Var to cases of high summer ozone
7 concentrations based on ground observations over Europe, and ozone sonde observations
8 from other locations. The results of 6-h data assimilation showed improved predictability. In
9 addition, they also examined the sensitivities of model simulation to data assimilation based on
10 the radius of the influenced area when data assimilation was performed.

11 Chai et al. (2007) analyzed the effects of observations from various observation systems, such
12 as ground, civil aviation, ship, ozone sonde, and lidar, on data assimilation. The ICARTT
13 (International Consortium for Atmospheric Research on Transport and Transformation) data
14 was obtained and used in the above research. In particular, they proposed a method to
15 calculate background errors, which had not been addressed in detail in the previous research,
16 and verified its performance in the interested modeling area. Boissongontier et al. (2008)
17 assimilated tropospheric ozone concentrations in their regional ozone prediction study prior to
18 the launch of the MetOp Satellite of European Organisation for the Exploitation of
19 Meteorological Satellites (EUMETSAT) Polar System (EPS) in October 2006. Although the
20 study performed data assimilation using the column ozone data ranging over 0–6 km in the
21 troposphere, they expected that it would positively affect the accuracy of regional ozone
22 prediction. The chemical data assimilation has been conducted using NO₂ and HCHO from
23 the satellite, Scanning Imaging Absorption spectroMeter for Atmospheric CHartographY
24 (SCHIAMACHY), together with air quality observations at the ground level (Zhang et al.,
25 2008). The initial fields with assimilated observations were improved compared with that
26 generated without data assimilation.

27 Gou and Sandu (2011) indicated that there might exist differences in the gradient results
28 between discrete and continuous adjoint in the process of developing an adjoint model due to
29 the high non-linearity in the advection equation of the air quality model. As a result, they
30 argued that the discrete method is more accurate in the adjoint sensitivity study, and that the
31 continuous method is faster in minimizing the cost function in the 4D-Var data assimilation.
32 In their study of the background pollutants affecting ground ozone concentrations in western

1 America during the summer, Huang et al. (2013) applied data assimilation not only to
2 numerical simulations, but also to evaluation of the concentrations associated with transport.
3 Based on analysis of the ground-observed ozone concentration, they suggested that the
4 simulated surface O₃ error decreased by an average of 5 ppb and the reduction can be up to a
5 maximum of 17 ppb with application of data assimilation. The estimated background O₃ that
6 was transported from the eastern Pacific Ocean is about 3 ppb higher due to the application of
7 data assimilation.

8 Most of the previous studies for chemical data assimilation have focused on a phenomena of
9 meteorologically synoptic scale using satellite-based observation as well as ground-based data.
10 The transport of air pollution forced by a local circulation such as land-sea breeze is poorly
11 examined.

메모 포함[p1]: Response to the comment #3 of reviewer 1

12 One of the important elements affecting results of data assimilation in the 4D-Var process is
13 the background errors of the model (Talagrand and Courtier, 1987). Many previous research
14 have treated the background errors as scalar quantities with a Gaussian distribution, whereas
15 there is a lack of research applying them in a matrix form and consider the three-dimensional
16 covariance (Constantinescu et al., 2007; Singh et al., 2011; Sliver et al., 2013).

17 In this study, the region centered in the capital area of South Korea, where the ground
18 observation sites are densely distributed, is selected for the study of data assimilation. The
19 previously developed 4D-Var code has been modified to treat background errors in matrix
20 forms, and various numerical tests have been conducted. The results are evaluated using an
21 idealized covariance function. The realistic background errors are then obtained for the region
22 around the capital of South Korea using long-term modeling results. Characteristics of the
23 background errors generated in this study is analyzed. Also, the predictability of high ozone
24 concentration was investigated by setting the initial ozone concentration as control variables
25 in the cost function for the 4D-Var data assimilation.

26 **2 Methods**

27 **2.1 4D-Var data assimilation**

28 The variational method solves data assimilation problem from an optimal control framework
29 (Penenko and Obraztsov, 1976; Courtier and Talagrand, 1987; Le-Dimet and Talagrand,
30 1986). We aim to find control variables that minimize the difference between the model
31 predictions and observations. In the frame of strongly-constrained 4D-Var data assimilation,

1 the observational data at all times within the assimilation window are simultaneously
 2 considered. The control variables become the initial concentration distribution \mathbf{c}_0 , and all
 3 results at future times are uniquely determined from this in the model.

4 In the maximum likelihood approach, the 4D-Var data assimilation gives the maximum a
 5 posteriori estimator of the true initial concentration distribution, which is obtained by
 6 minimizing the cost function:

$$\begin{aligned} \mathcal{J}(\mathbf{c}_0) = & \frac{1}{2}(\mathbf{c}_0 - \mathbf{c}_0^b)^T \mathbf{B}_0^{-1}(\mathbf{c}_0 - \mathbf{c}_0^b) \\ & + \frac{1}{2} \sum_{k=1}^F (\mathcal{H}(\mathbf{c}_k) - \mathbf{c}_k^{obs})^T \mathbf{R}_k^{-1} (\mathcal{H}(\mathbf{c}_k) - \mathbf{c}_k^{obs}). \end{aligned} \quad (1)$$

7 Before data assimilation is performed, the current state that best estimates the true state is
 8 called a priori or background state \mathbf{c}_0^b . The random background errors are assumed to be
 9 unbiased and to have a normal distribution and \mathbf{B}_0 refers to the background error covariance
 10 (BEC). The observed value at time k is \mathbf{c}_k^{obs} . In general, the observational data are not
 11 accurately represented at the model grids. Additionally, in some cases, the observation
 12 instruments do not measure the meteorological variables directly (e.g., weather radar and
 13 satellite). Therefore, an observation operator \mathcal{H} that converts a model space to an observation
 14 space is required. The observation error includes both measurement (instrument) error and
 15 representativeness error. The representativeness error occurs because of the error included in
 16 the observation operator itself and because the input data of \mathcal{H} is not exactly the true state.
 17 Similar to the background error, the observation error is assumed to be unbiased and have a
 18 normal distribution. It is independent of other observation times, and usually is assumed to be
 19 spatially uncorrelated. Under this assumption, observation error covariance \mathbf{R}_k becomes a
 20 diagonal matrix. In addition, the observation error and background error are assumed to be
 21 independent of each other. The interpretation of this equation is that the deviation of initial
 22 concentration \mathbf{c}_0 from the background field \mathbf{c}_0^b is weighted by the inverse matrix of the
 23 background error covariance, whereas the differences between the model predictions $\mathcal{H}(\mathbf{c}_k)$
 24 and observations \mathbf{c}_k^{obs} during assimilation windows are weighted by the inverse of error
 25 observation covariance matrix.

26 The 4D-Var analysis can be obtained by the initial concentration that minimizes (1) with
 27 respect to the model equation.

$$\mathbf{c}_0^a = \arg \min \mathcal{J}(\mathbf{c}_0) \quad \text{subject to } \mathbf{c}_t = \mathcal{M}_{t_0 \rightarrow t}(\mathbf{c}_0),$$

$$t = 1, \dots, F \quad (2)$$

1 Here \mathcal{M} represents the model solution operator and includes an atmospheric forcing, the
 2 emission rates, the chemical kinetics, and all the other parameters. Furthermore, the model
 3 provides analysis within the assimilation window using the optimal initial conditions: $\mathbf{c}_t^a =$
 4 $\mathcal{M}_{t_0 \rightarrow t}(\mathbf{c}_0^a)$. Formally, a gradient-based optimization procedure is used to obtain minimum
 5 value. Assuming a linear observation operator $\mathbf{H}_k = \mathcal{H}'(\mathbf{c}_t)$, the gradient of (1) with respect
 6 to \mathbf{c}_0 is

$$\nabla_{\mathbf{c}_0} \mathcal{J}(\mathbf{c}_0) = \mathbf{B}_0^{-1}(\mathbf{c}_0 - \mathbf{c}_0^b) + \sum_{k=1}^F \left(\frac{\partial \mathbf{c}_k}{\partial \mathbf{c}_0} \right)^T \mathbf{H}_k^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{c}_k - \mathbf{c}_k^{obs}). \quad (3)$$

7 In the gradient of 4D-Var cost function, $(\partial \mathbf{c}_k / \partial \mathbf{c}_0)^T$ is a transposed derivative of future states
 8 with respect to the initial concentration. At this point, the adjoint model is used and through
 9 the solution of adjoint equation at t_0 , the gradient of the cost function at the initial
 10 concentration is provided. The gradient for the 4D-Var's cost function can be effectively
 11 obtained by forcing the adjoint model with observation increments and calculating it
 12 backwards. When the forward and reverse adjoint models are performed, i.e., \sum in the Eq. (3)
 13 is finished, it results in the problem of solving the following equation:

$$\nabla_{\mathbf{c}_0} \mathcal{J}(\mathbf{c}_0) = \mathbf{B}_0^{-1}(\mathbf{c}_0 - \mathbf{c}_0^b) + \boldsymbol{\lambda}_0 = 0 \quad (4)$$

14 $\boldsymbol{\lambda}_0$ is the sensitivity of the cost function (1) defined for 4D-Var with respect to the initial
 15 concentration \mathbf{c}_0 . Since \mathbf{B}_0^{-1} , \mathbf{c}_0^b , and $\boldsymbol{\lambda}_0$ values are known matrices and vectors, if the value of
 16 \mathbf{c}_0 that satisfies Equation (4) is found, it becomes the analysis field \mathbf{c}_0^a . Solving the above
 17 equation is similar to solving a linear-algebraic problem such as $\mathbf{A}\mathbf{x} = \mathbf{b}$, and the solution can
 18 be obtained by various minimization algorithms (e.g., steepest descent, conjugate gradient and
 19 quasi-Newton methods)

20 2.2 Background error covariance

21 Accurate error covariances for background and observation are important for the quality of
 22 data assimilation. A reasonable analysis may deteriorate because of misunderstanding of these
 23 covariances (Daescu 2008). The Background Error Covariance (BEC) is of utmost importance,
 24 as it weights the model error against the competing observation error, spreads information

1 from observations to the adjacent area, and influences several parameters such as temperature
2 and wind fields or chemical constituents. (Elbern and Schmidt, 2001)

3 The adjoint code for CMAQ (CMAQ-ADJ) model was implemented from the project H98
4 (University of Huston, 2009) by Huston Advanced Research Center / Texas Environ mental
5 Research Consortium (HARC/TERC). The validation and several numerical tests of this code
6 are well described in Hakami et al. (2007). Below is the defined cost function in CMAQ-ADJ
7 to optimize initial condition, which refers to concentration at the initial time.

$$J(\mathbf{c}_0) = \frac{1}{2(\sigma_0^b)^2} (\mathbf{c}_0 - \mathbf{c}_0^b)^T (\mathbf{c}_0 - \mathbf{c}_0^b) + \frac{1}{2(\sigma_k^{obs})^2} \sum_{k=1}^N (\mathbf{H}_k \mathbf{c}_k - \mathbf{c}_k^{obs})^T (\mathbf{H}_k \mathbf{c}_k - \mathbf{c}_k^{obs}) \quad (5)$$

8 This form only considers the model and observation errors as its variance, i.e. a constant value
9 of $(\sigma_0^b)^2$ and $(\sigma_k^{obs})^2$ with Gaussian distribution.

10 If a BEC is to be correctly adopted, a cost function should be defined in the form of a matrix;
11 this is denoted by the first term on the right hand side in eq. (1). The background part and its
12 gradient of the cost function, written in Fortran codes, have been revised in this study to make
13 the matrix operation possible. A numerical test is conducted to validate the suitability and
14 effects of the revised codes .

15 The methods for obtaining the BEC of a numerical model are mainly divided into two types: a
16 NMC method (Parrish and Derber, 1992) that defines the model error as the difference
17 between the forecasting results at different initial times, and an ensemble method that uses a
18 perturbed forecast. Recently, Kucukkaraca and Fisher (2003) introduced a technique for
19 modeling a flow-dependent BEC. In Constantinescu et al. (2007), an autoregressive model
20 was proposed for flow-dependent BEC in air quality data assimilation.

21 In this study, the BEC of the model is constructed by using the NMC method, which is the
22 most intuitive and easily applied method.

23 **3 Experimental design**

24 If the observatory sites are distributed unevenly, results of data assimilation based on the
25 variational theory will have low reliability, and it is difficult to minimize the cost function
26 (Courtier and Talagrand, 1987). For this reason, the capital region of South Korea is selected
27 for the present data assimilation study because measurement sites are relatively evenly
28 distributed in this area. Figure 1 depicts the study area (d03), i.e. the capital region of South

1 Korea along with the domain configuration for the other two nesting domains of coarse
2 resolution. A total of 120 observatory sites are evenly distributed in the areas of Seoul (SU),
3 Gyeonggi-do (GG), Gangwon-do (GW), Chungcheongnam-do (CN), and Chungcheongbuk-
4 do (CB). The innermost domain, d03, is located in a geographical area with coasts to the west
5 and the topography gradually rises towards the east. The Weather Research and Forecasting
6 (WRF) model (Skamarock et al., 2008) is a mesoscale atmospheric model that has been
7 widely used to simulate local circulation pattern and provide the meteorological input data for
8 air quality model. The chemical formation and transportation of ozone is simulated by the
9 Model-3 Community Multiscale Air Quality (CMAQ) model (Byun and Ching, 1999). This
10 model simulates gas-phase chemistry using the Carbon Bond IV (CB-IV) photochemical
11 mechanisms (Grey et al., 1989). To describe the chemical transformation, Euler Backward
12 Iterative (EBI) (Hertel et al., 1993) solver is implemented. The advection is calculated by the
13 Piecewise-Parabolic Method (PPM) (Colella and Woodward, 1984), which is based on the
14 finite volume subgrid definition of the advected scalar. The vertical diffusion in the planetary
15 boundary layer is calculated following the approach in the Regional Acid Deposition Model,
16 RADM (Chang et al., 1987), which is based on the similarity theory. Detailed settings used
17 for the atmospheric and air quality model systems in the present study are presented in Table
18 1 and Table 2, respectively. All the time mentioned in this paper except those in Table 2 are
19 local standard time (LST), which is 9 hours earlier than the Coordinated Universal Time
20 (UTC).

21 The experiment without assimilation was conducted as a forward run (FWD), which covers
22 four days from 09 LST on August 3 to 09 LST on August 7. In addition, data assimilation
23 (4DV) was performed within the 12hour time-window from 09 LST to 21 LST on August 5.
24 Figure 2 illustrates the spatial distribution of the total NO_x and VOCs pollutants, which are
25 out of the 24 emitted substances used in the CMAQ model. The domain d27 is located in the
26 East Asian monsoon region, which includes most of China and Japan. The Intercontinental
27 Chemical Transport Experiment-Phase B (INTEX-B, Zhang et al. (2009)) 2006 data were
28 used as emissions; high emissions are mostly found over major cities of each country. The
29 emissions applied to domains d09 and d03 are extracted from the CAPSS 2007 data (Lee et
30 al., 2011).

31 The results of the WRF simulation for the synoptic pattern of surface pressure during the
32 study period are presented in Figure 3, along with the weather charts. The vector indicates

1 surface wind, and the values of the contours are the concentrations of O₃. The model has
2 successfully simulated the North Pacific high-pressure system, and adequately describes the
3 local high-pressure system that developed in and around the East Sea on August 4, as well as
4 the high-pressure system that developed in and around the southwestern coastal region on
5 August 5. A clockwise synoptic flow developed because of the well-developed North Pacific
6 high-pressure system. As a result, the long-distance transport from the pollution sources in
7 China had little impact on the simulated pollutants.

8 Figure 4 shows the horizontal distributions of simulated ozone concentration and surface wind
9 from 06 LST to 21 LST on August 5 at three-hour intervals. At 06 LST, a southeasterly to
10 easterly wind developed along the western coast, and the overall ozone concentration was low
11 in this region. Accompanied with the increase in solar radiation after sunrise, the ozone
12 concentration began to increase, and an onshore sea-breeze developed after 12 LST in the
13 western coast. This sea-breeze lasted from 18 LST to 21 LST. After sunset, the influence of
14 the sea-breeze can be identified over areas where the ozone concentration decreased due to
15 NO_x-titration. Afterwards, the dominant wind direction changed in a clockwise direction
16 (figure omitted), and the local circulation did not extend far enough beyond the GG region.

17 **4 Results**

18 **4.1 Effects of an idealized BEC**

19 Two simple yet popular covariance models are Gaussian and Balgovind (Balgovind et al.,
20 1983) functions expressed as:

$$\omega(r) = EXP\left(-\frac{r^2}{2L^2}\right), \text{ Gaussian} \quad (6)$$

$$\omega(r) = \left(1 + \frac{r}{L}\right) EXP\left(-\frac{r}{L}\right), \text{ Balgovind} \quad (7)$$

21 To examine the appropriation of modified code, the Balgovind distribution expressed in Eq.
22 (7) is selected for constructing the BEC that has the components of matrix form. Figure 5
23 shows the distribution patterns for Gaussian and Balgovind with respect to the distance
24 between two grid points (r) and the characteristic length or radius of influence (L).

25 Table 3 summarizes a suite of numerical tests with and without data assimilation. In the tests
26 with application of data assimilation, a matrix is constructed assuming that the BEC of the
27 model has the form of a Balgovind function. The model domain is the innermost domain as

1 illustrated in Figure 1. The FWD test is conducted without data assimilation, and the other test
2 is performed with data assimilation. The two types of test are named as EXP_A and EXP_B,
3 respectively.

4 EXP_A is a test that can be used to evaluate the characteristics of the BEC based on a single
5 observation experiment. In this experiment, 100 ppb of O₃ was incorporated as an arbitrary
6 value rather than actual observation data at the initial time at the center of the model domain.
7 To emphatically show the background part of the cost function, the value 8.00, which is much
8 larger than the basic value (0.08), is applied to σ_k^{obs} in Equation (5). Using the function that
9 sets the radius of influence to be 2, 5, and 10, the data assimilation characteristics for three
10 BECs were examined.

11 In EXP_B, which is the second test, the effect of BEC used in 4D-Var is examined. Real
12 observation data is used in EXP_B. The observation data include 12 h ozone concentration at
13 120 sites within the capital city regions. Two cases are investigated in the EXP_B (Table 3):
14 the XBE case only considers variance that is not in a matrix form, and the OBE case uses the
15 BEC in the matrix form that adopts the Balgovind function. In the XBE, two tests that takes
16 into consideration the different weighting between σ_0^B and σ_k^{obs} are conducted separately. In
17 XBE_r0.08, the observation data is assumed to be accurate and σ_k^{obs} is set to 0.08, which is
18 the basic value for this model. For XBE_r8.00, σ_k^{obs} is set to 8.00, indicating that the results
19 of the model are more important than the observation. For OBE_r8.00, σ_k^{obs} and L are set to
20 8.00 and 5, respectively. The result of OBE_r0.08 is not analysed because it is similar to the
21 result of XBE_r0.08.

22 Among the results of the EXP_A, horizontal distributions of the analysis increment with
23 respect to the radius of influence (L) are illustrated in Figure 6. At the model grid point (29,
24 31), where arbitrary observation data were applied, all three tests showed an O₃ increment of
25 about 50.0 ppb. The background concentration of O₃ at the grid was 40.1 ppb, but the value
26 was up to about 90 ppb in the analysis when the synthetic observation of 100 ppb was applied.
27 However, as the value of L increased, the O₃ increment in the analysis occurs at more
28 surrounding grids. Particularly, the analysis increments shown along the east-west cross-
29 section (Figure 7) are distinguished on the 2D graph according to the L values. This result is
30 attributed to the ideal function that is used, in which the error covariance information is
31 expanded to the surrounding regions according to the L values. These results indicate that the

1 idealized BEC performs well in the revised codes, and proper analysis increments can be
2 achieved when the spatial correlation is taken into account.

3 Figure 8 shows the daily changes in ozone concentration simulated by each experiment in the
4 test EXP_B and from observations at selected site. Exact locations of these sites are marked in
5 Figure 1. At the site GG01, the observed (black solid line) concentration of ozone, which is
6 higher than 100 ppb, was not simulated in the FWD (blue solid line). In XBE_r0.08 (green
7 solid line), although the BEC is not applied, the simulated O₃ concentration is close to the
8 observation in almost all the time slots. Comparing results of the two experiments that applied
9 8.00 for σ_k^{obs} , the effect of BEC can be determined. In the case of XBE_r8.00 (red dotted line),
10 the simulated changes in O₃ concentration are similar to that simulated by the FWD because
11 the weighted value in the FWD is high. When the BEC is taken into consideration for the
12 same σ_k^{obs} (OBE_r8.00), the result is similar to that of XBE_r0.08. This result demonstrates
13 the effect of the spreading analysis increment to its surrounding region where the observation
14 sites are densely distributed. Although the weight of the observations is not set very high,
15 improvements in the field analysis by spatial correlation are still achieved. At the GG07 site,
16 this trend is quite significant with the OBE_r8.00 test, giving a result similar to that of
17 XBE_r0.08. At the GG60 position, the model results are significantly improved, but the
18 nighttime ozone is still over-estimated. However, at the GG28 site, which is located at a
19 region where observation sites are sparsely distributed, the BEC effect is barely observed. The
20 results of XBE_r8.00 are similar to those of OBE_r8.00, except after 18:00. This indicates
21 that the effect of BEC, which considers the spatial correlation, can be distinct mainly over
22 regions where the observation sites are densely distributed.

23 4.2 Development of realistic BEC

24 The BEC is obtained using the NMC (National Meteorological Center, now National Centers
25 for Environmental Prediction) approach (Parrish and Derber, 1992), which is based on a real
26 simulation for the realistic 4D-Var data assimilation study.

27 Figure 9 describes the method to define the model error. The error statistics for the CMAQ
28 model is defined by the differences between +48 hours and +24 hours forecast:

$$\epsilon^i = c_{+48h}^i - c_{+24h}^i. \quad (8)$$

1 The BEC matrix has 2,800,526,400 components for a 3-dimensional model with a number of
 2 grids $N_x * N_y * N_z = 60 * 63 * 14 = 52,920$. To avoid storing the error covariance matrix
 3 explicitly, we assume \mathbf{B} can be written as

$$\mathbf{B} = \mathbf{X} \otimes \mathbf{Y} \otimes \mathbf{Z} \otimes \mathbf{C}, \text{ (Chai et al., 2007)} \quad (9)$$

4 where, $\mathbf{X} = [N_x * N_x]$, $\mathbf{Y} = [N_y * N_y]$, and $\mathbf{Z} = [N_z * N_z]$, representing the error correlation
 5 in the three directions. \mathbf{C} is the error covariance matrix at a single grid point that refers to the
 6 error variances and correlation between different species. In this study, \mathbf{C} is considered to be
 7 constant, which means there is no correlation between the species.

8 It seems to be error-prone to invert ill-conditioned matrices. Based on the property of
 9 Kronecker product, \mathbf{B}^{-1} can be expressed as

$$\mathbf{B}^{-1} = (\mathbf{X} \otimes \mathbf{Y} \otimes \mathbf{Z})^{-1} = \mathbf{X}^{-1} \otimes \mathbf{Y}^{-1} \otimes \mathbf{Z}^{-1} \quad (10)$$

10 Singular Value Decomposition (SVD) is applied to \mathbf{B} matrix. For example, a general $m \times n$
 11 matrix \mathbf{A} can be written as

$$\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T. \quad (11)$$

12 For the symmetric matrices, such as $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$

$$\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{U}^T. \quad (12)$$

13 Then the inverse of \mathbf{A} is easily calculated:

$$\mathbf{A}^{-1} = \mathbf{U} \mathbf{\Sigma}^{-1} \mathbf{U}^T. \quad (13)$$

14 The accuracy of inverted BEC through these process has been confirmed by a algebraic
 15 calculation such as $\mathbf{B}^{-1} \mathbf{B} = \mathbf{I}$ and by comparing the vector \mathbf{x} between $\mathbf{B}\mathbf{x} = \mathbf{y}$ and $\mathbf{x} = \mathbf{B}^{-1}\mathbf{y}$.

16 The error correlations between the vertical layers of the model are given in Figure 10. Moving
 17 further away from a pertinent layer, the error correlation decreases. Judging from the
 18 diagonalized structure of errors, the correlation was found to be roughly a function of the
 19 physical distance between the layers. Examining the vertical error correlations for the
 20 magnitude of the boundary layer, high correlations can be found up to the fourth layer for the
 21 correlations in the vicinity of ground surface. This result indicates that an improvement in the
 22 model simulation can be achieved in the neighboring layers by performing DA using the
 23 observation data of upper layers that are located from the surface to the boundary layer.

1 In Figure 11, the error correlations are plotted as a function of distance between two layers.
2 When the distribution of correlations versus distance is fitted to a simple function, $e^{-\frac{\Delta z^{1.2}}{l_z^{1.2}}}$, the
3 vertical length scale is $l_z = 300$ m. Although some high values deviate from this function,
4 generally low correlation coefficients agree well with this function. The correlation
5 coefficients versus the horizontal distance are illustrated in Figure 12. On average, for both
6 the north-south and east-west directions, l_h is identified to be 10 km, and a function $e^{-\frac{\Delta z^{1.0}}{l_z^{1.0}}}$
7 fits well with the results. Particularly, the correlation coefficient for the east-west direction is
8 somewhat higher than that for the south-north direction. This is partly attributed to the effect
9 of middle latitude synoptic westerly and partly due to the land- sea breeze that occurs
10 frequently in August in the capital city region, which produces circulation in the east-west
11 direction.

12 4.3 Validation time results

13 4D-Var experiments are performed in this study, using actual observations with the
14 distribution of the initial concentration of O_3 as the control variable. The observed hourly O_3
15 concentrations at 120 sites located within the domain d03 are used. In formula (1), \mathbf{c}_0 of
16 ozone is considered as the control variable, and the BEC established in 4.2 is applied as the
17 model error (\mathbf{B}_0^{-1}). The representativeness error is not considered, because the observatory
18 sites are manually placed on grids close to the measurement sites. The observation error \mathbf{R}_k^{-1}
19 is a diagonal matrix that has same diagonal components, which is 1% of the observed
20 concentration.

21 The observation results of the diurnal variation of O_3 at several sites during the 12-hour time-
22 window are shown in Figure 13, along with results of the FWD and 4DV experiments. The
23 sites are selected in accordance with the administrative districts as shown in in Figure 1. The
24 daytime high concentrations of O_3 above 100 ppb are not well simulated in the FWD, whereas
25 they are captured in the 4DV experiment. At almost all the sites the high values of O_3
26 concentration simulated by the 4DV experiment are found to be close to the observational
27 values. Looking at the results of the FWD, it is found that the ozone concentration at GW04
28 and CB06 is above 80 ppb at 09 LST, while the 4DV significantly reduces the errors in the
29 initial condition. However, 4DV cannot properly simulate the high concentrations of O_3 in the
30 early afternoon at some sites, for example at the site GG76, and the high concentration of O_3

1 at SU21 remains underestimated. These problems are caused by uncertainties in ozone
2 precursors that exist in both the initial conditions and in the emissions. This can probably be
3 solved by changing the control variables and optimizing the amounts of emissions and by
4 improving initial concentrations of the pollutants. In addition, the accuracy of the simulation
5 for the ozone concentration in Incheon areas is directly affected by the pollutants coming
6 from the Yellow Sea. Hence it is necessary to optimize the boundary data.

메모 포함[p2]: Response to the comment #3 of reviewer 2

7 The Root Mean Square Error (RMSE) and Index Of Agreement (IOA) of simulated results at
8 each iteration step of 4D-Var using observation data from all sites were calculated, and the
9 results are shown in Figure 14 (the definitions of statistical variables used in this research are
10 listed in Table 4). Results at the starting point, i.e. iteration=1, is the statistical results of the
11 FWD results, where RMSE and IOA are 35.1 ppb and 0.576, respectively. After
12 approximately 20 iterations, RMSE decreases to 20 ppb or less, and IOA increases to 0.9 or
13 more. Thereafter, there are little changes in these statistical variables, implying that the results
14 of 4DV have converged.

15 Figure 15 gives the diurnal variations of the two statistical variables. As the statistical results
16 are derived from 120 observatory sites over a fixed period of time, they actually represent the
17 errors and general agreement in spatial distribution of O₃ concentration. The FWD results
18 show a decrease in RMSE and an increase in IOA until 11 LST, but a rapid increase in RMSE
19 and a decrease in IOA occur after 11 LST. This is caused by the inaccurate simulation of high
20 ozone concentrations during the daytime. The value of RMSE then decreases again after 16
21 LST, but large errors of O₃ concentration up to 30 ppb or more are still evident. In contrast, in
22 the 4DV results, the RMSE and IOA for the initial concentration of O₃ are 2.9 ppb and 0.954
23 respectively, suggesting that the errors in the initial state are significantly reduced. Afterwards,
24 IOA continues to decrease and reach the value of 0.543 at 21 LST, but this value is still higher
25 than that in the FWD result (i.e., 0.363). The value of RMSE increases at the beginning 2 h
26 and is close to the FWD result, but it never becomes larger than 20 ppb thereafter. In
27 particular, the RMSE shows the maximum decrease of 27.4 ppb at 16 LST, which means that
28 the accuracy of the simulation for high daytime ozone concentration has been substantially
29 improved.

30 Table 5 shows the statistical results based on simulations with the 12-hour assimilation
31 periods and from the 120 observatory sites. The simulation result of the 4DV experiment is
32 61.4 ppb, which is close to the average concentration of observed ozone of 63.6 ppb. A 49.4%

1 decrease in RMSE and a 59.9% increase in IOA in the results of the 4DV (i.e., the difference
2 between FWD and 4DV) demonstrate the great improvement caused by data assimilation.
3 Mean Bias, normalized by the average observed concentration (MMB), was -21.2% in FWD,
4 and -3.4% in 4DV. This result of NMB implies that the tendency to underestimate daytime
5 ozone is mitigated by application of data assimilation.

6 To compare the spatial distribution of the simulated O₃ with that of the observed
7 concentrations, the 4DV results are presented in Figure 16. The concentrations of O₃ at
8 observatory sites are indicated with colored circles using the same color scales as the contours.
9 At 09 LST, 4DV shows a homogeneous distribution, with concentrations of O₃ in and around
10 Seoul to be almost zero. However, in eastern GG, GW, and CB, where the observatory sites
11 are sparsely distributed, the concentration of O₃ decreases to zero only near the observatory
12 sites. For the high concentration of ozone, i.e. 100 ppb or higher, which appears at 15 LST,
13 the FWD results are approximately 50–60 ppb in Seoul (Figure 4), and the 4DV results are
14 consistent with the observed concentrations. However, at 18 LST, the difference between
15 FWD and 4DV results grows more remarkable. Low ozone concentration appears even in
16 central Seoul and in southeastern GG in the FWD concentration simulation at 21 LST, which
17 is attributed to excessive NO_x-titration. However, for the 4DV results, the distribution of O₃
18 concentration in Seoul areas shows a pattern similar to that of the observations.

19 Figure 17 shows the difference between results of FWD and 4DV (4DV results minus that of
20 the FWD). These differences can be regarded as analysis increments and their effects during
21 assimilation windows. At 09 LST, the analysis increments are negative in most of the area,
22 but are positive over some of the western coast area and the CN area, which is affected by the
23 clockwise circulation of the sea-breeze. These analysis increments, which are also evident in
24 the result of the reanalysis of initial conditions, are transported to inland areas by the local
25 circulation. As a result, the differences between the FWD and 4DV experiments become
26 larger, and the areas of positive values become larger too, encompassing the SU and GG areas.
27 This process makes it possible to simulate the high concentration of daytime ozone.

28 **4.4 Predictability of ozone**

29 The direct comparison with the observation data used during the assimilation window has a
30 limit in the verification of results. Forecasts of FWD and 4DV with different initial conditions
31 after the time-window (Table 2) are performed in this part. Figure 18 (a) depicts the temporal

1 variation of ozone concentration, which is obtained by averaging the results of all the
2 observatory sites and those of corresponding model grids during the 12-hour assimilation
3 period and the 12-hour forecast. During the period for validation, the FWD overestimates O₃
4 in the morning and underestimates it after 12 LST while the 4DV shows a tendency that
5 almost conforms to that of the observations. The forecast is initialized at 21 LST, on August 5,
6 and run for 24 h. The results of the first 12 h are plotted in the figure. Both experiments show
7 a tendency to forecast high levels of nighttime ozone. However, while the FWD shows a
8 rising tendency after 21 LST, the 4DV gives a declining ozone tendency and therefore
9 provides a better forecast than the FWD. Figure 19 (b) indicates the reduced forecast errors in
10 the results of the 4DV, along with the time variations of statistical variables, for the forecast
11 period. At 21 LST, the 4DV error is only 19.8 ppb, much smaller than that of the FWD. This
12 is attributed to the initial condition that is 10.0 ppb less than that of FWD. After 21 LST, the
13 effect of improved initial condition diminishes gradually, although the RMSE in the 4DV
14 results is still smaller than that in the FWD results. To quantitatively evaluate the overall
15 improved predictability, the ratio of the reduced RMSE in the 4DV to that in the FWD
16 experiments is calculated. Results indicate that the ratio is 8% for the +24 hours, and 13% for
17 the +12 hours. This improvement in the forecast accuracy is achieved solely by using the
18 assimilated initial condition, and more improvements are therefore expected by further
19 optimizing the amount of parameters such as emissions and boundary conditions.

20 The above result shows a forecast for the nighttime ozone with application of the daytime
21 data assimilation. However, high concentrations of ozone that have harmful effects to human
22 health are often found during daytime. Therefore, the effects of the assimilation over a time-
23 window in the nighttime upon the forecast accuracy of daytime ozone concentration are also
24 carried out. The period for validation of data assimilation is set to be 12 h, from 12 UTC on
25 August 5 to 00 UTC on August 6 (Table 2). The +12 h forecast period for 4DV in Figure 18
26 (a) corresponds to that of the FWD during the validation period in Figure 18 (b). In the results
27 with assimilation of nighttime ozone, the estimated ozone concentration approaches that of
28 the observation, and the variation tendency conforms to the observation. In the ensuing
29 forecast period, both of the experiments show a diurnal variation in the simulated ozone, but
30 the FWD results demonstrate deviations from the observation, which are caused by the
31 overestimated initial concentration at 09 LST. In the morning, the maximum reduced RMSE
32 (Figure 19 (b)) is 13.6 ppb, and all the reductions of RMSEs are more than 10.0 ppb. After 09
33 LST, the value of the reduced RMSE decreases. The improvement in forecast accuracy,

메모 포함[p3]: Response to the comment #3 of reviewer 2

1 obtained by calculating the ratio of reduced errors, is 11% for +24 h, and 17% for +12 h,
2 indicating that the improvement achieved by the nighttime assimilation is higher than that by
3 the daytime assimilation. However, the effects of the improved initial condition by 4D-Var in
4 the daytime ozone forecast cannot last for more than 12 h.

5 Optimized ozone after data assimilation didn't show a significant change in the other
6 chemical components (not shown here). Ozone is a secondary produced pollutant, and has no
7 direct emission sources. Other components, especially, the precursors of ozone, are mostly
8 dependent on its emission information. Our next study will be optimizing the initial condition
9 for NO_x and VOCs to improve the predictability of O₃. If the multivariate background error
10 covariance is well established, this optimization will be achieved although the control variable
11 is different from the observed variables.

메모 포함 [p4]: Response to the comment #2 of reviewer 2

12 5 Conclusions

13 In this study, we presented an approach that uses an adjoint model in data assimilation. To
14 incorporate observation data in a numerical model, the 4D-Var that is designed to improve
15 predictability of ozone concentration is conducted by optimization of the initial values. The
16 model systems used in the present study includes WRF, CMAQ and CMAQ-ADJ.

17 The previously developed adjoin code for 4D-Var considers the background error of the
18 model in the cost function as a constant. In this study, the code is revised to reflect the
19 information of errors belonging to the actual subject areas. Verification of the revised code are
20 conducted. Two numerical experiments are first performed by defining an ideal matrix with
21 the assumption that the background error has a Balgovind function distribution. The results
22 are verified. It is found that synthetic observation information are effectively spread over the
23 neighboring areas.

24 In order to define the realistic model error, the NMC method that is widely used in
25 meteorological DA is adopted in this study. The background error covariance is constructed
26 based on the 29 differences between 48h forecasts and 24h forecasts, which are taken as the
27 model error. The forecasts are performed over August, with daily initialization and a forecast
28 period of 48-hour. . The vertical correlation of the model results is constructed as a diagonal
29 and symmetric matrix; the length scale in the correlation analysis of vertical distance is about
30 300 m, and the scale of length in the averaged east-west and south-north correlation is about
31 10 km (the east-west correlation is higher than the north-south correlation).

1 The generated background error of the model simulation is applied in the 4D-Var research,
2 and the surface observation is incorporated by DA to optimize the initial concentration of
3 ozone. As a result of DA in a 12-h time-window during the daytime of August 5, the 4DV
4 experiment shows a diurnal variation of O_3 concentration that conforms well to the
5 observation, while the experiment without DA (FWD) either overestimates or underestimates
6 the O_3 concentration. In the statistical result, the RMSE decreases by about 49.4%, and the
7 IOA increases by 59.9%, suggesting that the initial conditions of ozone concentration are
8 successfully improved by application of DA. The analysis increments, which are the extents
9 of improvement of the initial conditions, spread along the route of the sea breeze that blows in
10 from Incheon during the daytime and blows out during the evening, causing an improvement
11 in the statistical results for the calculation area over 12 h. In addition, a potential improvement
12 for the ozone predictability is presented using the optimized initial condition after the time-
13 window. In particular, a larger improvement in the predictability of daytime ozone
14 concentration is expected if DA is performed over the nighttime than in the daytime.

15 Data assimilation has been playing an essential role in air quality modelling study. For this
16 reason, the following studies need to be conducted for further operational applications of data
17 assimilation.

- 18 1. In addition to ground data, other observations such as the data from ozone sonde,
19 airplanes, and satellites, need to be exploited.
- 20 2. In the case of long-range transport, the inflow boundary condition needs to be
21 optimized by considering it as a control variable in 4D-Var data assimilation.
- 22 3. Instead of using the averaged values of BEC data (which is used in the present
23 research) to easily obtain the inverse matrix, the error correlation with different
24 length scales at each grid should be considered. For this purpose, the preconditioning
25 procedure, which modifies the form of the cost function, should be applied.
- 26 4. When considering the error covariance used in the modelling study, it is possible to
27 conduct DA research using observation variables that are different to the control
28 variables.

29 The study proposes a method to improve predictability by applying DA technology to air
30 quality forecasts. Results of the present study provide helpful information to policy makers in
31 charge of emission regulation. With more information related to a variety of air pollutants

1 become available in the future, for example data from the geostationary orbit environmental
2 satellite that is planned to operate in 2018 (Lee et al., 2009) and other observation systems, it
3 is necessary to handle vast amount of observation data for better chemical weather forecasting
4 (Carmichael et al., 2008). This study can be considered to be a preliminary research in this
5 aspect.

6

7 **Acknowledgements**

8

1 **References**

- 2 Balgovind, R., Dalcher, A., Ghil, M., and Kalnay, E.: A Stochastic-Dynamic Model for the
3 Spatial Structure of Forecast Error Statistics, *Monthly Weather Review*, 111, 701-722, Doi
4 10.1175/1520-0493(1983)111<0701:Asdmft>2.0.Co;2, 1983.
- 5 Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Žabkar, R., Carmichael, G. R., Flemming, J.,
6 Inness, A., Pagowski, M., Pérez Camaño, J. L., Saide, P. E., San Jose, R., Sofiev, M., Vira, J.,
7 Baklanov, A., Carnevale, C., Grell, G., and Seigneur, C.: Data assimilation in atmospheric
8 chemistry models: current status and future prospects for coupled chemistry meteorology
9 models, *Atmospheric Chemistry and Physics*, 15, 5325-5358, 10.5194/acp-15-5325-2015,
10 2015.
- 11 Boisgontier, H., Mallet, V., Berroir, J. P., Bocquet, M., Herlin, I., and Sportisse, B.: Satellite
12 data assimilation for air quality forecast, *Simulation Modelling Practice and Theory*, 16,
13 1541-1545, 10.1016/j.simpat.2008.01.008, 2008.
- 14 Byun, D. W. and Ching, J. K. S.: Science algorithms of the EPA models-3 Community
15 Multiscale Air Quality (CMAQ) modeling system, EPA/600/R-99/030, US EPA, Research
16 Triangle Park, USA, 1999.
- 17 Carmichael, G. R., Sandu, A., Chai, T., Daescu, D. N., Constantinescu, E. M., and Tang, Y.:
18 Predicting air quality: Improvements through advanced methods to integrate models and
19 measurements, *Journal of Computational Physics*, 227, 3540-3571, 10.1016/j.jcp.2007.02.024,
20 2008.
- 21 Chai, T., Carmichael, G. R., Tang, Y., Sandu, A., Hardesty, M., Pilewskie, P., Whitlow, S.,
22 Browell, E. V., Avery, M. A., Nédélec, P., Merrill, J. T., Thompson, A. M., and Williams, E.:
23 Four-dimensional data assimilation experiments with International Consortium for
24 Atmospheric Research on Transport and Transformation ozone measurements, *Journal of*
25 *Geophysical Research*, 112, D12S15, 10.1029/2006jd007763, 2007.
- 26 Constantinescu, E. M., Chai, T., Sandu, A., and Carmichael, G. R.: Autoregressive models of
27 background errors for chemical data assimilation, *Journal of Geophysical Research*, 112,
28 D12309, 10.1029/2006jd008103, 2007.
- 29 Courtier, P., and Talagrand, O.: Variational Assimilation of Meteorological Observations
30 With the Adjoint Vorticity Equation. Ii: Numerical Results, *Quarterly Journal of the Royal*
31 *Meteorological Society*, 113, 1329-1347, 10.1002/qj.49711347813, 1987.

1 Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljevic, D., Hamrud, M.,
2 Hollingsworth, A., Rabier, E., and Fisher, M.: The ECMWF implementation of three-
3 dimensional variational assimilation (3D-Var). I: Formulation, Quarterly Journal of the Royal
4 Meteorological Society, 124, 1783-1807, DOI 10.1002/qj.49712455002, 1998.

5 Daescu, D. N.: On the Sensitivity Equations of Four-Dimensional Variational (4D-Var) Data
6 Assimilation, Monthly Weather Review, 136, 3050-3065, 10.1175/2007mwr2382.1, 2008.

7 Daley, R.: Atmospheric Data Analysis, Cambridge University Press, Cambridge, UK, 1991.

8 Elbern, H., Schmidt, H., and Ebel, A.: Variational data assimilation for troospheric chemistry
9 modeling, Journal of Geophysical Research, 102, 15,967-915,985, 1997.

10 Elbern, H., and Schmidt, H.: Ozone episode analysis by four-dimensional variational
11 chemistry data assimilation, Journal of Geophysical Research, 106, 3569-3590, Doi
12 10.1029/2000jd900448, 2001.

13 Evensen, G.: Data Assimilation: The Ensemble Kalman Filter, Springer Berlin Heidelberg,
14 2009.

15 Gou, T., and Sandu, A.: Continuous versus discrete advection adjoints in chemical data
16 assimilation with CMAQ, Atmospheric Environment, 45, 4868-4881,
17 10.1016/j.atmosenv.2011.06.015, 2011.

18 Hakami, A., Henze, D. K., Seinfeld, J. H., Singh, K., Sandu, A., Kim, S., Byun, D., and Li, Q.:
19 The Adjoint of CMAQ, Environ Sci Technol, 41, 7807-7817, 10.1021/es070944p, 2007.

20 Huang, M., Carmichael, G. R., Chai, T., Pierce, R. B., Oltmans, S. J., Jaffe, D. A., Bowman,
21 K. W., Kaduwela, A., Cai, C., Spak, S. N., Weinheimer, A. J., Huey, L. G., and Diskin, G. S.:
22 Impacts of transported background pollutants on summertime western US air quality: model
23 evaluation, sensitivity analysis and data assimilation, Atmospheric Chemistry and Physics, 13,
24 359-391, 10.5194/acp-13-359-2013, 2013.

25 Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, Cambridge
26 University Press, Cambridge, UK, 2003.

27 Kucukkaraca, E. and Fisher, M.: Use of Analysis Ensembles in Estimating Flow-dependent
28 Background Error Variances, European Centre for Medium-Range Weather Forecasts,
29 ECMWF technical memorandum, 429, 2006.

1 Le-dimet, F. X., and Talagrand, O.: Variational algorithms for analysis and assimilation of
2 meteorological observations, *Tellus*, 38A, 97-110, 1986.

3 Lee, D.-G., Lee, Y.-M., Jang, K.-W., Yoo, C., Kang, K.-H., Lee, J.-H., Jung, S.-W., Park, J.-
4 M., Lee, S.-B., Han, J.-S., Hong, J.-H., and Lee, S.-J.: Korean National Emissions Inventory
5 System and 2007 Air Pollutant Emissions, *Asian Journal of Atmospheric Environment*, 5,
6 278-291, 10.5572/ajae.2011.5.4.278, 2011.

7 Navon, I.: Data Assimilation for Numerical Weather Prediction: A Review, in: *Data
8 Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, edited by: Park, S., and
9 Xu, L., Springer Berlin Heidelberg, 21-65, 2009.

10 Parrish, D. F., and Derber, J. C.: The National Meteorological Center's Spectral Statistical-
11 Interpolation Analysis System, *Monthly Weather Review*, 120, 1747-1763, 10.1175/1520-
12 0493(1992)120<1747:TNMCS>2.0.CO;2, 1992.

13 Penenko, V. V. a. O., N. N.: A variational initialization method for the fields of meteorologica
14 l elements, *Soviet Meteor. Hydrol.*, 11, 1-11, 1976.

15 Penenko, V., Baklanov, A., and Tsvetova, E.: Methods of sensitivity theory and inverse mode
16 ling for estimation of source parameters, *Future Generation Computer Systems*, 18, 661-671,
17 [http://dx.doi.org/10.1016/S0167-739X\(02\)00031-6](http://dx.doi.org/10.1016/S0167-739X(02)00031-6), 2002.s

18 Rabier, F., Jarvinen, H., Klinker, E., Mahfouf, J. F., and Simmons, A.: The ECMWF
19 operational implementation of four-dimensional variational assimilation. I: Experimental
20 results with simplified physics, *Quarterly Journal of the Royal Meteorological Society*, 126,
21 1143-1170, Doi 10.1256/Smsqj.56414, 2000.

22 Sandu, A., Daescu, D. N., Carmichael, G. R., and Chai, T.: Adjoint sensitivity analysis of
23 regional air quality models, *Journal of Computational Physics*, 204, 222-252,
24 10.1016/j.jcp.2004.10.011, 2005.

25 Sandu, A., and Chai, T.: Chemical Data Assimilation—An Overview, *Atmosphere*, 2, 426-
26 463, 10.3390/atmos2030426, 2011.

27 Silver, J. D., Brandt, J., Hvidberg, M., Frydendall, J., and Christensen, J. H.: Assimilation of
28 OMI NO₂ retrievals into the limited-area chemistry-transport model DEHM
29 (V2009.0) with a 3-D OI algorithm, *Geoscientific Model Development*, 6, 1-16,
30 10.5194/gmd-6-1-2013, 2013.

1 Singh, K., Jardak, M., Sandu, A., Bowman, K., Lee, M., and Jones, D.: Construction of non-
2 diagonal background error covariance matrices for global chemical data assimilation,
3 Geoscientific Model Development, 4, 299-316, 10.5194/gmd-4-299-2011, 2011.

4 Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang,
5 X.-Y., Wang, W., and Powers, J. G.: A Description of the Advanced Research WRF Version
6 3, National Center for Atmospheric Research, Boulder, Colorado, USA, 2008.

7 Talagrand, O., and Courtier, P.: Variational Assimilation of Meteorological Observations
8 With the Adjoint Vorticity Equation. I: Theory, Quarterly Journal of the Royal
9 Meteorological Society, 113, 1311-1328, 10.1002/qj.49711347812, 1987.

10 University of Houston: Air Quality Modeling of TexAQS-II Episodes with Data Assimilation,
11 TERC Project H98, Final Report, Houston Advanced Research Center (HARC), 2009.

12 Wang, K. Y., Lary, D. J., Shallcross, D. E., Hall, S. M., and Pyle, J. A.: A review on the use
13 of the adjoint method in four-dimensional atmospheric-chemistry data assimilation, Quarterly
14 Journal of the Royal Meteorological Society, 127, 2181-2204, Doi 10.1256/Smsqj.57615,
15 2001.

16 Zhang, L., Constantinescu, E. M., Sandu, A., Tang, Y., Chai, T., Carmichael, G. R., Byun, D.,
17 and Olaguer, E.: An adjoint sensitivity analysis and 4D-Var data assimilation study of Texas
18 air quality, Atmospheric Environment, 42, 5787-5804, 10.1016/j.atmosenv.2008.03.048, 2008.

19 Zhang, Q., Streets, D. G., Carmichael, G. R., He, K. B., Huo, H., Kannari, A., Klimont, Z.,
20 Park, I. S., Reddy, S., Fu, J. S., Chen, D., Duan, L., Lei, Y., Wang, L. T., and Yao, Z. L.:
21 Asian emissions in 2006 for the NASA INTEX-B mission, Atmospheric Chemistry and
22 Physics, 9, 5131-5153, 2009.

23

1

2 Table 1. Configuration of WRF modeling system

WRF	d27	d09	d03
Horizontal Grid	123 × 130	72 × 84	65 × 68
Horizontal resolution	27 km	9 km	3 km
Vertical layers	33 layers (top: 50hPa)		
Physical options	WSM5 scheme Kain-Fritsch scheme Noah LSM Yonsei University PBL RRTM Longwave Dudhia Shortwave		
Initial data	NCEP FNL data		
Time period	00 UTC 03 August ~ 00 UTC 07 August, 2008		

3

4 Table 2. Configuration of CMAQ 4D-Var modeling system

CMAQ	d27	d09	d03
Meteorological input	correspond to each WRF domain		
Horizontal Grid	118 × 125	67 × 79	60 × 63
Horizontal resolution	27 km	9 km	3 km
Vertical layers	15 layers (top: 20 km)		
Other options	CB IV Chemical Mechanism PPM Advection Multiscale Horizontal Diffusion Eddy Vertical Diffusion RADM Cloud scheme		
Emission data	INTEX-B	CAPSS	CAPSS
Forward	00 UTC 03 ~ 00 UTC 07 August, 2008 (4 days)		
Time periods	4D-Var	day time assimilation	00 UTC 05 ~ 12 UTC 05 August, 2008 (12 hours, analysis)
		night time assimilation	12 UTC 05 ~ 12 UTC 06 August, 2008 (24 hours, forecast)
		night time assimilation	12 UTC 05 ~ 00 UTC 06 August, 2008 (12 hours, analysis)
		night time assimilation	00 UTC 06 ~ 00 UTC 07 August, 2008 (24 hours, forecast)

5

6

7

8

9

1

2

3 Table 3. Experimental design for the idealized background error covariance test. The FWD
4 case is conducted and the results are compared with that of the 4D-Var run.

Assimilation	Case	Observation data	Radius of Influence	σ_0^B	σ_k^{obs}	
Forward run	FWD	n/a	n/a	n/a	n/a	
4D-Var run	EXP_A (single obs.)	L02	100 ppb at (29,31)	L=02	BEC	8.00
		L05		L=05	BEC	8.00
		L10		L=10	BEC	8.00
	EXP_B	XBE_r0.08	12 hours O ₃ at all 120 sites	n/a	1.00	0.08
		XBE_r8.00		n/a	1.00	8.00
		OBE_r8.00		L=05	BEC	8.00

5

6 Table 4. Statistics of the model results.

Description	Variable	Statistic definition*
Mean obs.	\bar{O}	$(1/N) \sum_{i=1}^N O_i$
Mean model	\bar{M}	$(1/N) \sum_{i=1}^N M_i$
Mean Bias	MB	$(1/N) \sum_{i=1}^N (M_i - O_i)$
Normalized Mean Bias	NMB(%)	$(1/N) \sum_{i=1}^N (M_i - O_i) / \bar{O} \times 100$
Root Mean Square Error	RMSE	$\sqrt{(1/N) \sum_{i=1}^N (M_i - O_i)^2}$
Index Of Agreement	IOA	$1 - \frac{\sum_{i=1}^N (M_i - O_i)^2}{\sum_{i=1}^N (M_i - \bar{O} + O_i - \bar{O})^2}$

7 *(M = modelled, O = observed)

8

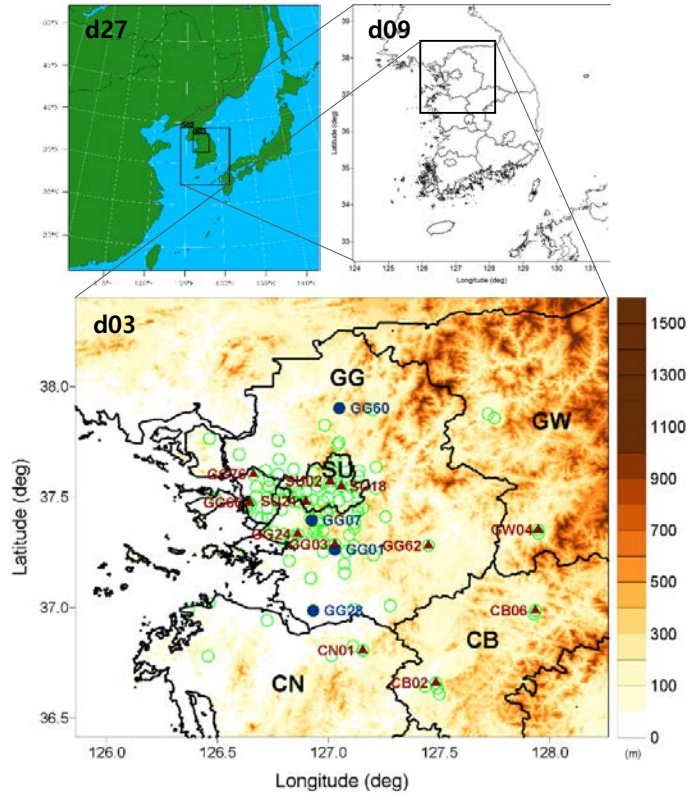
9

10

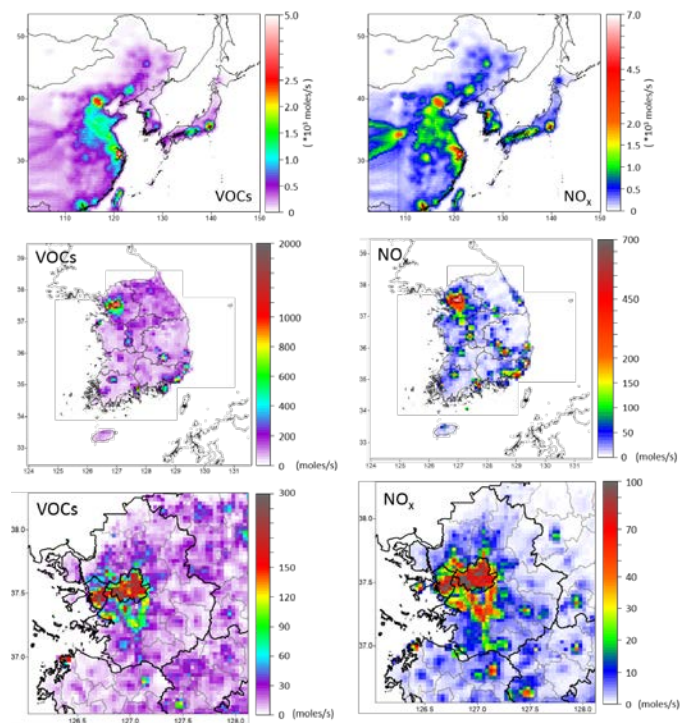
1
2
3
4
5
6
7
8

Table 5. Statistics for the observed (OBS) and simulated (FWD and 4DV) results. The FWD indicates the simulation without data assimilation. 4DV results are obtained by assimilating all observed surface O₃ with realized background error covariance matrix during 12h time-windows.

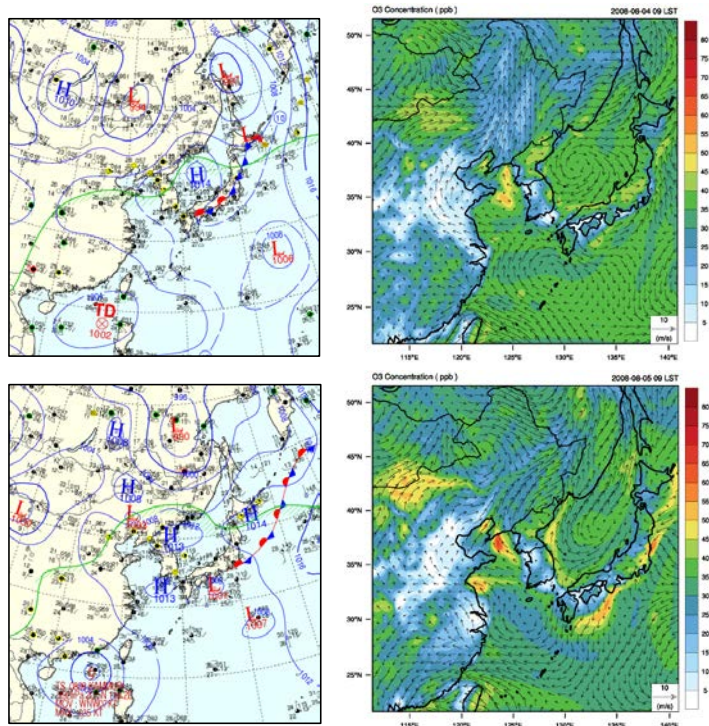
Statistics	FWD	4DV	OBS
Mean (ppb)	50.1	61.4	63.6
RMSE (ppb)	35.1	17.8	
IOA	0.576	0.921	
MB (ppb)	-13.5	-2.1	
NMB (%)	-21.2	-3.4	



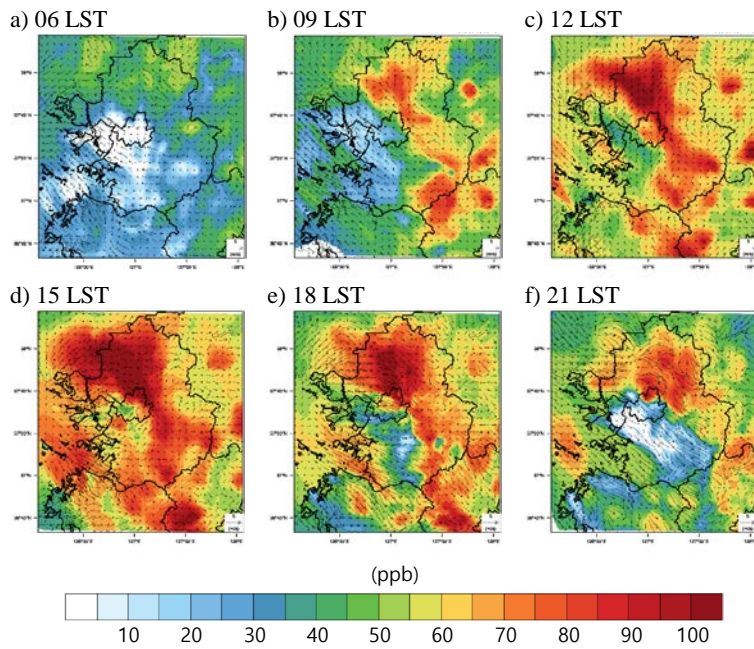
1
 2 Figure 1. The model domains (d27, d09, and d03) for WRF. The domain size of CMAQ is
 3 mostly the same except that it has five grids fewer than WRF at lateral boundaries. The air
 4 quality monitoring sites at ground level are marked by green blank circles. Blue filled circles
 5 and red filled triangles indicate the selected locations for the idealized and realized
 6 background error covariance experiments, respectively. These experiments are conducted to
 7 investigate the diurnal variation of ozone during the assimilation window. Administrative
 8 district in the areas of Seoul, Gyeonggi-do, Gangwon-do, Chungcheongnam-do, and
 9 Chungcheongbuk-do is abbreviated to SU, GG, GW, CN, and CB, respectively, and also
 10 represented on the map.
 11



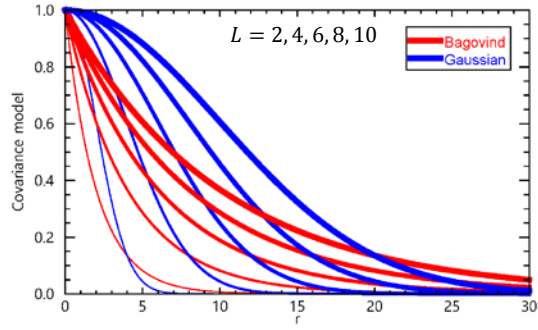
1
 2 Figure 2. Horizontal distributions of emission rate for domain d27 (top), d09 (middle), and
 3 d03 (bottom). The left and right panels are for VOCs and NO_x emission rates, respectively.
 4



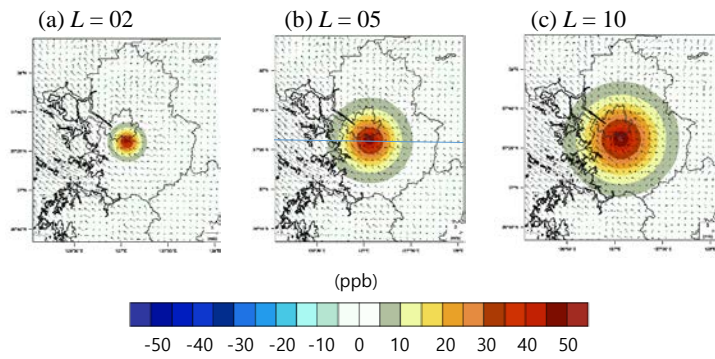
1 Figure 3. Synoptic weather charts (left) and simulated results (right) on 04 (upper) and 05
 2 (lower) August. Filled contours and vectors represent ozone concentration and winds,
 3 respectively
 4



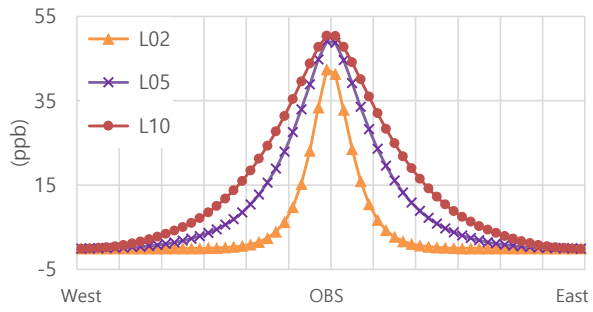
1 Figure 4. Diurnal variations of horizontal distribution of ozone (contour) and wind (vector) at
 2 3-hour interval starting from 06 LST on 5 August.
 3



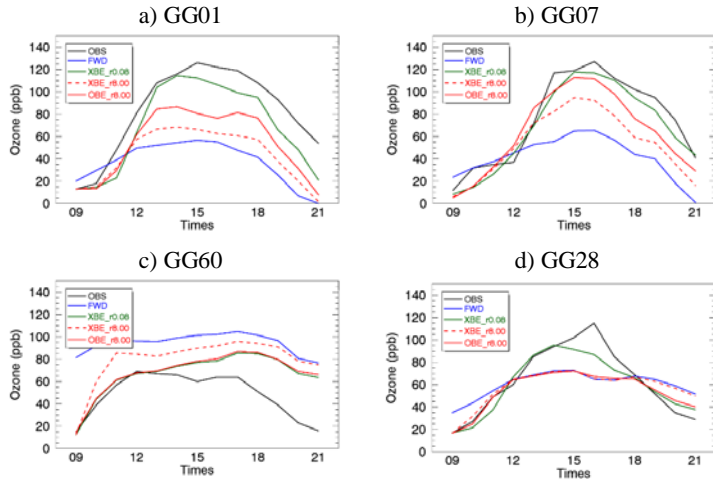
1
 2 Figure 5. Covariance distribution for Gaussian (blue) and Balgovind (red) functions with
 3 respect to the distance (r) and the values of radius of influence (L).
 4



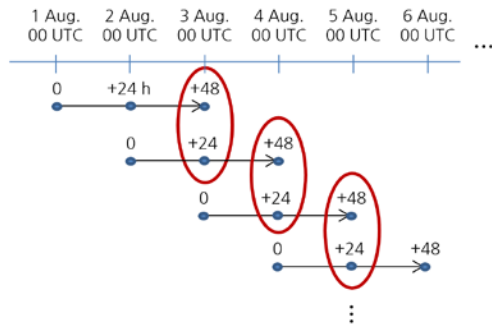
1 Figure 6. Horizontal distribution of analysis increments at surface resulted from the single
 2 observation experiment (EXP_A) with respect to radius of influence (L). Blue line on the (b)
 3 stands for the location where the cross-sectional values of analysis increments is examined.
 4



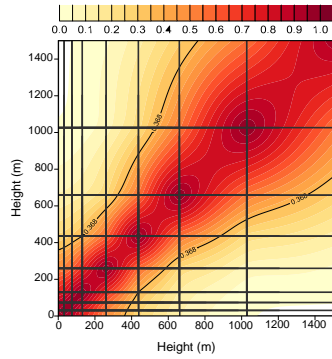
1
 2 Figure 7. Cross-section of analysis increments along the blue line in Figure 6. (b) as the radius
 3 of influence (L) values are increase.
 4



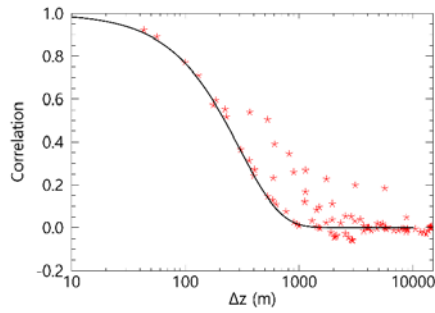
1 Figure 8. Diurnal variations of surface ozone from the results of EXP_B at a) GG01, b) GG07,
 2 c) GG60, and d) GG28. Black and blue solid lines indicate observation (OBS) and results of
 3 forward run (FWD), respectively. XBE_r0.08 (green solid), XBE_r8.00 (red dashed), and
 4 OBE_r8.00 (red solid) represent 4D-Var run results with and without considering the
 5 background error in matrix form where the observation error (σ_k^{obs}) is 0.08 and 8.00.
 6



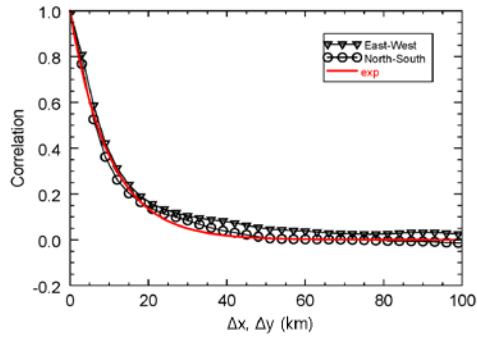
- 1
- 2 Figure 9. Schematic illustration for the NMC approach to obtain the background error
- 3 covariance (BEC) matrix.
- 4



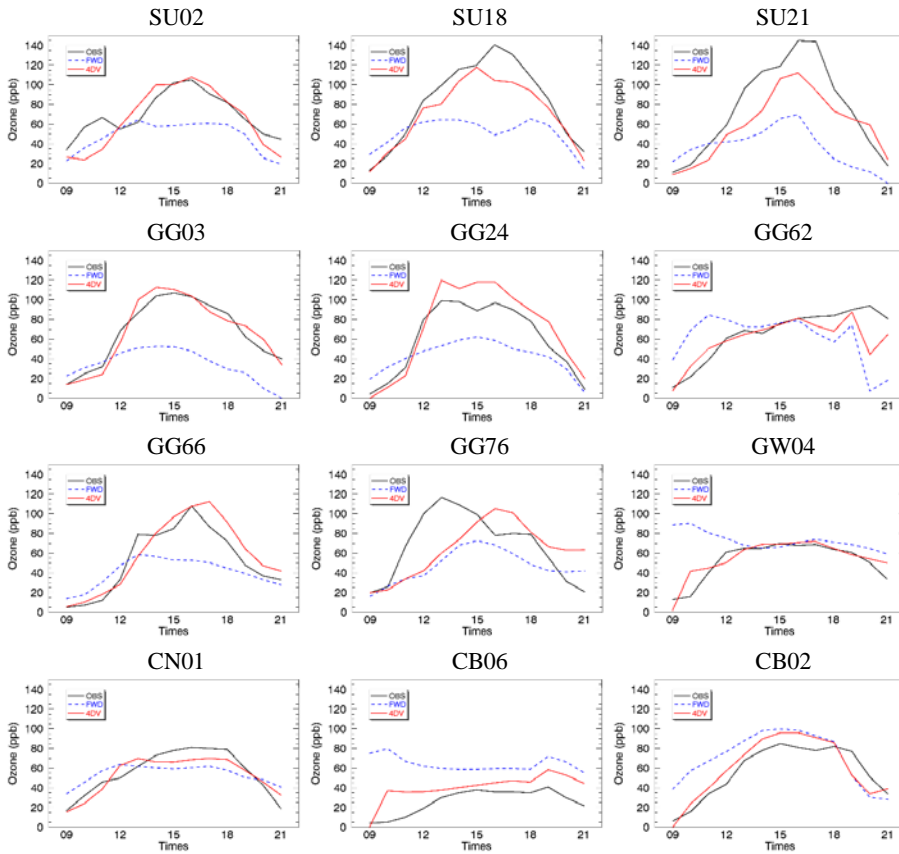
1
 2 Figure 10. Model error correlation coefficients between vertical levels. The physical height of
 3 each level is indicated by the non-uniform grid line only in the layer below 1553 m, which is
 4 the 8th layer of CMAQ.
 5



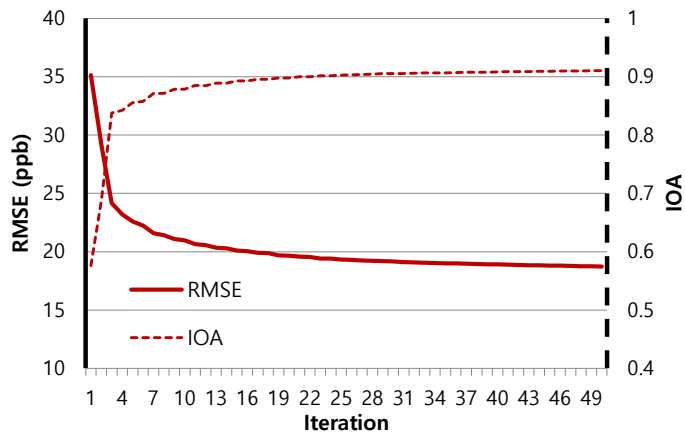
1
2 Figure 11. Model error correlation coefficients between two layers, as a function of Δz (the
3 distance between two levels). The fitted line is $R = e^{-\frac{\Delta z^{1.2}}{l_z^{1.2}}}$, where $l_z = 300$ m.
4



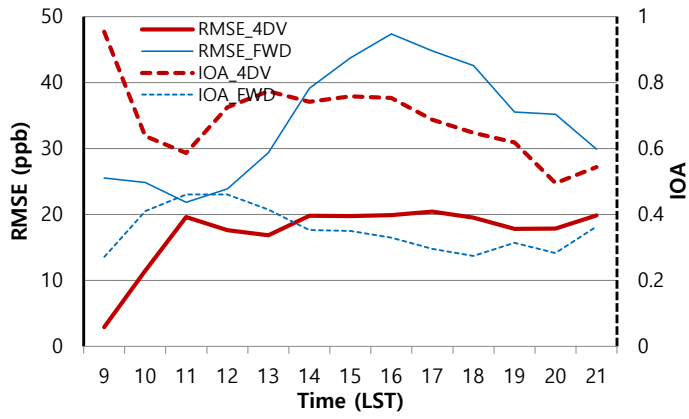
1
 2 Figure 12. Model error correlation coefficients as a function of horizontal distance Δx or Δy ,
 3 which is correspond to East-West (revert triangles) and North-South (blank circles) direction,
 4 respectively. They can be fitted to $R = e^{-\frac{\Delta h^{1.0}}{l_h^{1.0}}}$, where $l_h = 10$ km.
 5



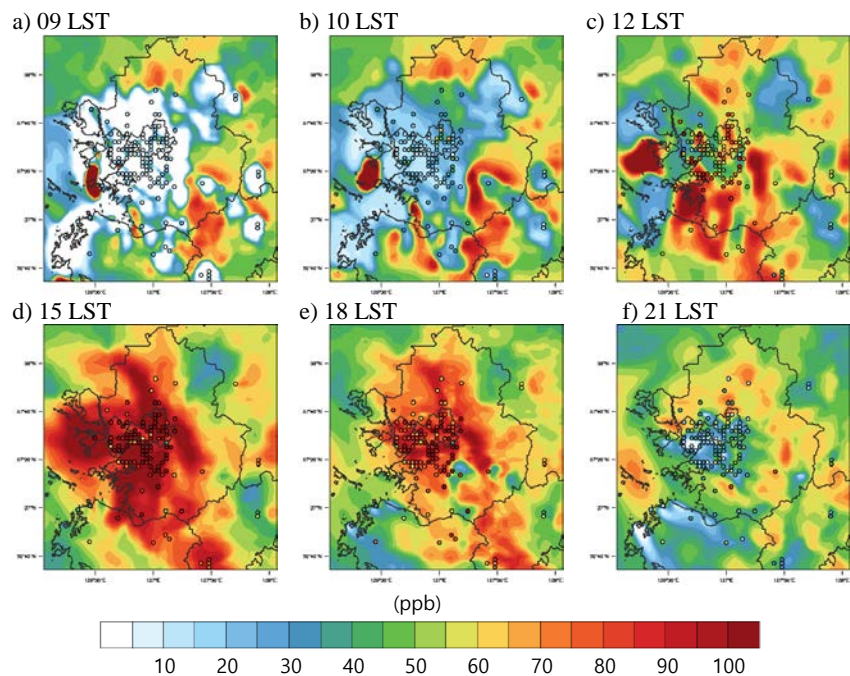
1 Figure 13. Time variations of surface ozone concentration at selected sites whose specific
 2 locations are shown in Figure 1 with red filled triangles during daytime on 5 August. Black
 3 solid lines are observed results, and blue dashed and red solid lines indicate simulated results
 4 from the FWD and 4DV, respectively.
 5



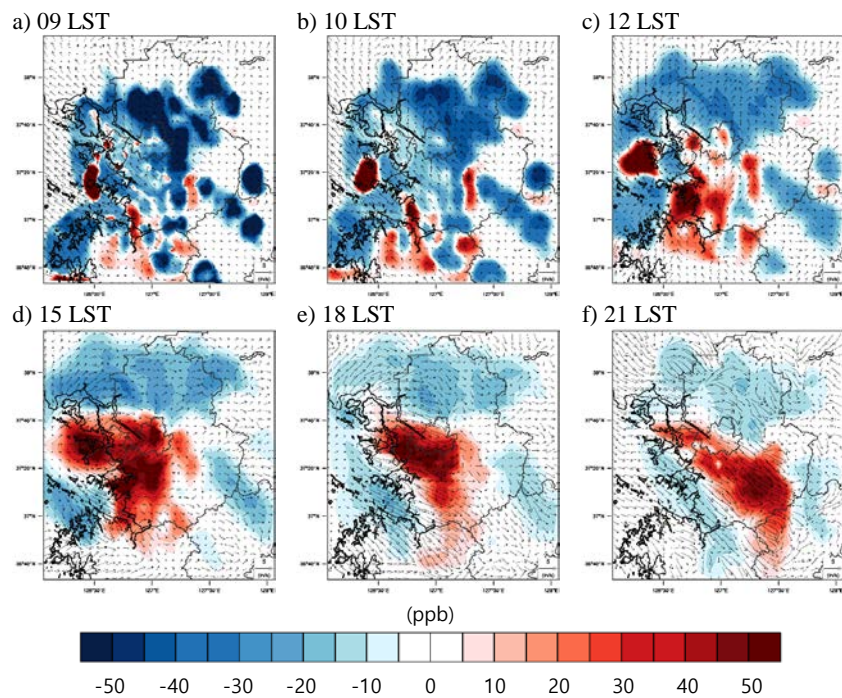
1
 2 Figure 14. Decreasing Root Mean Square Error (RMSE, solid line) and increasing Index Of
 3 Agreement (IOA, dashed line) with respect to each iteration step. The RMSE and IOA are
 4 calculated by comparing 4D-Var data assimilation (4D-Var) results during time-window with
 5 observed O₃ concentration.
 6



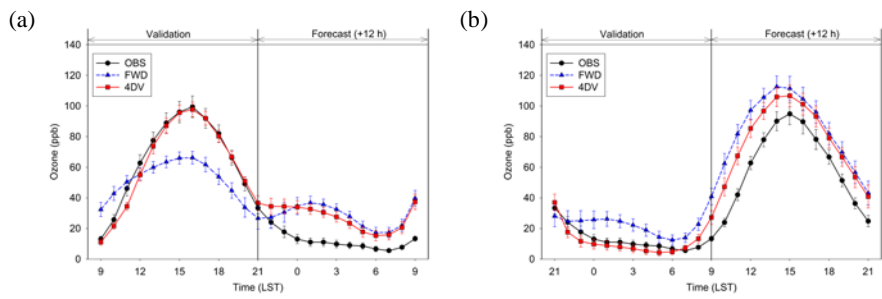
1
 2 Figure 15. Diurnal variations of statistical results of IOA (dashed) and RMSE (solid) during
 3 the assimilation time-window. The results with assimilation (4DV) are indicated by red and
 4 thick lines, and those without assimilation (FWD) are the blue and thin lines.
 5



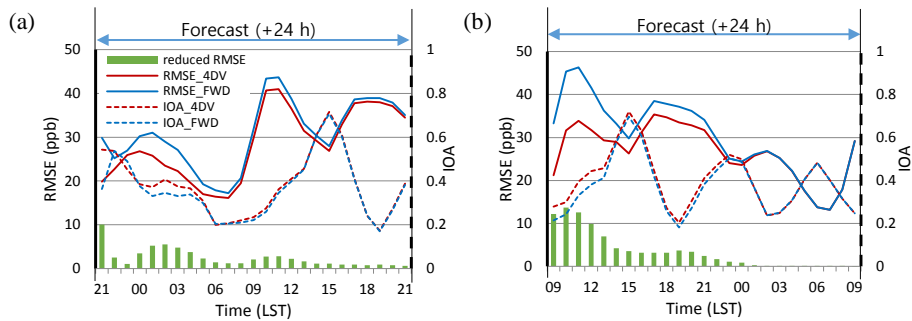
1 Figure 16. Horizontal distributions of surface ozone and its time variations. The plotted time is
 2 valid at a) 09, b) 10, c) 12, d) 15, e) 18, and f) 21 LST on 5 August. Contour value stands for
 3 simulated results of 4DV experiment and the filled circles with the same colour scale as the
 4 contours indicate observed values.
 5



1 Figure 17. The same as Figure 16 except that the contour value is analysis increments
 2 (leftmost in the upper panels) and its impact on daytime ozone.
 3



1 Figure 18. Time variations of observed and forecast ozone concentration after (a) daytime and
 2 (b) nighttime assimilation. All 120 sites data are averaged and its 3 standard errors also
 3 displayed with vertical bars. Triangle over blue dashed line, circle over red solid line, and dot
 4 over black solid line stand for forward run (FWD), 4D-Var run (4DV), and observation (OBS)
 5 results, respectively.
 6



1 Figure 19. Time variations of RMSE (solid lines) and IOA (dashed lines) for 24 hours
 2 forecast after (a) daytime and (b) nighttime assimilation. Red and Blue lines indicate the
 3 statistical results for 4D-Var run (4DV) and forward run (FWD), respectively. Hourly reduced
 4 RMSE values are also marked along the axis of abscissas.