



Supplement of

Implementation of an ensemble Kalman filter in the Community Multiscale Air Quality model (CMAQ model v5.1) for data assimilation of ground-level PM_{2.5}

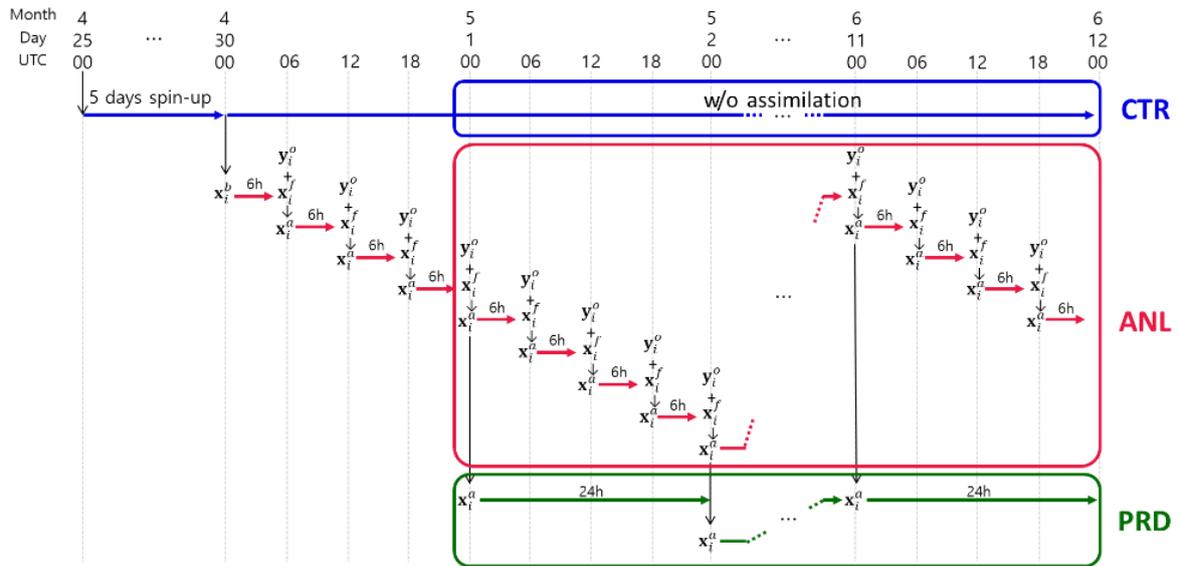
Soon-Young Park et al.

Correspondence to: Chul Han Song (chsong@gist.ac.kr)

The copyright of individual parts of the supplement might differ from the article licence.

Table S1. Definitions for statistical metrics used in this study for model evaluations.

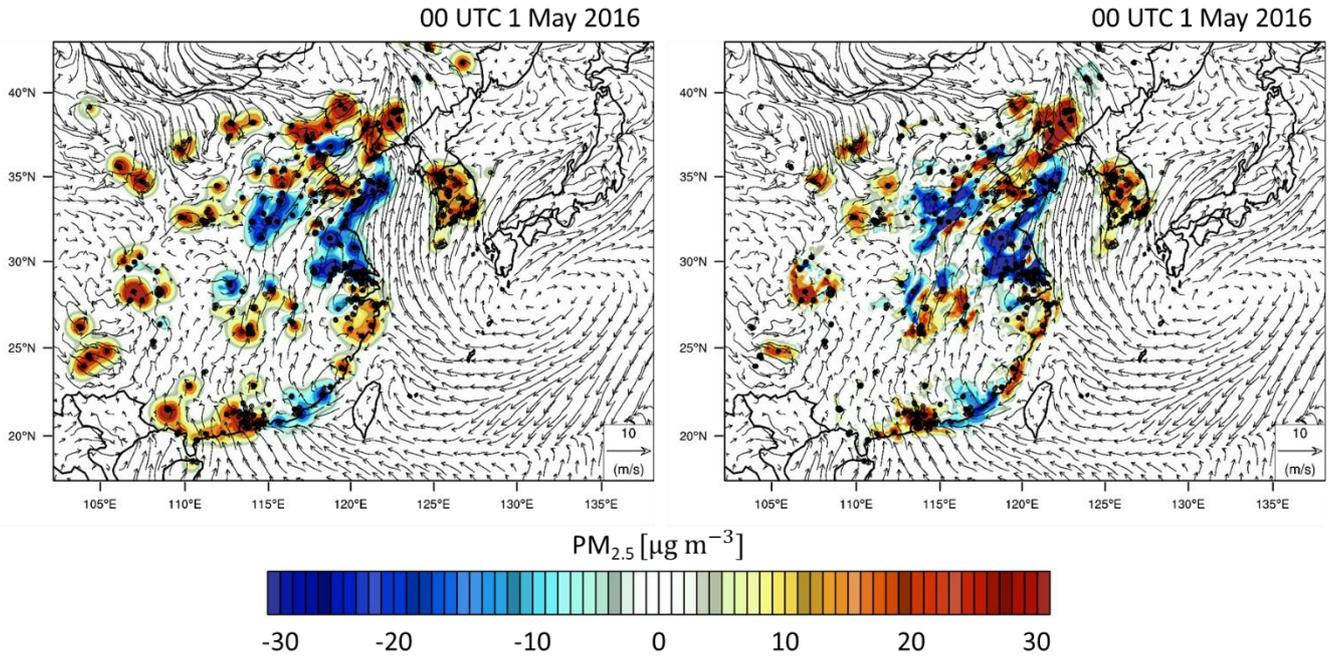
| Descriptions | Metrics | Definition* |
|-------------------------|----------------|---|
| Mean observation | \bar{O} | $(1/N) \sum_{i=1}^N O_i$ |
| Mean model | \bar{M} | $(1/N) \sum_{i=1}^N M_i$ |
| Correlation Coefficient | R | $\frac{\sum_{i=1}^N (M_i - \bar{M})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^N (M_i - \bar{M})^2} \sqrt{\sum_{i=1}^N (O_i - \bar{O})^2}}$ |
| Index Of Agreement | IOA | $1 - \frac{\sum_{i=1}^N (M_i - O_i)^2}{\sum_{i=1}^N (M_i - \bar{O} + O_i - \bar{O})^2}$ |
| Root Mean Squared Error | RMSE | $\sqrt{(1/N) \sum_{i=1}^N (M_i - O_i)^2}$ |
| Mean Bias | MB | $(1/N) \sum_{i=1}^N (M_i - O_i)$ |
| Normalised Mean Bias | NMB | $(1/N) \sum_{i=1}^N (M_i - O_i) / \bar{O} \times 100$ |



5 **Figure S1.** Simulation strategy for data assimilation cycles and 1-day predictions. The case names of CTR, ANL, and PRD indicate the control, reanalysis, and prediction run, respectively.

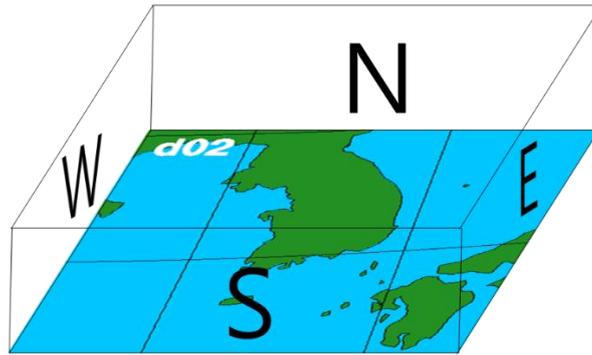
(a) 3DVAR

(b) EnKF



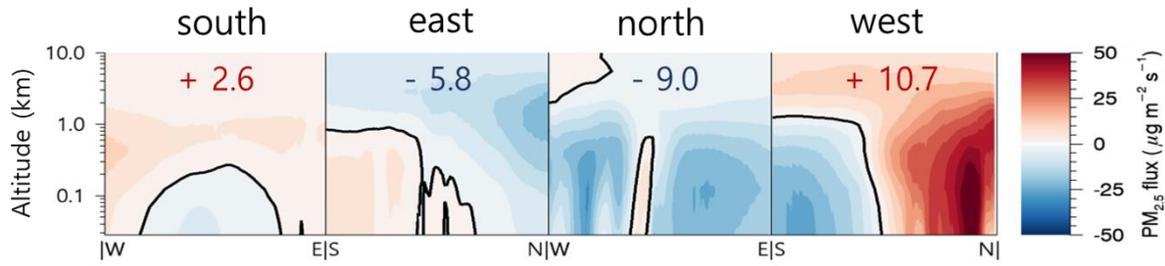
10

Figure S2. Snapshot of two analysis increments and wind vectors from (a) 3DVAR and (b) EnKF DA methods at ground level at 00 UTC, 1 May 2016. The blank circles with dot in the centre indicate the observation sites. Note that the increments are different along the marginal area in the domain 1 because several parameters were considered differently such as observation operator, number of observation data, and observation error.

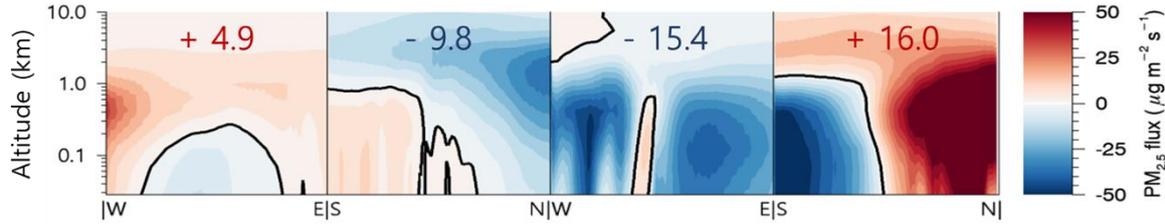


15 **Figure S3. Three-dimensional view of the four lateral boundaries of domain 2 (D2). N, S, E, and W indicate north, south, east, and west planes, respectively.**

(a) DA_ic



(b) DA_icbc



(c) Difference (DA_icbc - DA_ic)

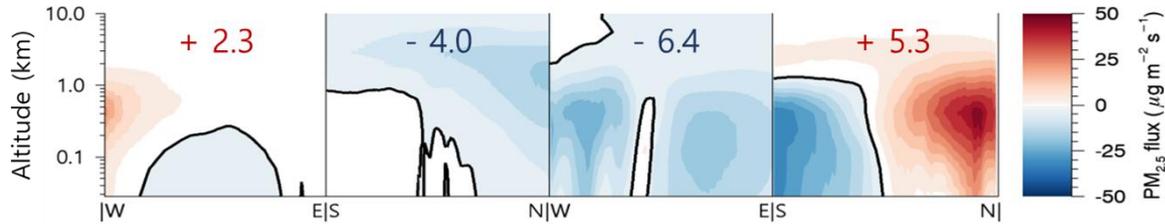
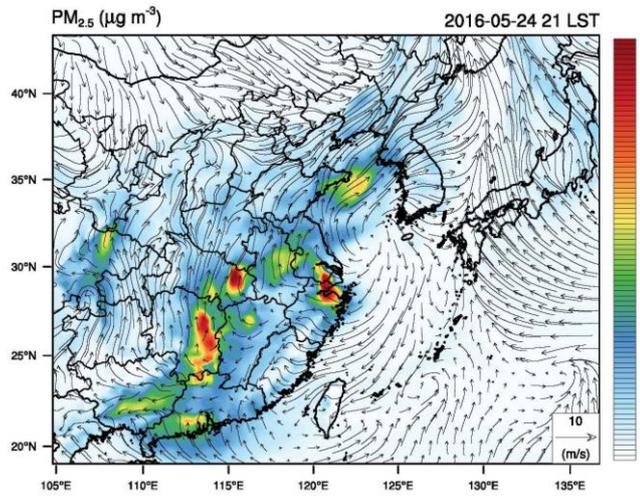


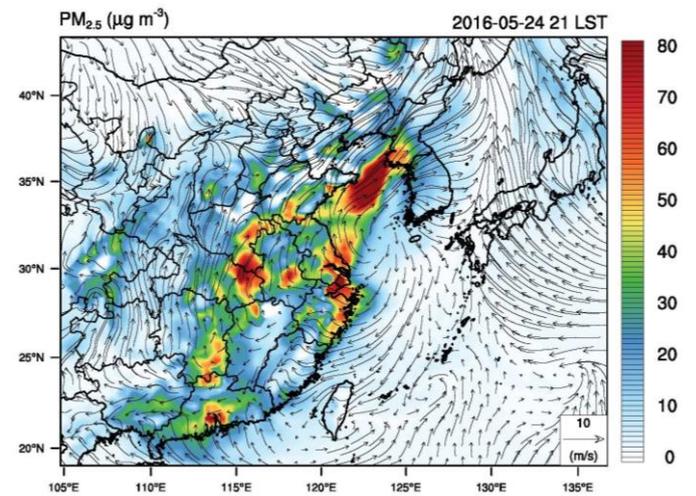
Figure S4. Averaged $PM_{2.5}$ fluxes in the four lateral boundary planes (south, east, north, and west from left to right, refer Fig. S2) for simulation periods along the four perimeters of the fine domain (D2). The thick black lines indicate zero flux. The positive and negative values on each boundary present the plain averaged fluxes. (a) DA_ic, and (b) DA_icbc, are the averaged $PM_{2.5}$ fluxes without, and with, the EnKF data assimilation in China (D1), respectively. The increments in DA_icbc experiment are also presented at the bottom (c). Note that the y-axis for the height is log-scale, to show the results below the planetary boundary layer in detail.

20

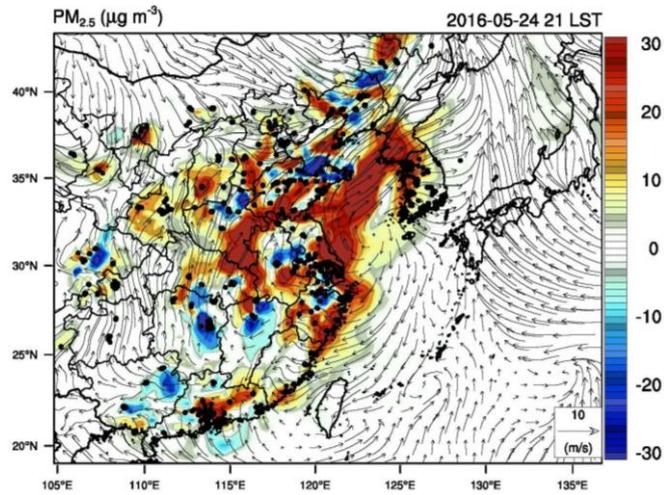
(a) Boundary condition used for DA_ic



(b) Boundary condition used for DA_icbc



(c) Differences in boundary condition between DA_icbc and DA_ic



25 Figure S5. Horizontal distributions of PM_{2.5} in the domain 1 (D1) at 21 LST, 24 May 2016, which provided (a) DA_ic and (b) DA_icbc with the BCs for the domain 2 (D2). The differences from DA_icbc to DA_ic is also shown (c). The blank circles with dot in the centre indicate the observation sites.