This presentation participates in OSPP

Outstanding Student Poster & PICO Contest

Variational Data Assimilation for the Optimized Ozone Initial State and the Short-Time Forecasting

Soon-Young Park[1], **Dong-Hyeok Kim**[1], **Soon-Hwan Lee**[2], **and Hwa Woon Lee**[3]

[1]Institute of Environmental Studies, Pusan National University, Busan, Republic of Korea

[2]Department of Earth Science Education, Pusan National University, Busan, Republic of Korea

[3]Division of Earth Environmental System, Pusan National University, Busan, Republic of Korea

CONTACT: soon@pusan.ac.kr

ATMOSPHERIC DISPERSION LAB.

PUSAN NATIONAL UNIVERSITY

CC BY

1. Introduction

Data assimilation (DA) provides a consistent representation of the physical state by blending imperfect model predictions and noisy observations. With more chemical observations available in recent years data assimilation is expected to improve the predictability of air quality.

- One of the important elements affecting results of data assimilation in the 4D-Var process is the background error covariance (BEC) of the model.
- In this study, the previously developed 4D-Var code has been modified to treat background errors in matrix forms, and various numerical tests have been conducted.
- Characteristics of the background errors obtained through long-term modeling results are analyzed. Also, the predictability of high ozone concentration was investigated.

2. 4D-Var Data Assimilation

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

$$\mathcal{J}(\boldsymbol{c}_0) = \frac{1}{2}\left(\boldsymbol{c}_0 - \boldsymbol{c}_0^b\right)^{\mathrm{T}}\mathbf{B}_0^{-1}\left(\boldsymbol{c}_0 - \boldsymbol{c}_0^b\right) + \frac{1}{2}\sum_{k=1}^{F}\left(\mathcal{H}(\boldsymbol{c}_k) - \boldsymbol{c}_k^{obs}\right)^{\mathrm{T}}\mathbf{R}_k^{-1}\left(\mathcal{H}(\boldsymbol{c}_k) - \boldsymbol{c}_k^{obs}\right). \quad (1)$$

The 4D-Var analysis can be obtained by the initial concentration that minimizes (1) with respect to the model equation. Formally, a gradient-based optimization procedure is used to obtain minimum value. Assuming a linear observation operator $\mathbf{H}_k = \mathcal{H}'(\boldsymbol{c}_t)$, the gradient of (1) with respect to $\boldsymbol{c}_0$ is

$$\nabla_{\boldsymbol{c}_0}\mathcal{J}(\boldsymbol{c}_0) = \mathbf{B}_0^{-1}\left(\boldsymbol{c}_0 - \boldsymbol{c}_0^b\right) + \sum_{k=1}^{F}\left(\frac{\partial \boldsymbol{c}_k}{\partial \boldsymbol{c}_0}\right)^{\mathrm{T}}\mathbf{H}_k^{\mathrm{T}}\mathbf{R}_k^{-1}\left(\mathbf{H}_k\boldsymbol{c}_k - \boldsymbol{c}_k^{obs}\right). \quad (2)$$

The gradient for the 4D-Var cost function can be effectively obtained by forcing the adjoint model with observation increments and calculating it backwards. When the forward and reverse adjoint models are performed, it results in the problem of solving the following equation:

$$\nabla_{\boldsymbol{c}_0}\mathcal{J}(\boldsymbol{c}_0) = \mathbf{B}_0^{-1}\left(\boldsymbol{c}_0 - \boldsymbol{c}_0^b\right) + \boldsymbol{\lambda}_0 = 0. \quad (3)$$

$\boldsymbol{\lambda}_0$ is the sensitivity of the cost function (1) with respect to the initial concentration $\boldsymbol{c}_0$. Since $\mathbf{B}_0^{-1}$, $\boldsymbol{c}_0^b$, and $\boldsymbol{\lambda}_0$ values are known, if the value of $\boldsymbol{c}_0$ that satisfies equation (3) is found, it becomes the analysis field $\boldsymbol{c}_0^a$.

The adjoint code for CMAQ (CMAQ-ADJ) model was implemented from the project H98 (University of Huston, 2009) by Huston Advanced Research Center / Texas Environmental Research Consortium (HARC/TERC). The validation and several numerical tests of this code are well described in Hakami et al. (2007). This model only considers the model and observation errors as its variance, i.e. a constant value of σ_0^B and σ_k^{obs} with Gaussian distribution.

$$\mathcal{J}(\boldsymbol{c}_0) = \frac{1}{2\left(\sigma_0^b\right)^2}\left(\boldsymbol{c}_0 - \boldsymbol{c}_0^b\right)^{\mathrm{T}}\left(\boldsymbol{c}_0 - \boldsymbol{c}_0^b\right) + \frac{1}{2\left(\sigma_k^{obs}\right)^2}\sum_{k=1}^{N}\left(\mathbf{H}_k\boldsymbol{c}_k - \boldsymbol{c}_k^{obs}\right)^{T}\left(\mathbf{H}_k\boldsymbol{c}_k - \boldsymbol{c}_k^{obs}\right) \quad (4)$$

3. Experimental Design

The capital region of South Korea is selected for the present data assimilation study because measurement sites are relatively evenly distributed.

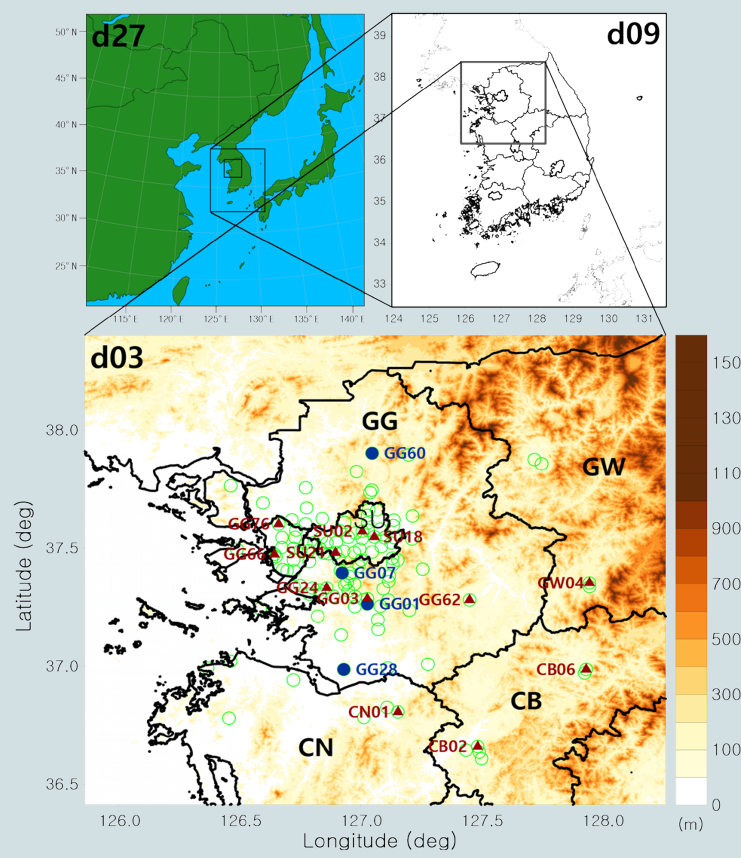

Figure 1. The model domains for WRF and CMAQ. The air quality monitoring sites at ground level are marked by green blank circles.

Table 1. Configuration of WRF modeling system

WRF	d27	d09	d03
Horizontal grid	123 × 130	72 × 84	65 × 68
Grid resolution	27 km	9 km	3 km
Vertical layers	33 layers (top: 50 hPa)		
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		

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
Grid resolution			27 km	9 km	3 km
Vertical layers			15 layers (top: 20 km)		
Emission data			INTEX-B	CAPSS	CAPSS
Time periods	Forward (FWD)		00 UTC 03 ~ 00 UTC 07 August, 2008 (4 days)		
	4D-Var (4DV)	day time	00 UTC 05 ~ 12 UTC 05 August, 2008 (analysis) 12 UTC 05 ~ 12 UTC 06 August, 2008 (forecast)		
		night time	12 UTC 05 ~ 00 UTC 06 August, 2008 (analysis) 00 UTC 06 ~ 00 UTC 07 August, 2008 (forecast)		

4. Idealized BEC Results

Simple yet popular covariance model is Balgovind functions expressed as:

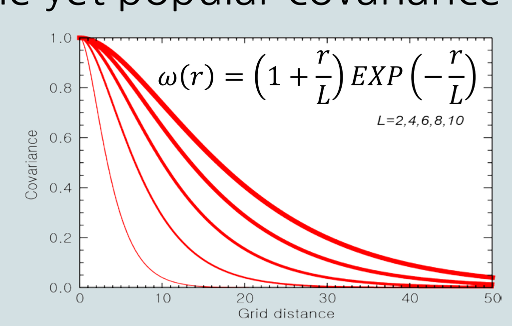

Figure 2. Covariance distribution for Balgovind functions with respect to the grid distance (r) and the values of radius of influence (L).

Table 3. Experimental design for the idealized background error covariance test.

Case	Obs. Data	Radius of Influence	σ_0^B	σ_k^{obs}
L02	100 ppb at (29,31)	L=02	BEC	8.00
L05		L=05	BEC	8.00
L10		L=10	BEC	8.00

This experiment is used to evaluate the characteristics of the BEC based on a single observation experiment. In this experiment, 100 ppb of O_3 was incorporated as an arbitrary value rather than actual observation data at the initial time at the center of the model domain.

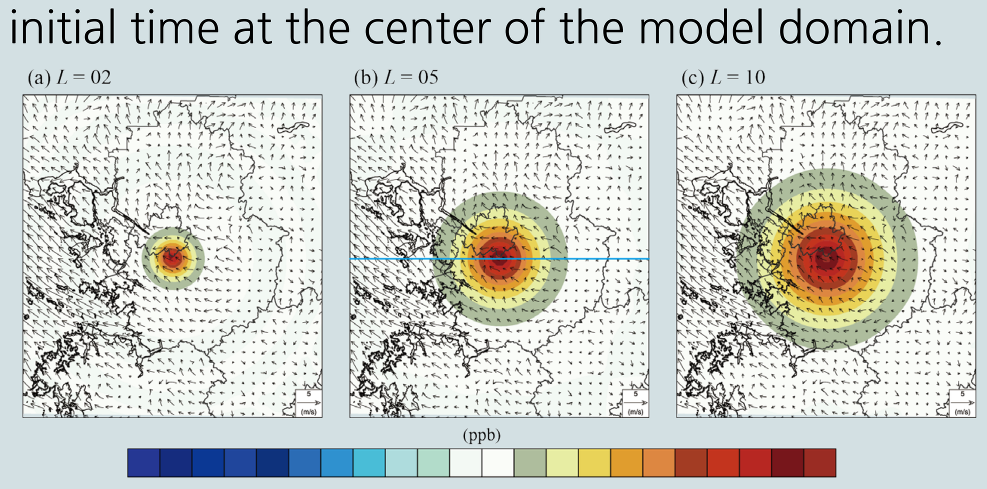

Figure 3. Horizontal distribution of analysis increments at surface resulted from the single observation experiment with respect to radius of influence (L).

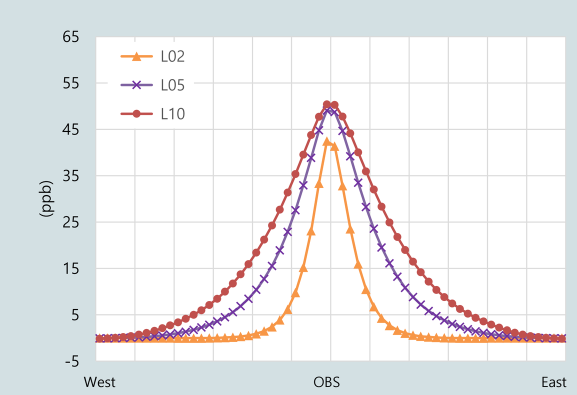

Figure 4. Cross-section of analysis increments along the blue line in Figure 6. (b) as the radius of influence (L) values are increase.

At the model grid point (29, 31), where arbitrary observation data were applied, all three tests showed an O_3 increment of about 50.0 ppb. The background concentration of O_3 at the grid was 40.1 ppb, but the value was up to about 90 ppb in the analysis when the synthetic observation of 100 ppb was applied. However, as the value of L increased, the O_3 increment in the analysis occurs at more surrounding grids. These results indicate that the idealized BEC performs well in the revised codes, and proper analysis increments can be achieved when the spatial correlation is taken into account.

5. Development of Realistic BEC

The BEC is obtained using the NMC approach, which is based on a real simulation for the realistic 4D-Var data assimilation study. The error statistics for the CMAQ model is defined by the differences between +48 hours and +24 hours forecast, $\boldsymbol{\epsilon}^i = \boldsymbol{c}_{+48h}^i - \boldsymbol{c}_{+24h}^i$. We assume $\mathbf{B}$ can be written as $\mathbf{B} = \mathbf{X} \otimes \mathbf{Y} \otimes \mathbf{Z} \otimes \mathbf{C}$ where, $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ representing the error correlation in the three directions. $\mathbf{C}$ is the error covariance matrix at a single grid point that refers to the error variances and correlation between different species. In this study, $\mathbf{C}$ is considered to be no correlation between the species. It seems to be error-prone to invert ill-conditioned matrices. Based on the property of Kronecker product, $\mathbf{B}^{-1}$ can be expressed as $\mathbf{B}^{-1} = (\mathbf{X} \otimes \mathbf{Y} \otimes \mathbf{Z})^{-1} = \mathbf{X}^{-1} \otimes \mathbf{Y}^{-1} \otimes \mathbf{Z}^{-1}$. Singular Value Decomposition (SVD) is applied to $\mathbf{B}$ matrix.

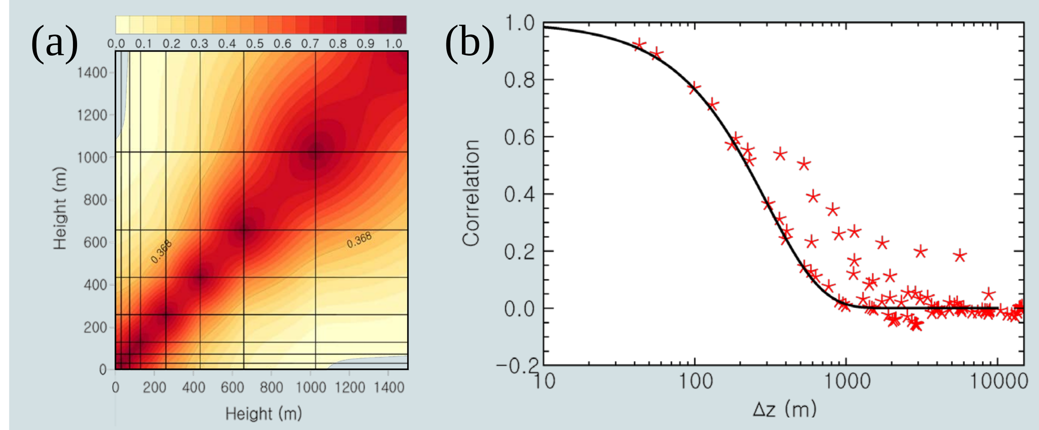

Figure 5. Model error correlation coefficients (a) between vertical levels and (b) between two layers as a function of Δz. The fitted line is $R = exp(-\Delta z^{1.2}/l_z^{1.2})$, where l_z=300 m.

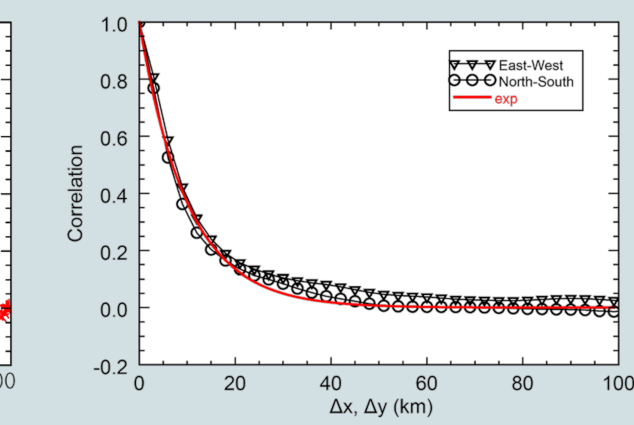

Figure 6. The error correlation as a function of horizontal distance Δx or Δy. The red line corresponds to $R = exp(-\Delta h^{1.0}/l_h^{1.0})$, where l_h=10 km.

The correlation coefficient for the east-west direction is somewhat higher than that for the south-north direction.

6. Validation and Forecasting Results

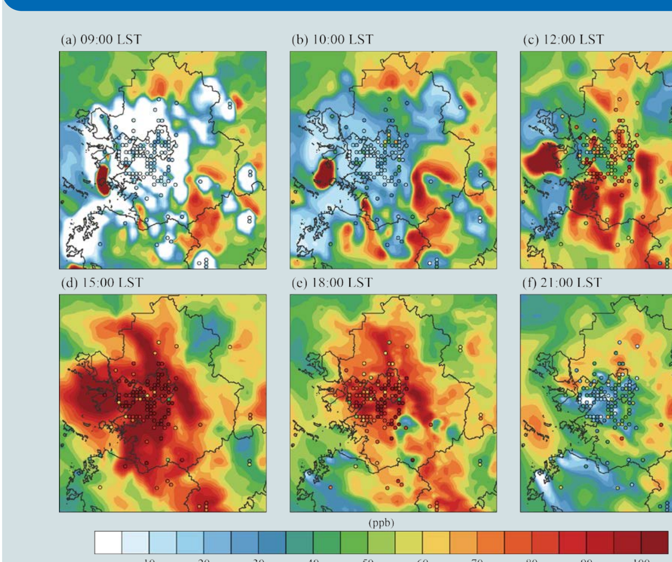

Figure 7. Horizontal distributions of surface ozone and its time variations.

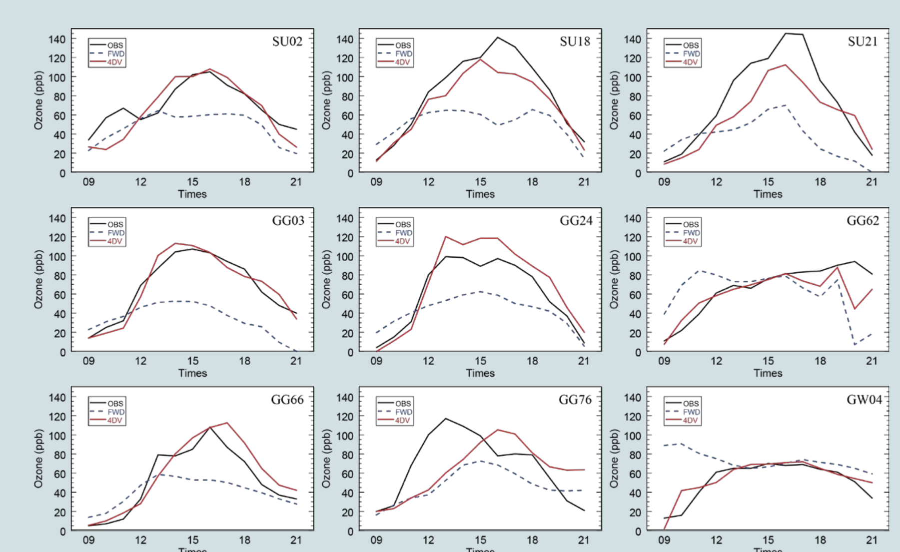

Figure 8. Time variations of surface ozone concentration at selected sites.

The 4DV experiment shows a diurnal variation of O_3 concentration that conforms well to the observation. RMSE decreases by about 49.4%, and the IOA increases by 59.9%, suggesting that the initial conditions of ozone concentration are successfully improved by application of DA.

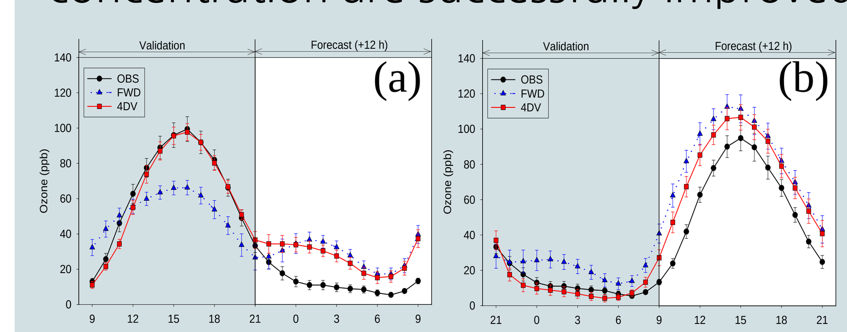

Figure 9. Time variations of observed and forecast ozone concentration after (a) daytime and (b) nighttime assimilation.

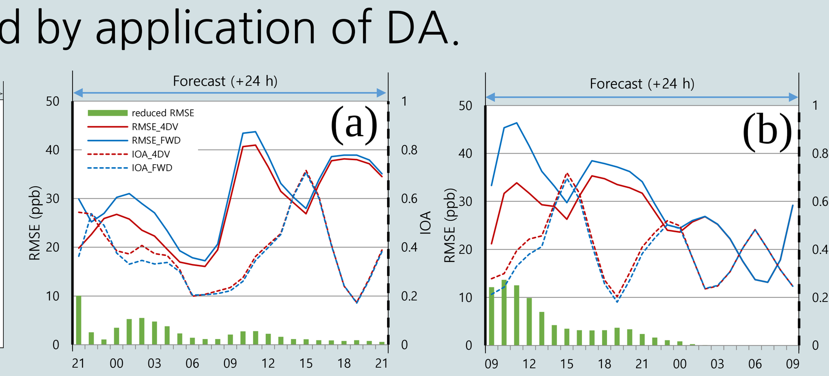

Figure 10. Time variations of RMSE and IOA for 24 h forecast after (a) daytime and (b) nighttime assimilation.

A potential improvement for the ozone predictability is presented using the optimized initial condition after the time-window. In particular, a larger improvement in the predictability of daytime ozone concentration is expected if DA is performed over the nighttime than in the daytime.