Influence of CO_2 observations on the optimized CO_2 flux in the CarbonTracker framework

JinWoong Kim[1], Hyun Mee Kim[1]*, and Chun-Ho Cho[2]

[1]*Department of Atmospheric Sciences, Yonsei University, Seoul, Republic of Korea*
[2]*National Institute of Meteorological Research, Jeju, Republic of Korea*
**Corresponding author : khm@yonsei.ac.kr*

1. INTRODUCTION

- Atmospheric CO_2 observations can be used to quantitatively estimate the source and sink of surface carbon fluxes.
- Cardinali et al. (2004) suggested a method for calculating the influence matrix within the general data assimilation framework.
- Liu et al. (2009) suggested a method for calculating self-sensitivity and cross-sensitivity within the ensemble Kalman filter (EnKF) framework.
- In this study, the effect of CO_2 observations on an analysis of surface CO_2 flux was calculated using an influence matrix in the CarbonTracker, which is an inverse modeling system for estimating surface CO_2 flux based on an EnKF.

2. METHODOLOGY

Analysis equation in data assimilation

General analysis equation for data assimilation $\quad \mathbf{x}^a = \mathbf{K}\mathbf{y}^o + (\mathbf{I}_n - \mathbf{K}\mathbf{H})\mathbf{x}^b$

The projection of analysis equation onto the observation space $\quad \mathbf{H}\mathbf{x}^a = \mathbf{y}^a = \mathbf{H}\mathbf{K}\mathbf{y}^o + (\mathbf{I}_p - \mathbf{H}\mathbf{K})\mathbf{y}^b$

Influence matrix

➔ The influence matrix reflects the regression fit of the analysis to observations.

Analysis sensitivity with respect to observations $\quad \mathbf{S}^o = \dfrac{\partial \mathbf{y}^a}{\partial \mathbf{y}^o} = \mathbf{K}^T\mathbf{H}^T = \mathbf{R}^{-1}\mathbf{H}\mathbf{P}^a\mathbf{H}^T$

Analysis sensitivity with respect to background $\quad \mathbf{S}^b = \dfrac{\partial \mathbf{y}^a}{\partial \mathbf{y}^b} = \mathbf{I}_p - \mathbf{K}^T\mathbf{H}^T = \mathbf{I}_p - \mathbf{S}^o$

Influence matrix calculation in EnKF framework

Self-sensitivity (diagonal terms of influence matrix)
➔ the self-sensitivity gives a measure of the analysis sensitivity to the observations.

$$\mathbf{S}^o_{jj} = \frac{\partial \mathbf{y}^a_j}{\partial \mathbf{y}^o_j} = \left(\frac{1}{m-1}\right)\frac{1}{\sigma_j^2}\sum_{i=1}^{m}(\mathbf{H}\mathbf{X}^a_i)_j \times (\mathbf{H}\mathbf{X}^a_i)_j$$

Cross-sensitivity (off-diagonal terms of influence matrix)

$$\mathbf{S}^o_{jl} = \frac{\partial \mathbf{y}^a_l}{\partial \mathbf{y}^o_j} = \left(\frac{1}{m-1}\right)\frac{1}{\sigma_j^2}\sum_{i=1}^{m}(\mathbf{H}\mathbf{X}^a_i)_j \times (\mathbf{H}\mathbf{X}^a_i)_l$$

Information content

➔ Information content reflects the amount of information that the analysis extracts from a subset of observations during data assimilation.

Globally averaged observation influence $\quad GAI = \dfrac{\mathrm{tr}(\mathbf{S}^o)}{p}$

Partial influence for selected observations $\quad PAI = \dfrac{\sum_{ii\in I}\mathbf{S}^o_{ii}}{p_I}$

Cumulative impact

➔ Cumulative impact of the influence matrix for the 5 weeks of lag can be calculated because the background in the lagged window already includes the effect from previous observations.

The analysis sensitivity to background at the first week

$$\mathbf{S}^b(1) = (1-\mathbf{S}^o(1))(1-\mathbf{S}^o(2))(1-\mathbf{S}^o(3))(1-\mathbf{S}^o(4))(1-\mathbf{S}^o(5))$$

The cumulative impact of the influence matrix

$$\mathbf{S}^o_{\mathrm{cum}} = 1-\mathbf{S}^b(1) = 1-(1-\mathbf{S}^o(1))(1-\mathbf{S}^o(2))(1-\mathbf{S}^o(3))(1-\mathbf{S}^o(4))(1-\mathbf{S}^o(5))$$

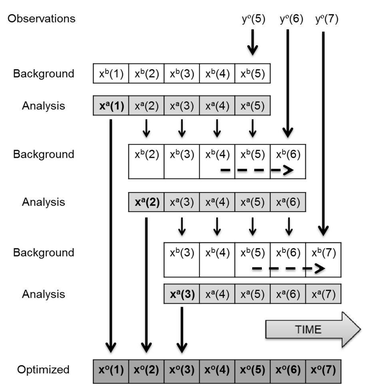

Figure 1. Schematic diagram of the assimilation process employed in CarbonTracker. In each analysis cycle, observations made within one week are used to update the state vectors with a five-week lag. The dashed line indicates how the simple dynamic model uses analysis state vectors from the previous on and two weeks to produce a new background state vector for the current analysis time. The TM5 model is used as the observation operator to calculate the model CO_2 concentration for each corresponding observation location and time.

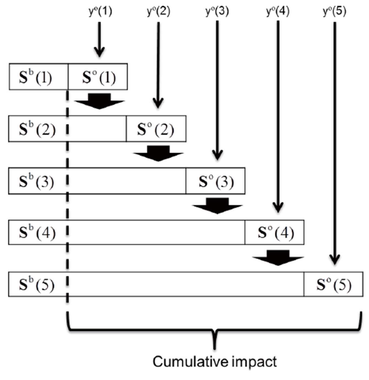

Figure 2. Schematic diagram of calculating cumulative impact in CarbonTracker. $\mathbf{S}^b(\bullet)$ indicates the analysis sensitivity to background at each analysis cycle within 5 weeks of lag, where $\bullet$ denotes each week from 1 to 5.
$\mathbf{S}^o(\bullet)$ indicates the analysis sensitivity to observation at each analysis cycle.

3. EXPERIMENTAL FRAMEWORK

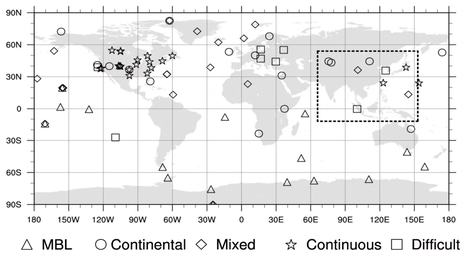

Figure 3. Observation network of CO_2 concentrations around the globe and the nested domain of the TM5 transport model over Asia (dashed box). Each observation site is assigned to different categories (△: MBL; ○: Continental; ◇: Mixed land/ocean and mountain; ☆: Continuous; □: Difficult)

The experimental period is from January 2000 to December 2009. The results for the year 2000 were excluded because 2000 is considered as spin-up period.

4. RESULTS - I

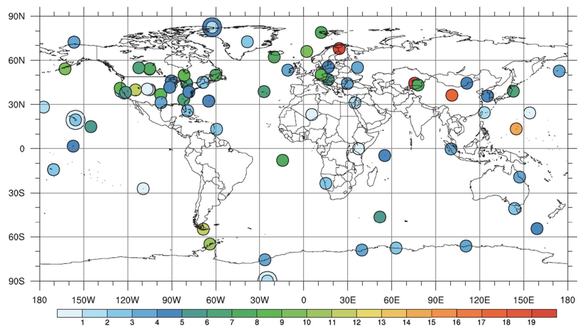

Figure 4. Average self-sensitivity at each observation site from 2001 to 2009. The overlapping observation sites at the same locations or at close locations are distinguished by different sizes of circles.

➔ The average global self-sensitivity is 4.8%.
➔ The cumulative impact over 5 weeks is 19.1% much greater than 4.8 %, which only represents the most recent week of each cycle.

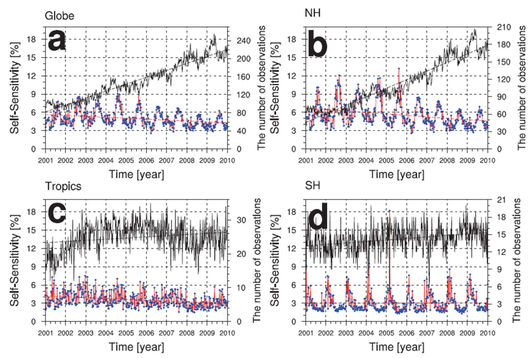

Figure 5. Time series of the average self-sensitivity (red solid line with blue dots) and the number of observations (black solid line) with a weekly temporal resolution **(a)** around the globe and in the **(b)** Northern Hemisphere, **(c)** Tropics, and **(d)** Southern Hemisphere from 2001 to 2009. The dashed lines represent the regression line of the average self-sensitivity (red dashed line) and the number of observations (black dashed line).

➔ The average self-sensitivity decreases as the number of observations increases, showing an inversely proportional relationship.
➔ There is seasonal variability in the average self-sensitivity, showing high values in summer and low values in winter.

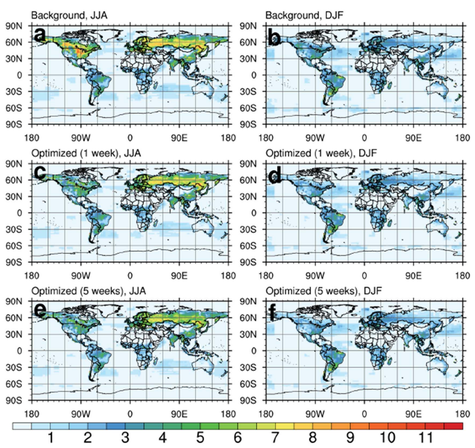

Figure 6. Average standard deviation of background biosphere and ocean fluxes optimized by one-week observations in **(a)** JJA and **(b)** DJF; the posterior biosphere and ocean fluxes optimized by one-week observations in **(c)** JJA and **(d)** DJF; and the posterior biosphere and ocean fluxes optimized by five weeks of observations in **(e)** JJA and **(f)** DJF. The units are g C m^{-2} $week^{-1}$.

➔ The ensemble spread of the background surface CO_2 fluxes reflects the uncertainties, which are projected onto the ensemble spread of the background and analysis CO_2 concentrations by the transport model.

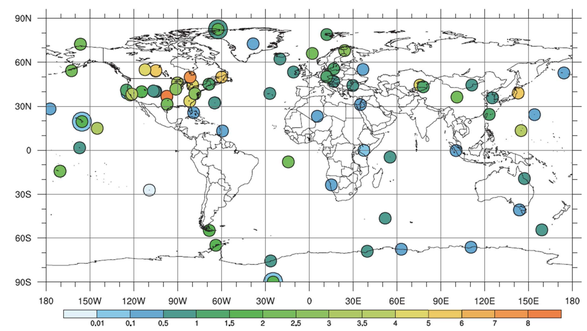

Figure 7. Average normalized information content for each observation site from 2001 to 2009. The overlapping observation sites at the same locations or at close locations are distinguished using different sizes of circles.

➔ The magnitude of information content at one observation site is proportional to the self-sensitivity and the number of observations, the observation sites with a high average self-sensitivity or a large number of observations show high information content.

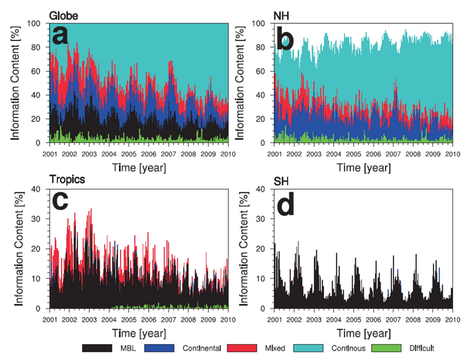

Figure 8. Time series of the average information content for each observation site category **(a)** around the globe and in the **(b)** Northern Hemisphere, **(c)** tropics, and **(d)** Southern Hemisphere from 2001 to 2009.

➔ The proportion of the information content of the Continuous site category increases steadily over time.

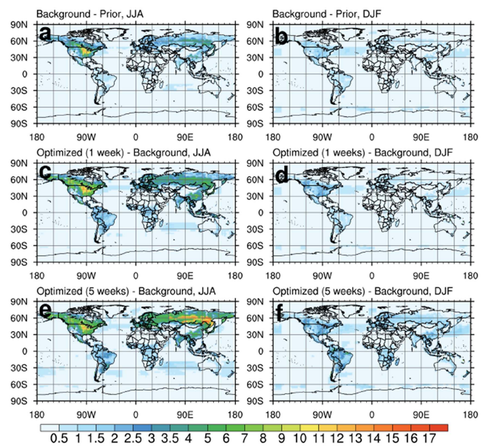

Figure 9. Root mean square difference (RMSD) between the background flux and prior flux in **(a)** JJA and **(b)** DJF; RMSD between the background flux and posterior flux optimized by one-week observations in **(c)** JJA and **(d)** DJF; and RMSD between the background flux and posterior flux optimized by five weeks of observations in **(e)** JJA and **(f)** DJF. The unit is g C m^{-2} $week^{-1}$.

➔ The region with a high average information content are consistent with the regions with a high root mean square differences (RMSD) of the surface CO_2 fluxes.

5. RESULTS - II

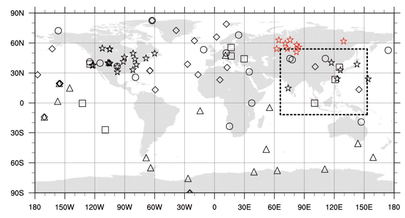

Figure 10. Observation network of CO_2 concentrations around the globe and the nested domain of the TM5 transport model over Asia (dashed box). The sites over Asia in red color indicates the additional observations used in this study.

➔ Two experiments were conducted. The **CNTL** experiment was conducted without JR-STATION (Sasakawa et al., 2013) observations. The **JR** experiment was conducted with the JR-STATION observations in Asia (red color in Fig. 10).

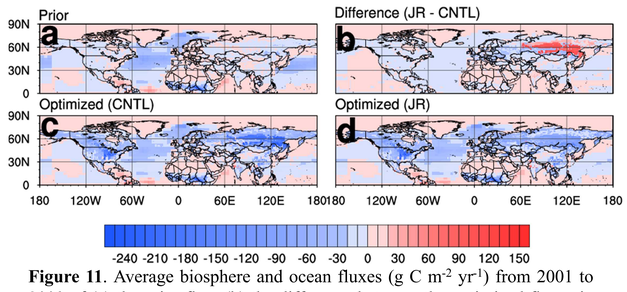

Figure 11. Average biosphere and ocean fluxes (g C m^{-2} yr^{-1}) from 2001 to 2009 of (a) the prior flux, (b) the difference between the optimized fluxes in the JR and CNTL experiment, (c) the optimized flux in the CNTL experiment, and (d) optimized flux in the JR experiment.

➔ While the magnitude of the optimized surface CO_2 flux uptake in Eurasian Boreal (Siberia) was decreased for the JR experiment, the magnitude of the optimized surface CO_2 flux uptake in Eurasia Temperate and Europe was increased.

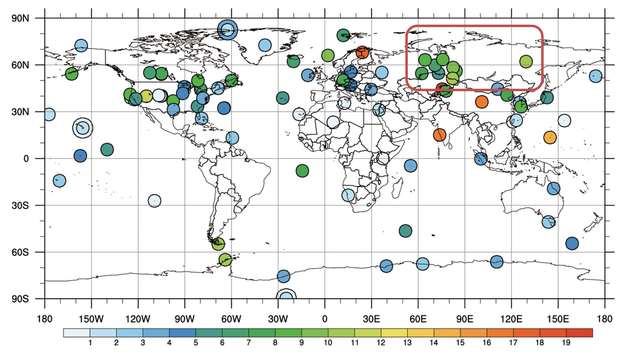

Figure 12. Average self-sensitivity at each observation site from 2001 to 2009. The overlapping observation sites at the same locations or at close locations are distinguished by different sizes of circles.

6. SUMMARY

- In this study, the effect of CO_2 concentration observations on an analysis of surface CO_2 flux was calculated using an influence matrix in the CarbonTracker.
- The self-sensitivity is inversely proportional to the number of observations used in the assimilation.
- The time series of globally averaged self-sensitivities shows seasonal variations, with greater sensitivities in summer and lower sensitivities in winter, which is attributed to the surface CO_2 flux uncertainties.
- The observation sites with a high average self-sensitivity or a large number of observations show high information content.
- The strong correlation between the information content and the optimized surface CO_2 fluxes exists.
- The cumulative impact over 5 weeks is 19.1% much greater than 4.8 %.
- The observation impact of the Siberian observation data is as large as other continuous measurements (e.g., tower measurements in North America).
- More comprehensive results can be found in Kim et al. (2014).

Acknowledgment

The authors thank Andrew R. Jacobson for providing the resources necessary for this study. This work was funded by the Korea Meteorological Administration Research and Development Program under Grant KMIPA 2015-2021.

References

Cardinali, C., S. Pezzulili, and E. Andersson (2004), Influence-matrix diagnostic of a data assimilation system, *Q. J. R. Meteorol. Soc.*, **130**, 2767-2876.

Kim, J., H. M. Kim, and C.-H. Cho (2014), Influence of CO_2 observations on the optimized CO_2 flux in an ensemble Kalman filter, *Atmos. Chem. Phys. Discuss.*, **14**, 1-42, doi:10.5194/acpd-14-1-2014

Liu, J, E. Kalnay, T. Miyoshi, and C. Cardinali (2009), Analysis sensitivity calculation in an ensemble Kalman filter, *Q. J. R. Meteorol. Soc.*, **135**, 1842-1851.

Sasakawa, M, T. Machida, N. Tsuda, M. Arshinov, D. Davydov, A. Fofonov, and O. Krasnov (2013), Aircraft and tower measurements of CO_2 concentration in the planetary boundary layer and the lower free troposphere over southern taiga in West Siberia: Long-term records fro 2002 to 2011, *J. Geophys. Res. Atmos.*, **118**, 9489-9498, doi:10.1002/jgrd.50755.