Atmos. Chem. Phys. Discuss., 14, 13561–13602, 2014
www.atmos-chem-phys-discuss.net/14/13561/2014/
doi:10.5194/acpd-14-13561-2014

Atmospheric Chemistry and Physics Discussions

Influence of CO_2 observations on the optimized CO_2 flux in an ensemble Kalman filter

J. Kim[1], H. M. Kim[1], and C.-H. Cho[2]

[1]Atmospheric Predictability and Data Assimilation Laboratory, Department of Atmospheric Sciences, Yonsei University, Seoul, Republic of Korea
[2]National Institute of Meteorological Research, Jeju, Republic of Korea

Received: 25 April 2014 – Accepted: 12 May 2014 – Published: 26 May 2014

Correspondence to: H. M. Kim (khm@yonsei.ac.kr)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Various data assimilation schemes have been applied in studies on atmospheric CO_2 inversion. An influence matrix based on the linear statistical analysis scheme can diagnose the impact of individual observations on a particular analysis. In this study, to estimate the effect of CO_2 observations on an analysis of surface CO_2 flux, both the analysis sensitivity and the information content were calculated using the influence matrix in the CarbonTracker, which is an inverse modeling system for estimating surface CO_2 flux based on an ensemble Kalman filter. The experimental period was from January 2000 to December 2009. The global average self-sensitivity is 4.8 %, which implies that the analysis extracts 4.8 % of the information from the observations and 95.2 % from the background each assimilation cycle. Because the surface CO_2 flux in each week is optimized by five weeks of observations, the cumulative impact over five weeks would be greater than 4.8 %. The analysis sensitivity is inversely proportional to the number of observations used in the assimilation, which is distinctly apparent in continuous observation categories with a sufficient number of observations. The time series of the globally averaged analysis sensitivities shows seasonal variations, with greater sensitivities in summer and lower sensitivities in winter, which is attributed to the surface CO_2 flux uncertainty. The time-averaged analysis sensitivities in the Northern Hemisphere are greater than those in the Tropics and the Southern Hemisphere. The information content indicates an imbalance between the observation coverage in North America and that in other regions. Approximately half of the total observational information is provided by continuous observations, mainly from North America, which indicates that continuous observations are the most informative and that comprehensive coverage of additional observations in other regions is necessary to estimate the surface CO_2 flux in these areas as accurately as in North America. In addition, the uncertainty of the surface CO_2 flux in Asia, where observations are sparse, is reduced by assimilating five weeks of observations as opposed to one week of observations in

ACPD
14, 13561–13602, 2014

North America, which indicates that a longer assimilation window with a lag is necessary to optimize the surface CO_2 flux in Asia.

1 Introduction

Atmospheric CO_2 observations can be used to quantitatively estimate the sources and sinks of surface carbon fluxes. Thus, atmospheric CO_2 inversion studies using atmospheric CO_2 observations have been conducted (Gurney et al., 2002; Ciais et al., 2010; Peylin et al., 2013). Various studies applying state-of-the-art data assimilation methods have been carried out to estimate the surface carbon cycle at global and regional scales. The methods employed for the atmospheric CO_2 inversion studies include variational data assimilation methods (Chevallier et al., 2005, 2009a, b; Baker et al., 2006, 2010; Basu et al., 2013), ensemble Kalman filter (EnKF) (Peters et al., 2005, 2007, 2010; Feng et al., 2009; Miyazaki et al., 2011; Kang et al., 2011, 2012; Chatterjee et al., 2012; Kim et al., 2012, 2014), and maximum likelihood ensemble filter (Zupanski et al., 2007; Lokupitiya et al., 2008). These studies have applied the data assimilation method used in numerical weather prediction (NWP) to estimate surface CO_2 fluxes.

Recent studies on atmospheric CO_2 inversion have focused on analyzing the difference between prior and optimized surface CO_2 fluxes obtained by using new inversion methods or observations (Chevallier et al., 2009a; Basu et al., 2013), as well as the carbon cycle based on optimized surface CO_2 fluxes. By contrast, the impact of various atmospheric CO_2 observations on the estimation of surface CO_2 fluxes has rarely been studied. One method employed to evaluate the impact of observations on atmospheric CO_2 inversion is the calculation of the uncertainty reduction (Peters et al., 2005; Meirink et al., 2008; Chevallier et al., 2009b; Feng et al., 2009), which is a ratio between the variances of the prior and posterior state vectors. A large uncertainty reduction implies that observations have a large impact on the estimation of surface CO_2 fluxes. However, the uncertainty reduction cannot measure the impact of individual observations on the estimated (i.e., analyzed) surface CO_2 fluxes. Another method

ACPD
14, 13561–13602, 2014

for assessing the impact of observations is to calculate the information content, which is the amount of information obtained from observations (Rodgers, 2000). Engelen and Stephen (2004) calculated the information content of infrared satellite sounding observations on atmospheric CO_2 concentrations. To estimate the impact of simulated CO_2 observations on surface flux analysis, Zupanski et al. (2007) calculated the information content using the information matrix in the ensemble subspace. However, similar to the uncertainty reduction, these methods calculate the impact of all observations, rather than calculating the impact of individual observations on surface CO_2 flux analysis.

Data assimilation algorithms are fundamentally based on a linear statistical assumption (Talagrand, 1997). Both sequential and variational algorithms combine background and observation information to estimate parameters based on the linear assumption. According to the linear assumption, the influence matrix that measures the impact of individual observations on estimated parameters can be calculated in the observation space. Cardinali et al. (2004) suggested a method for calculating the influence matrix within the general data assimilation framework and applied the method to a forecast model of the European Centre for Medium Weather Forecasts (ECMWF). The diagonal elements of the influence matrix are the analysis sensitivities (i.e., self-sensitivity), which are proportional to the spread of the analysis and are inversely proportional to the predetermined observation error. The trace of the diagonal elements of the influence matrix reflects the information content, which is the amount of information extracted from observations. The influence matrix provides objective diagnostics regarding the impact of observations on the analysis and hence the performance of the data assimilation system because inaccurate observations can be identified by analyzing the observation impact (Cardinali et al., 2004). Liu et al. (2009) suggested a method for calculating self-sensitivity and cross-sensitivity (i.e., off-diagonal elements of the influence matrix) within the EnKF framework and diagnosed the relative importance of individual observations within an observation system using the idealized Lorenz-40 model and the simplified hydrostatic model.

ACPD
14, 13561–13602, 2014

Although Cardinali et al. (2004) and Liu et al. (2009) suggested methods for calculating the impact of individual observations on an analysis, their studies focused on NWP. Therefore, the impact of individual observations on surface CO_2 flux analysis has not been diagnosed in a study on atmospheric CO_2 inversion using the state-of-the-art data assimilation method. Because the analysis is more important than the forecast in atmospheric CO_2 inversion, the methods suggested by Cardinali et al. (2004) and Liu et al. (2009) can be applied to diagnose the impact of observations on the analysis.

CarbonTracker is a system developed by the National Oceanic and Atmospheric Administration (NOAA), which optimizes the surface CO_2 flux by assimilating mole fraction observations (i.e., concentration) of surface CO_2 (Peters et al., 2005). CarbonTracker has been applied in studies on atmospheric CO_2 inversion in North America (Peters et al., 2010), Europe (Peters et al., 2010), and Asia (Kim et al., 2014). To develop CarbonTracker for use in Asia, Kim et al. (2012) performed an experiment employing CarbonTracker in this region and demonstrated that CarbonTracker produces optimized surface CO_2 fluxes for Asia. Kim et al. (2014) showed that the estimates of the surface CO_2 flux are more consistent with observed CO_2 concentrations in Asia when using the nesting domain of the transport model on Asia in CarbonTracker. Zhang et al. (2013) conducted a study on the assimilation of aircraft CO_2 observations from the Comprehensive Observation Network for TRace gases by AIrLiner (CONTRAIL) (Machida et al., 2008) in Asia using CarbonTracker.

In this study, an influence matrix is calculated in CarbonTracker to evaluate the impact of mole fraction observations of CO_2 on the analyzed surface CO_2 fluxes. The relative importance of each observation site and each observation site category is evaluated by analyzing the self-sensitivity and information content, and the characteristics of the self-sensitivity and information content are subsequently investigated. Section 2 presents the experimental framework, which includes CarbonTracker, EnKF, observations, the methodology for calculating the influence matrix, and the experimental framework. Section 3 presents the results, and Sect. 4 provides a summary and conclusion.

ACPD
14, 13561–13602, 2014

2 Methodology

2.1 CarbonTracker

CarbonTracker is an atmospheric CO_2 inversion system that estimates the surface CO_2 flux consistent with CO_2 observations. In CarbonTracker, the optimized flux with a $1^\circ \times 1^\circ$ horizontal resolution is calculated by

$$F(x,y,t) = \lambda_r \cdot F_{\text{bio}}(x,y,t) + \lambda_r \cdot F_{\text{ocn}}(x,y,t) + F_{\text{ff}}(x,y,t) + F_{\text{fire}}(x,y,t), \quad (1)$$

where $F_{\text{bio}}(x,y,t)$ is the prescribed prior biosphere flux from the Carnegie Ames Stanford Approach Global Fire Emissions Database (CASA GFED) version 3.1 (van der Werf et al., 2010); $F_{\text{ocn}}(x,y,t)$ is the prescribed prior ocean flux based on Jacobson et al. (2007); $F_{\text{ff}}(x,y,t)$ is the prescribed prior fossil fuel flux determined using the Carbon Dioxide Information and Analysis Center (CDIAC) and the Emission Database for Global Atmospheric Research (EDGAR) inventories; $F_{\text{fire}}(x,y,t)$ is the prescribed prior fire flux derived from CASA GFED version 2 (van der Werf et al., 2006); and λ_r is the scaling factor to be optimized in the data assimilation process, corresponding to 156 ecoregions around the globe. CarbonTracker adopts a smoother window to reflect the transport speed of CO_2, which is based on the temporal relationship between the surface CO_2 flux and atmospheric CO_2 observations, as found in Bruhwiler et al. (2005) (Peters et al., 2005). For this reason, the scaling factor is optimized for five weeks of lag, which implies that the observations made in the most recent week affect the optimized surface CO_2 flux in the preceding four weeks. The optimization of the scaling factor during the data assimilation process is presented in Fig. 1. In each assimilation cycle, five weeks of analysis scaling factors are estimated by observations from the most recent week. After the fifth cycle, the scaling factor estimated by these five weeks of observations is saved as the optimized scaling factor and used to calculate the optimized surface CO_2 flux in Eq. (1). During this process, a new mean background scaling factor for the next week is calculated by the estimated mean scaling factors of

ACPD
14, 13561–13602, 2014

the previous two weeks using a simple dynamic model, as follows:

$$\lambda_t^{\mathrm{b}} = \frac{\left(\lambda_{t-2}^{\mathrm{a}} + \lambda_{t-1}^{\mathrm{a}} + \lambda^{\mathrm{p}}\right)}{3}, \tag{2}$$

where λ_t^{b} is a prior mean scaling factor for the new analysis week; $\lambda_{t-2}^{\mathrm{a}}$ and $\lambda_{t-1}^{\mathrm{a}}$ are posterior mean scaling factors estimated two weeks and one week previous, respectively; and λ^{p} is a prior value fixed as 1. Thus, the information from the previous observations is included in λ_t^{b}.

The TM5 model (Krol et al., 2005) is used as a transport model that calculates model CO_2 concentrations corresponding to the observed CO_2 concentrations. The TM5 model uses the surface CO_2 fluxes calculated from Eq. (1) and the ECMWF meteorological field to calculate model CO_2 concentrations and is used as the observation operator, which will be explained in Sect. 2.2.

2.2 Ensemble Kalman Filter

The EnKF data assimilation method used in CarbonTracker is the ensemble square root filter (EnSRF) suggested by Whitaker and Hamill (2002). The analysis equation for data assimilation is expressed as

$$\boldsymbol{x}^{\mathrm{a}} = \mathbf{K}\boldsymbol{y}^{\mathrm{o}} + (\mathbf{I}_n - \mathbf{KH})\boldsymbol{x}^{\mathrm{b}}, \tag{3}$$

where $\boldsymbol{x}^{\mathrm{a}}$ is the n-dimensional analysis (posterior) state vector; $\boldsymbol{y}^{\mathrm{o}}$ is the p-dimensional observation vector; $\mathbf{K}$ is the $n \times p$ dimensional Kalman gain; $\mathbf{I}_n$ is the identical matrix; $\mathbf{H}$ is the linearized observation operator, which transforms the information in the model space to the information in the observation space; and $\boldsymbol{x}^{\mathrm{b}}$ is the background state vector. In EnSRF, the ensemble mean and perturbed state vectors are updated independently using the following equations:

$$\boldsymbol{x}^{\mathrm{a}} = \boldsymbol{x}^{\mathrm{b}} + \mathbf{K}(\boldsymbol{y}^{\mathrm{o}} - \mathbf{H}\boldsymbol{x}^{\mathrm{b}}), \tag{4}$$

$$x_i'^{\mathrm{a}} = x_i'^{\mathrm{b}} - \tilde{\mathbf{k}}\mathbf{H}x_i'^{\mathrm{b}}, \tag{5}$$

ACPD
14, 13561–13602, 2014

where $\boldsymbol{x}^{\mathrm{a}}$ and $\boldsymbol{x}^{\mathrm{b}}$ are mean state vectors of the analysis and background, respectively, and $x_i'^{\mathrm{a}}$ and $x_t'^{\mathrm{b}}$ are perturbation state vectors of the analysis and background, respectively. In CarbonTracker, the state vector corresponds to the scaling factor, as described in Sect. 2.1. $\mathbf{K}$ and the reduced Kalman gain, $\tilde{\mathbf{k}}$, are defined as

$$\mathbf{K} = (\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}})(\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}, \tag{6}$$

$$\tilde{\mathbf{k}} = \mathbf{K} \cdot \alpha, \tag{7}$$

where $\mathbf{P}^{\mathrm{b}}$ is the background error covariance; $\mathbf{R}$ is the observation error covariance, which is predefined at each observation site; and α is a scalar value that varies whenever each observation is used in the analysis process and is calculated as

$$\alpha = \left(1 + \sqrt{\frac{\mathbf{R}}{\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}}}\right)^{-1}, \tag{8}$$

$\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}$ and $\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}$ in Eqs. (6) and (8) can be calculated as

$$\mathbf{P}\mathbf{H}^{\mathrm{T}} \approx \frac{1}{m-1}\left(x_1', x_2', \ldots, x_m'\right) \cdot \left(\mathbf{H}x_1', \mathbf{H}x_2', \ldots, \mathbf{H}x_m'\right)^{\mathrm{T}}, \tag{9}$$

$$\mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}} \approx \frac{1}{m-1}\left(\mathbf{H}x_1', \mathbf{H}x_2', \ldots, \mathbf{H}x_m'\right) \cdot \left(\mathbf{H}x_1', \mathbf{H}x_2', \ldots, \mathbf{H}x_m'\right)^{\mathrm{T}}, \tag{10}$$

where m is the number of ensembles.

To reduce the sampling error and filter divergence due to the underestimation of background error covariance in EnSRF, the covariance localization method is used (Houtekamer and Mitchell, 2001). Because the physical distance between the scaling factors cannot be defined in CarbonTracker, correlations between the ensemble of the scaling factor and the ensemble of the model CO_2 concentration are calculated, and a statistical significance test is performed on the correlations. Then, the Kalman gain

ACPD
14, 13561–13602, 2014

which has an insignificant statistical value is set to zero. This type of localization is applied to all observation sites except for Marine Boundary Layer (MBL) sites, because the observations at MBL sites are considered to include information on large footprints of flux signals (Peters et al., 2007).

2.3 Influence matrix

The influence matrix for EnKF is calculated as in Liu et al. (2009). The projection of Eq. (1) onto the observation space becomes

$$\mathbf{H}\boldsymbol{x}^{\mathrm{a}} = \boldsymbol{y}^{\mathrm{a}} = \mathbf{HK}\boldsymbol{y}^{\mathrm{o}} + (\mathbf{I}_p - \mathbf{HK})\boldsymbol{y}^{\mathrm{b}}, \tag{11}$$

where $\boldsymbol{y}^{\mathrm{a}}$ is the analysis value in the observation space and the projection of the state vector $\boldsymbol{x}^{\mathrm{a}}$ on the observation space. The influence matrix $\mathbf{S}^{\mathrm{o}}$, representing the sensitivity of the analysis state vector $\boldsymbol{y}^{\mathrm{a}}$ to the observation vector $\boldsymbol{y}^{\mathrm{o}}$ (i.e., analysis sensitivity to observation) in the observation space, is calculated as follows:

$$\mathbf{S}^{\mathrm{o}} = \frac{\partial \boldsymbol{y}^{\mathrm{a}}}{\partial \boldsymbol{y}^{\mathrm{o}}} = \mathbf{K}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} = \mathbf{R}^{-1}\mathbf{H}\mathbf{P}^{\mathrm{a}}\mathbf{H}^{\mathrm{T}}, \tag{12}$$

where $\mathbf{S}^{\mathrm{o}}$ is proportional to the analysis error covariance and is inversely proportional to the observation error covariance. By contrast, the analysis sensitivity to background is

$$\mathbf{S}^{\mathrm{b}} = \frac{\partial \boldsymbol{y}^{\mathrm{a}}}{\partial \boldsymbol{y}^{\mathrm{b}}} = \frac{\partial \boldsymbol{y}^{\mathrm{a}}}{\partial (\mathbf{H}\boldsymbol{x}^{\mathrm{b}})} = \mathbf{I}_p - \mathbf{K}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} = \mathbf{I}_p - \mathbf{S}^{\mathrm{o}}, \tag{13}$$

where $\boldsymbol{y}^{\mathrm{b}}$ is the projection of the background on the observation space, and $\mathbf{I}_p$ is an identity matrix with the size of the number of observations. Consequently, the sum of the analysis sensitivity to observation in Eq. (12) and the analysis sensitivity to background in Eq. (13) is one.

ACPD
14, 13561–13602, 2014

Substituting Eq. (10) into Eq. (12) becomes

$$\mathbf{S}^{\mathrm{o}} = \mathbf{R}^{-1}\mathbf{H}\mathbf{P}^{\mathrm{a}}\mathbf{H}^{\mathrm{T}} = \frac{1}{m-1}\mathbf{R}^{-1}(\mathbf{H}\mathbf{X}^{\mathrm{a}})(\mathbf{H}\mathbf{X}^{\mathrm{a}})^{\mathrm{T}}, \tag{14}$$

where $\mathbf{HX}^{\mathrm{a}}$ is the analysis ensemble perturbation matrix in the observation space, and the ith column of $\mathbf{HX}^{\mathrm{a}}$ is calculated as

$$\mathbf{HX}_i^{\mathrm{a}} \cong h\left(x_i^{\mathrm{a}}\right) - \frac{1}{m}\sum_{i=1}^{m} h\left(x_i^{\mathrm{a}}\right), \tag{15}$$

where x_i^{a} is the ith analysis ensemble member; m is the number of ensembles (i.e., 150); and $h(\cdot)$ is the linear or nonlinear observation operator. More specifically, the diagonal elements of the influence matrix (i.e., self-sensitivity) are calculated as

$$\mathbf{S}^{\mathrm{o}}_{jj} = \frac{\partial \boldsymbol{y}_j^{\mathrm{a}}}{\partial \boldsymbol{y}_j^{\mathrm{o}}} = \left(\frac{1}{m-1}\right)\frac{1}{\sigma_j^2}\sum_{i=1}^{m}\left(\mathbf{HX}_i^{\mathrm{a}}\right)_j \times \left(\mathbf{HX}_i^{\mathrm{a}}\right)_j, \tag{16}$$

and the cross-sensitivity, which is the off-diagonal elements of the influence matrix, is calculated as

$$\mathbf{S}^{\mathrm{o}}_{jl} = \frac{\partial \boldsymbol{y}_l^{\mathrm{a}}}{\partial \boldsymbol{y}_j^{\mathrm{o}}} = \left(\frac{1}{m-1}\right)\frac{1}{\sigma_j^2}\sum_{i=1}^{m}\left(\mathbf{HX}_i^{\mathrm{a}}\right)_j \times \left(\mathbf{HX}_i^{\mathrm{a}}\right)_l, \tag{17}$$

where σ_j^2 is the error variance of the jth observation.

The information content (i.e., degrees of freedom for signal), which is a measure of the information extracted from the observations, is calculated by the trace of the influence matrix. As suggested by Cardinali et al. (2004), the globally averaged influence of the observations can be calculated by averaging the global self-sensitivities as

$$\mathrm{GAI} = \frac{\mathrm{tr}(\mathbf{S}^{\mathrm{o}})}{p}, \tag{18}$$

ACPD
14, 13561–13602, 2014

where p is the total number of observations used in each assimilation cycle. The partial influence of a subset of observations is calculated as

$$\mathrm{PAI} = \frac{\sum_{i \in I} S^{\mathrm{o}}_{ii}}{p_I}, \tag{19}$$

where p_I represents the number of observations in subset I, which can either be set as specific observation types or as specific vertical and horizontal domains.

2.4 Observations

The observations used in this study are surface CO_2 mole fraction data observed at sites distributed around the globe (Table 1 and Fig. 2). These data were observed by NOAA, the Commonwealth Scientific and Industrial Research Organization (CSIRO), Environment Canada (EC), the National Center for Atmospheric Research (NCAR), and Lawrence Berkeley National Laboratory (LBNL) (Masarie et al., 2011). Observations from three additional sites made by the Japan Meteorological Agency (JMA) are also used in this study. The site categories and model–data mismatch values (i.e., observation error) are shown in Table 2. The model–data mismatch is determined as the innovation χ^2 in Eq. (20) becomes one at each observation site (Peters et al., 2007).

$$\chi^2 = \frac{(\boldsymbol{y}^{\mathrm{o}} - \mathbf{H}\boldsymbol{x}^{\mathrm{b}})^2}{\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}}, \tag{20}$$

The innovation χ^2 statistics for each observation site in Asia during the experimental period are presented in Table 3. The model–data mismatch for the TAP site (Tae-ahn peninsula, South Korea; 36.73° N, 126.13° E, 20 m) was changed from the value of 7.5 ppm used in previous studies to 5 ppm because the innovation χ^2 value obtained using 5 ppm was closer to one. However, TAP was still included in the Difficult category in the statistical analysis in Sect. 3. The model–data mismatches of the three JMA sites were set to 3 ppm, as in Zhang et al. (2014).

ACPD
14, 13561–13602, 2014

2.5 Experimental framework

The surface carbon flux analysis system used in this study is based on the CarbonTracker 2010 release (CT2010). However, the system employed in this study is different from CT2010 in two aspects: first, the nesting domain of the TM5 model, with $1^\circ \times 1^\circ$ horizontal resolution, is centered in Asia rather than in North America, which enables a more detailed analysis of the surface CO_2 fluxes over Asia, as shown in Kim et al. (2014); second, as mentioned in Sect. 2.4, three new JMA observation sites are added in this system, which also enhances the analysis of surface CO_2 fluxes over Asia. The global horizontal resolution is $3^\circ \times 2^\circ$, as in CT2010. The experimental period is from 1 January 2000 to 31 December 2009. The number of ensembles is 150, and the scaling factor includes five weeks of lag, as in Peters et al. (2007, 2010) and Kim et al. (2012, 2014).

3 Results

3.1 Validation

Cardinali et al. (2004) showed that the influence matrix is calculated approximately in the four-dimensional variational data assimilation method (4DVAR) because the analysis error covariance in 4DVAR is numerically calculated by the inverse of the Hessian matrix of the cost function. If the analysis error covariance is not calculated appropriately, the self-sensitivity can show a value greater than one. In contrast, the self-sensitivity in EnKF theoretically has a value lesser than one. Nevertheless, the self-sensitivity in this study can have a value greater than one because the observation operator used has nonlinearity in calculating the transport of CO_2 concentrations. In this study, only 13 observations from the total of 76 692 observations used for data assimilation present a value greater than one. This is only 0.02 % of the total number of observations, which implies that the calculated self-sensitivity is generally valid.

ACPD
14, 13561–13602, 2014

3.2 Self-sensitivity

3.2.1 Average self-sensitivity

Because the spatial coverage and number of observations varies during the experimental period, the average self-sensitivity throughout the experimental period was analyzed to evaluate the overall characteristics of the self-sensitivity at each observation site. As in previous studies (e.g., Peters et al., 2007, 2010; Kim et al., 2014), the results for the year 2000 were excluded from the data analysis because 2000 is considered as the spin-up period.

Figure 3 shows the average self-sensitivities at each observation site during the experimental period. Different sizes of circles are used in some locations to distinguish sites at the same location or at geographically close locations. In the globe, negative correlations between the spatial density of the observation sites and the self-sensitivities are not as apparent as those reported by Cardinali et al. (2004) and Liu et al. (2009). Negative correlations between the spatial density of the observation sites and the self-sensitivities are apparent in the Northern Hemisphere (NH). In particular, some observation sites in Asia show high sensitivities and a low spatial density of observation sites. The observation sites located in deserts, remote oceans, and high altitude regions generally exhibit low sensitivities.

The average self-sensitivities of each observation site category over the globe, in the NH, Tropics, and Southern Hemisphere (SH) are shown in Fig. 4. The average global self-sensitivity is 4.8 % (Fig. 4a), which implies that the analysis extracts 4.8 % of its information from the observations and 95.2 % from the background each assimilation cycle. Although the average self-sensitivity seems low, the background contains the observation information included in the previous analysis cycle, as reported in Cardinali et al. (2004). Moreover, the surface CO_2 fluxes in CarbonTracker are optimized by five weeks of observations during the assimilation process. Therefore, the cumulative impact over five weeks would be greater than 4.8 %, which only represents the most recent week of the cycle. In the globe, the Mixed site category shows the highest average

ACPD
14, 13561–13602, 2014

self-sensitivity, and the Difficult site category shows the lowest average self-sensitivity (Fig. 4a), which is related to the model–data mismatch values shown in Table 1. The model–data mismatch for the Mixed site category is relatively low, while that of the Difficult site category is high. Although the MBL site category has the lowest model data mismatch, the MBL site category does not show the highest average self-sensitivity due to the small spread of the analysis CO_2 concentrations at MBL sites. As shown in Eq. (17), the model–data mismatch and the spread of the analysis CO_2 concentrations are two factors determining the self-sensitivity. Because MBL sites are located far from strong source and sink regions of CO_2, the spread of the analysis CO_2 concentrations at these sites is small. The average self-sensitivity in the NH is 5.3 %, which is the highest of all global regions (Fig. 4b). Similar to the global results, the average self-sensitivity is highest for the Mixed site category, while that for the Difficult site category is lowest. The average self-sensitivity in the Tropics is 3.6 % (Fig. 4c); the Mixed site category shows the highest values, but they are not significantly higher than those of other categories. In the Tropics, there is no Continuous site category. In the SH, the average self-sensitivity is 3.0 %, which is the lowest among the global regions (Fig. 4d); the MBL site category shows the highest values, and there is no Continuous site category.

3.2.2 Time series of self-sensitivity

Figure 5 shows the time series of the average self-sensitivity and number of observations around the globe and in each region. Globally, two apparent characteristics can be identified in the time series (Fig. 5a): first, the average self-sensitivity decreases as the number of observations increases, showing an inversely proportional relationship; second, there is seasonal variability in the average self-sensitivity, showing high values in summer and low values in winter. In the NH, the above two features are more apparent than in other regions (Fig. 5b). Because most of the observation sites are located in the NH, characteristics of the average global self-sensitivity are mostly determined by those in the NH. As the number of observations in the Tropics increases

ACPD
14, 13561–13602, 2014

in the late 2000s, a slight inversely proportional relationship between the average self-sensitivity and the number of observations appears in the Tropics (Fig. 5c). However, the average self-sensitivity in the Tropics does not show distinct seasonal variability. In the SH, an inverse relationship between the average self-sensitivity and the number of observations is not clearly shown (Fig. 5d), which is due to the insufficient number of observations assimilated in the SH compared with the other regions. However, the seasonal variability of the average self-sensitivity appears clearly in the SH.

Figure 6 shows the average self-sensitivity for each observation site category. Although the MBL site category has the second largest number of observations, the average self-sensitivity shows little variation with respect to time (Fig. 6a). Similarly, the average self-sensitivity for the Continental site category does not vary with respect to time (Fig. 6b). The average self-sensitivity of the Mixed site category shows distinct seasonal variation (Fig. 6c). The Continuous site category displays distinct seasonal variability in the average self-sensitivity and an inversely proportional relationship between the average self-sensitivity and the number of observations (Fig. 6d). Because Continuous sites are mostly located in North America with relatively large numbers (Fig. 2), the impact of a single observation decreases as the number of observations increases. Therefore, the inversely proportional relationships between the average self-sensitivity and the number of observations around the globe (Fig. 5a) and in the NH (Fig. 5b) are mainly attributed to the Continuous site category. The Difficult site category shows a slight inverse relationship between the average self-sensitivity and the number of observations (Fig. 6e).

3.2.3 Effect of the ensemble spread of the model surface CO_2 flux on the average self-sensitivity

Despite the inversely proportional relationship between the self-sensitivity and the number of observations in the NH time series (Fig. 5a), the average self-sensitivity in the NH is higher than in the other regions (Fig. 4). In addition, the average self-sensitivities in the NH and SH are greater in summer than in winter (Fig. 5). The above

ACPD
14, 13561–13602, 2014

two characteristics imply that another factor in addition to the number of observations affects the self-sensitivity. As briefly mentioned in Sect. 3.2.1, another factor that affects the self-sensitivity is the spread of the analysis CO_2 concentrations. Therefore, the average standard deviations of surface CO_2 fluxes are evaluated in Fig. 7 to investigate the influence of the surface CO_2 flux uncertainties on the seasonal and regional characteristics of the self-sensitivities. 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 (i.e., $\mathbf{HX}^a$ in Eq. 16) by the transport model. The uncertainties of the background surface CO_2 fluxes over the terrestrial portion of the NH are high in summer months (i.e., June, July, and August: JJA) (Fig. 7a) compared with those in winter months (i.e., December, January, and February: DJF) (Fig. 7b). Due to the high surface CO_2 flux uncertainties in North America (Fig. 7a), the self-sensitivities in North America are not lower than those in the other regions (Fig. 3), regardless of the large number of observations in this region. By contrast, despite the high uncertainties of the surface CO_2 fluxes in the Eurasian Boreal region, the self-sensitivities in this region cannot be evaluated owing to the absence of observations. Instead, the self-sensitivities of the observation sites near the Eurasian Boreal region show high values (Fig. 3).

The uncertainties of the optimized biosphere and ocean fluxes by one week of observations, shown in Fig. 7c and d, are reduced compared with those of the background fluxes, shown in Fig. 7a and b. The magnitude of the reduction of the surface CO_2 flux uncertainties in North America is relatively greater than in other regions, which is consistent with the greater self-sensitivities found in North America. By contrast, when using five weeks of observations, the magnitude of the reduction of the surface CO_2 flux uncertainties is greater in Asia than in North America (Fig. 7e and f).

Therefore, the surface CO_2 flux uncertainty is one of the components to determine the magnitude of self-sensitivities, and the seasonal variability of the surface CO_2 flux uncertainties leads to the seasonal variation of the self-sensitivities.

ACPD
14, 13561–13602, 2014

3.3 Information content

3.3.1 Average information content

Figure 8 shows the average information content at each observation site during the experimental period. This value was calculated by averaging the ratio of information contents for each cycle at each site during the experimental period. Note that this average value is not the amount of information content extracted from observations but rather the relative ratio of each site's information content, normalized by the total influence of all observations. Because the magnitude of the 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. Therefore, the observation sites located in North America and Asia generally show high average information content.

To investigate the distribution of the information content in each region, histograms of the average information content around the globe and in the NH, Tropics, and SH were generated (Fig. 9). The average information content was 80.2 % in the NH, 13.3 % in the Tropics, and 6.5 % in the SH, which implies that the observations in the NH are the most informative. This is due to the large number of observations and high self-sensitivities in the NH. Around the globe, the most informative observation site category is the Continuous category (Fig. 9a). The MBL, Continental, and Mixed site categories show a similar magnitude of information content, but the Difficult site category shows the lowest information content. In the NH and for the globe as a whole, the Continuous site category is the most informative (Fig. 9b). In the current CarbonTracker system, the observation sites of the Continuous site category are mainly located in North America, except for the three JMA sites, which are located in Asia. Therefore, most of the information extracted from the Continuous site category is used to constrain the surface CO_2 fluxes of North America. In the Tropics, the MBL and Mixed site categories provide the most information (Fig. 9c). In the SH, the MBL site category provides the most

ACPD
14, 13561–13602, 2014

information, but information extracted from the Continental, Mixed, and Difficult site categories is rare (Fig. 9d).

3.3.2 Time series of information content

Figure 10 shows the time series of the weekly averaged information content for each site category in each region. In the globe, the proportion of the information content of the Continuous site category increases steadily over time (Fig. 10a), which is associated with the steady increase in the number of observations of the Continuous site category over time. In the NH, the increase of the proportion of the information content and the number of observations of the Continuous site category is more readily apparent (Fig. 10b). Because of the high self-sensitivity in summer in the NH, the proportion of the information content of the Continuous category in the NH is greater in summer than in winter. In the Tropics, the MBL and Mixed site categories provide the most information, while the Difficult site category provides limited information from 2004 onward (Fig. 10c) because, after this date, observations from only one Difficult observation site (Bukit Kotobang (BKT), Indonesia: 0.2° S, 100.32° E, 864 m) are used in the data assimilation. In the SH, most information is extracted from observations made in the MBL site category (Fig. 10d). Because the number of observations in the SH is much lower than in the other regions, the information content extracted from the observations made in this region is also lower. The information content in summer is greater than in winter in the SH owing to the seasonal variability in self-sensitivity.

To investigate the regional distribution of the information content in the NH, the time series of the information contents in Asia, North America, and Europe are shown in Fig. 11. The information content in North America is greater than that in the other regions because the self-sensitivities are high and the number of observations increases with time in North America. However, the rate of increase in the information content is lower than that of the number of observations because self-sensitivity decreases as the number of observations increases in North America.

ACPD
14, 13561–13602, 2014

3.3.3 Relationship between the information content and the optimized flux

Because CarbonTracker is a system that optimizes the surface CO_2 flux using measurements of surface CO_2 concentrations, the effect of the observations on the optimized surface CO_2 fluxes is important. To investigate the relationship between the information content and the optimized surface CO_2 fluxes, the root mean square differences (RMSDs) between the optimized surface CO_2 fluxes and the background fluxes were calculated (Fig. 12). The surface CO_2 fluxes predicted by the dynamic model in Eq. (2) (i.e., background) show a high RMSD in the NH, with the highest values in North America, followed by Asia (Fig. 12a). In terms of seasonal variation, the impact of the observations in JJA is greater than in DJF (Fig. 12a and b). The large difference between the prior fluxes and the surface CO_2 fluxes predicted by the dynamic model implies that the assimilation of previous observations substantially affects the results. The RMSD of the analyzed surface CO_2 fluxes constrained by one week of observations from the background fluxes in JJA is greater in the NH compared with the other regions. The JJA RMSD value for North America (especially in the mid-continental region of the US) is the highest in the NH (Fig. 12c). Although the RMSD of North America in DJF is lower than that in JJA, the RMSD of North America is still greater than that of other regions in DJF (Fig. 12d). The regions with a high average information content are consistent with the regions with a high RMSD (compare Figs. 8 and 12), which implies that the observations from North America provide more information in the first cycle than those from other regions because the observations in North America are characterized by high self-sensitivities and abundant observations. By contrast, the RMSD values obtained in the first cycle in other regions are not as high as those in North America. The RMSD in Asia and other regions increases after five weeks of optimization (Fig. 12e and f). In particular, the magnitude of the RMSD in the Eurasian Boreal region increases after five weeks of optimization (Fig. 12e), which implies that, by the transport of the CO_2 concentrations, the observation information from remote regions affects the optimization of the surface CO_2 fluxes in the Eurasian Boreal region. This

ACPD
14, 13561–13602, 2014

remote influence is due to the absence of observations in this region. In addition, the five-week assimilation lag is effective in optimizing the surface CO_2 flux in this region. Therefore, a longer, smoother window is necessary to optimize the surface CO_2 flux in Asia, where there are sparse observations; this implies that in the current version of CarbonTracker, the uncertainty of the surface CO_2 flux in Asia can be reduced when using a longer, smoother window than that used for North America.

4 Summary and conclusion

In this study, the effect of observations of CO_2 concentrations on the optimized surface CO_2 fluxes in CarbonTracker was evaluated by calculating the influence matrix for a 10 year period from 2000 to 2009. CarbonTracker is a system used to optimize the surface CO_2 flux using EnKF as a data assimilation algorithm. Most of the calculated influence values were in the range of the theoretical limit, from 0 to 1, which makes it possible to objectively diagnose the performance of the data assimilation system used in this study.

The average global self-sensitivity is 4.8 %, which implies that the impact of the background on the optimized flux is 95.2 %. The value of 4.8 % obtained in CarbonTracker is lower than the 15 % value obtained from NWP models, as reported by Cardinali et al. (2004) and Liu et al. (2009). However, as indicated by Cardinali et al. (2004), the background fluxes predicted by the dynamic model already include information extracted from earlier observations used in previous cycles. Because the state vector used in CarbonTracker includes five weeks of lag, the cumulative impact of the observations on the analysis is greater than the impact calculated for a single assimilation cycle. The large cumulative impact is confirmed by the RMSD of the surface CO_2 fluxes associated with each assimilation process.

The self-sensitivity and spatial coverage of the observation sites are inversely correlated in the NH, while these factors are not apparently related in the Tropics and SH. The lower correlation between the self-sensitivity and the spatial coverage of the

ACPD
14, 13561–13602, 2014

observation sites in the Tropics and SH is attributed to either the sparseness of the observation sites or the locations of the observation sites which are not appropriate for detecting the variability of CO_2 concentrations with a high temporal resolution but are appropriate for detecting the global trend of the background CO_2 concentrations. By contrast, the observation sites near the Eurasian Boreal region show high self-sensitivity because there are no available observations in the Eurasian Boreal region.

The self-sensitivity time series is characterized by seasonal variations. In both hemispheres, the self-sensitivity in summer is greater than in winter, which is clearly evident in the Mixed and Continuous site categories and is associated with the background surface CO_2 flux uncertainties. The number of observations used in data assimilation increases over time, which causes the average self-sensitivities to decrease. The decreasing trend of the self-sensitivity over time for the Continuous site observations in North America may indicate the limited impact of additional observations in this region. Schuh et al. (2013) reported that additional tower measurements (i.e., observations in the Continuous site category) in the Corn Belt region of the US did not significantly alter the surface CO_2 flux estimates for 2008, which is consistent with the low self-sensitivity detected over North America in the same period. Therefore, under the current CarbonTracker framework, to obtain the beneficial effect of additional observations on the surface CO_2 flux analysis, new observations should be added in regions with a low spatial density of observation sites (e.g., Asia).

The observation sites with a high average self-sensitivity and a small number of observations show low average information content, whereas the observation sites with a low average self-sensitivity and a large number of observations show high average information content. Therefore, the Continuous site category shows high average information content. In general, the information extracted from observations is concentrated in the NH, especially in North America. A strong correlation exists between the information content and the optimized surface CO_2 fluxes. The high information content found in regions with a large number of observations implies that much of the information is extracted from observations, and as a result, the fluxes are optimized quickly

ACPD
14, 13561–13602, 2014

in a relatively short period. However, the surface CO_2 fluxes in regions with no local observation sites (e.g., Siberia) are optimized by remote observations during relatively long assimilation windows with a lag.

The effect of various observations on the analyzed surface CO_2 fluxes can be calculated using the method suggested in this study. More CO_2 observations become available in data assimilation for estimating the surface CO_2 fluxes. These additional sources include CONTRAIL data, which are aircraft observations (Machida et al., 2008); column-averaged CO_2 concentrations retrieved from the Japanese Greenhouse gases Observing SATellite (GOSAT) (Yokoda et al., 2009); and data from the Total Carbon Column Observing Network (TCCON), which are observed by ground-based Fourier Transform Spectrometers (Wunch et al., 2011). As a next step, the impact of various observations on the optimization of surface CO_2 fluxes can be evaluated using the method suggested in this study.

Acknowledgements. This study was funded by the Korea Meteorological Administration Research and Development Program under the Grant CATER 2012-3032.

References

Baker, D. F., Doney, S. C., and Schimel, D. S.: Variational data assimilation for atmospheric CO_2, Tellus B, 58, 359–365, 2006.

Baker, D. F., Bösch, H., Doney, S. C., O'Brien, D., and Schimel, D. S.: Carbon source/sink information provided by column CO_2 measurements from the Orbiting Carbon Observatory, Atmos. Chem. Phys., 10, 4145–4165, doi:10.5194/acp-10-4145-2010, 2010.

Basu, S., Guerlet, S., Butz, A., Houweling, S., Hasekamp, O., Aben, I., Krummel, P., Steele, P., Langenfelds, R., Torn, M., Biraud, S., Stephens, B., Andrews, A., and Worthy, D.: Global CO_2 fluxes estimated from GOSAT retrievals of total column CO_2, Atmos. Chem. Phys., 13, 8695–8717, doi:10.5194/acp-13-8695-2013, 2013.

Bruhwiler, L. M. P., Michalak, A. M., Peters, W., Baker, D. F., and Tans, P.: An improved Kalman Smoother for atmospheric inversions, Atmos. Chem. Phys., 5, 2691–2702, doi:10.5194/acp-5-2691-2005, 2005.

ACPD
14, 13561–13602, 2014

Cardinali, C., Pezzulli, S., and Andersson, E.: Influence-matrix diagnostic of a data assimilation system, Q. J. Roy. Meteor. Soc., 130, 2767–2786, 2004.
Chatterjee, A., Michalak, A. M., Anderson, J. L., Mueller, K. L., Yadav, V.: Toward reliable ensemble Kalman filter estimates of CO_2 fluxes, J. Geophys. Res., 117, D22306, doi:10.1029/2012JD018176, 2012.
Chevallier, F., Fisher, M., Peylin, P., Seerrar, S., Bousque, P., Bréon, F.-M., Chédin, A., and Ciais, P.: Inferring CO_2 sources and sinks from satellite observations: method and application to TOVS data, J. Geophys. Res., 110, D24309, doi:10.1029/2005JD006390, 2005.
Chevallier, F., Engelen, R. J., Carouge, C., Conway, T. J., Peylin, P., Pickett-Heaps, C., Ramonet, M., Rayner, P. J., and Xueref-Remy, I.: AIRS-based versus flask-based estimation of carbon surface fluxes, J. Geophys. Res., 114, D20303, doi:10.1029/2009JD012311, 2009a.
Chevallier, F., Maksyutov, S., Bousquet, P., Bréon, F.-M., Saito, R., Yoshida, Y., and Yokota, T.: On the accuracy of the CO_2 surface fluxes to be estimated from the GOSAT observations, Geophys. Res. Lett., 36, L19807, doi:10.1029/2009GL040108, 2009b.
Ciais, P., Rayner, P., Chevallier, F., Bousquet, P., Logan, M., Peylin, P., and Ramonet, M.: Atmospheric inversions for estimating CO_2 fluxes: methods and perspectives, Climate Change, 103, 69–92, 2010.
Engelen, R. J. and Stephens, G. L.: Information content of infrared satellite sounding measurements with respect to CO_2, J. Appl. Meteorol., 43, 373–378, 2004.
Feng, L., Palmer, P. I., Bösch, H., and Dance, S.: Estimating surface CO_2 fluxes from space-borne CO_2 dry air mole fraction observations using an ensemble Kalman Filter, Atmos. Chem. Phys., 9, 2619–2633, doi:10.5194/acp-9-2619-2009, 2009.
Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y. H., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B. C., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C. W.: Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models, Nature, 415, 626–630, 2002.
Houtekamer, P. L. and Mitchell, H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation, Mon. Weather Rev., 129, 123–137, 2001.
Jacobson, A. R., Fletcher, S. E. M., Gruber, N., Sarmiento, J. L., and Gloor, M.: A joint atmosphere–ocean inversion for surface fluxes of carbon dioxide: 2. Regional results, Global Biogeochem. Cy., 21, GB1019, doi:10.1029/2006GB002703, 2007.

ACPD
14, 13561–13602, 2014

Kang, J. S., Kalnay, E., Liu, J., Fung, I., Miyoshi, T., and Ide, K.: "Variable localization" in an ensemble Kalman filter: application to the carbon cycle data assimilation, J. Geophys. Res., 116, D09110, doi:10.1029/2010JD014673, 2011.

Kang, J. S., Kalnay, E., Miyoshi, T., Liu, J., and Fung, I.: Estimation of surface carbon fluxes with an advanced data assimilation methodology, J. Geophys. Res., 117, D24101, doi:10.1029/2012JD018259, 2012.

Kim, J., Kim, H. M., and Cho, C.-H.: Application of Carbon Tracking System based on ensemble Kalman Filter on the diagnosis of Carbon Cycle in Asia, Atmosphere, 22, 415–447, 2012 (in Korean with English abstract).

Kim, J., Kim, H. M., and Cho, C.-H.: The effect of optimization and the nesting domain on carbon flux analyses in Asia using a carbon tracking system based on the ensemble Kalman filter, Asia-Pac. J. Atmos. Sci., 50, 327–344, doi:10.1007/s13143-014-0020-y, 2014.

Krol, M., Houweling, S., Bregman, B., van den Broek, M., Segers, A., van Velthoven, P., Peters, W., Dentener, F., and Bergamaschi, P.: The two-way nested global chemistry-transport zoom model TM5: algorithm and applications, Atmos. Chem. Phys., 5, 417–432, doi:10.5194/acp-5-417-2005, 2005.

Liu, J., Kalnay, E., Miyoshi, T., and Cardinali, C.: Analysis sensitivity calculation in an ensemble Kalman filter, Q. J. Roy. Meteor. Soc., 135, 1842–1851, 2009.

Lokupitiya, R., S., Zupanski, D., Denning, A. S., Kawa, S. R., Gurney, R., and Zupanski, M.: Estimation of global CO_2 fluxes at regional scale using the maximum likelihood ensemble filter, J. Geophys. Res., 113, D20110, doi:10.1029/2007JD009679, 2008.

Machida, T., Matsueda, H., Sawa, Y., Nakagawa, Y., Hirotani, K., Kondo, N., Goto, K., Nakazawa, T., Ishikawa, K., and Ogawa, T.: Worldwide measurements of atmospheric CO_2 and other trace gas species using commercial airlines, J. Atmos. Ocean. Tech., 25, 1744–1754, doi:10.1175/2008JTECHA1082.1, 2008.

Masarie, K. A., Pétron, G., Andrews, A., Bruhwiler, L., Conway, T. J., Jacobson, A. R., Miller, J. B., Tans, P. P., Worthy, D. E., and Peters, W.: Impact of CO_2 measurement bias on CarbonTracker surface flux estimates, J. Geophys. Res., 116, D17305, doi:10.1029/2011JD016270, 2011.

Meirink, J. F., Bergamaschi, P., Frankenberg, C., d'Amelio, M. T. S., Dlugokencky, E. J., Gatti, L. V., Houweling, S., Miller, J. B., Röckmann, T., Villani, M. G., and Krol, M. C.: Four-dimensional variational data assimilation for inverse modeling of atmospheric methane

ACPD
14, 13561–13602, 2014

emissions: analysis of SCIAMACHY observations, J. Geophys. Res., 113, D17301, doi:10.1029/2007JD009740, 2008.
Miyazaki, K., Maki, T., Patra, P., and Nakazawa, T.: Assessing the impact of satellite, aircraft, and surface observations on CO_2 flux estimation using an ensemble-based 4-D data assimilation system, J. Geophys. Res., 116, D16306, doi:10.1029/2010JD015366, 2011.
Talagrand, O.: Assimilation of observations, an introduction, J. Meteorol. Soc. Jpn., 75, 191–209, 1997.
Peters, W., Miller, J. B., Whitaker, J., Denning, A. S., Hirsch, A., Krol, M. C., Zupanski, D., Bruhwiler, L., and Tans, P. P.: An ensemble data assimilation system to estimate CO_2 surface fluxes from atmospheric trace gas observations, J. Geophys. Res., 110, D24304, doi:10.1029/2005JD006157, 2005.
Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Petron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, P. Natl. Acd. Sci. USA, 104, 18925–18930, 2007.
Peters, W., Krol, M. C., van der Werf, G. R., Houweling, S., Jones, C. D., Hughes, J., Schaefer, K., Masarie, K. A., Jacobson, A. R. Miller, J. B., Cho, C. H., Ramonet, M., Schmidt, M., Ciattaglia, L., Apadula, F., Heltai, D., Meinhardt, F., di Sarra, A. G., Piacentino, S., Sferlazzo, D., Aalto, T., Hatakka, J., Ström, J., Haszpra, L., Meijer, H. A. J., van der Laan, S., Neubert, R. E. M., Jordan, A., Rod, X., Morguí, J. A., Vermeulen, A. T., Popa, E., Rozanski, K., Zimnoch, M., Manning, A. C., Leuenberger, M., Uglietti, C., Dolman, A. J., Ciais, P. Heimann, M., and Tans, P. P.: Seven years of recent European net terrestrial carbon dioxide exchange constrained by atmospheric observations, Glob. Change Biol., 16, 1317–1337, doi:10.1111/j.1365-2486.2009.02078.x, 2010.
Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., Niwa, Y., Patra, P. K., Peters, W., Rayner, P. J., Rödenbeck, C., van der Laan-Luijkx, I. T., and Zhang, X.: Global atmospheric carbon budget: results from an ensemble of atmospheric CO_2 inversions, Biogeosciences, 10, 6699–6720, doi:10.5194/bg-10-6699-2013, 2013.
Rodgers, C. D.: Inverse Methods for Atmospheric Sounding, World Scientific, London, 2000.
Schuh, A. E., Lauvaux, T., West, T. O., Denning, A. S., Dvais, K. J., Miles, N., Richardson, S., Uliasz, M., Lokupitiya, E., Cooley, D., Andrews, A., and Ogle, S.: Evaluating atmospheric CO_2

ACPD
14, 13561–13602, 2014

inversions at multiple scales over a highly inventoried agricultural landscape, Global Change. Biol., 19, 1424–1439, doi:10.1111/gcb.12141, 2013.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Kasibhatla, P. S., and Arellano Jr., A. F.: Interannual variability in global biomass burning emissions from 1997 to 2004, Atmos. Chem. Phys., 6, 3423–3441, doi:10.5194/acp-6-3423-2006, 2006.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmos. Chem. Phys., 10, 11707–11735, doi:10.5194/acp-10-11707-2010, 2010.

Whitaker, J. S. and Hamill, T. M.: Ensemble data assimilation without perturbed observations, Mon. Weather Rev., 130, 1913–1924, 2002.

Wunch, D., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Stephens, B. B., Fischer, M. L., Uchino, O., Abshire, J. B., Bernath, P., Biraud, S. C., Blavier, J.-F. L., Boone, C., Bowman, K. P., Browell, E. V., Campos, T., Connor, B. J., Daube, B. C., Deutscher, N. M., Diao, M., Elkins, J. W., Gerbig, C., Gottlieb, E., Griffith, D. W. T., Hurst, D. F., Jiménez, R., Keppel-Aleks, G., Kort, E. A., Macatangay, R., Machida, T., Matsueda, H., Moore, F., Morino, I., Park, S., Robinson, J., Roehl, C. M., Sawa, Y., Sherlock, V., Sweeney, C., Tanaka, T., and Zondlo, M. A.: Calibration of the Total Carbon Column Observing Network using aircraft profile data, Atmos. Meas. Tech., 3, 1351–1362, doi:10.5194/amt-3-1351-2010, 2010.

Yokota, T., Yoshida, Y., Eguchi, N., Ota, Y., Tanaka, T., Watanabe, H., Maksyutov, S.: Global concentrations of CO_2 and CH_4 retrieved from GOSAT: first preliminary results, Sci. Online Lett. Atmos., 5, 160–163, doi:10.2151/sola.2009-041, 2010.

Zhang, H. F., Chen, B. Z., van der Laan-Luijkx, I. T., Machida, T., Matsueda, H., Sawa, Y., Fukuyama, Y., Labuschagne, C., Langenfelds, R., van der Schoot, M., Xu, G., Yan, J. W., Zhou, L. X., Tans, P. P., and Peters, W.: Estimating Asian terrestrial carbon fluxes from CONTRAIL aircraft and surface CO_2 observations for the period 2006 to 2010, Atmos. Chem. Phys. Discuss., 13, 27597–27639, doi:10.5194/acpd-13-27597-2013, 2013.

Zupanski, D., Denning, A. S., Uliasz, M., Zupanski, M., Schuh, A. E., Rayner, P. J., Peters, W., and Corbin, K. D.: Carbon flux bias estimation employing Maximum Likelihood Ensemble Filter (MLEF), J. Geophys. Res., 112, D17107, doi:10.1029/2006JD008371, 2007.

ACPD
14, 13561–13602, 2014

Table 1. Information on the observation sites used in this study. MDM represents the model–data mismatch, which is the observation error.

Site code	Location	Latitude	Longitude	Height	Laboratory	MDM
ALT_01D0	Alert, Nunavut, Canada	82.45° N	62.51° W	200 m	ESRL	1.5
ALD_06C0	Alert, Nunavut, Canada	82.45° N	62.51° W	200 m	ESRL	2.5
AMT_01C3	Argyle, Maine, US	45.03° N	68.68° W	50 m	ESRL	3
AMT_01P0	Argyle, Maine, US	45.03° N	68.68° W	50 m	ESRL	3
ASC_01D0	Ascension Island, UK	7.92° S	14.42° W	54 m	ESRL	0.75
ASK_01D0	Assekrem, Algeria	23.18° N	5.42° E	2728 m	ESRL	1.5
AZR_01D0	Terceira Island, Azores, Portugal	38.77° N	27.38° W	40 m	ESRL	1.5
BAL_01D0	Baltic Sea, Poland	55.35° N	17.22° E	3 m	ESRL	7.5
BAO_01C3	Boulder Atmospheric Observatory, Colorado, US	40.05° N	105.00° W	1584 m	ESRL	3
BAO_01P0	Boulder Atmospheric Observatory, Colorado, US	40.05° N	105.00° W	1584 m	ESRL	3
BKT_01D0	Bukit Kotobang, Indonesia	0.20° S	100.32° E	864 m	ESRL	7.5
BME_01D0	St. Davids Head, Bermuda, UK	32.27° N	64.65° E	30 m	ESRL	1.5
BMW_01D0	Tudor Hill, Bermuda, UK	32.27° N	64.88° E	30 m	ESRL	1.5
BRW_01D0	Barrow, Alaska, US	71.32° N	156.61° W	11 m	ESRL	1.5
BRW_01C0	Barrow, Alaska, US	71.32° N	156.61° W	11 m	ESRL	2.5
BSC_01D0	Black Sea, Constanta, Romania	44.17° N	28.68° E	3 m	ESRL	7.5
CBA_01D0	Cold Bay, Alaska, US	55.21° N	162.72° W	21 m	ESRL	1.5
CDL_06C0	Candle Lake, Saskatchewan, Canada	53.99° N	105.12° W	600 m	ESRL	3
CFA_02D0	Cape Ferguson, Queensland, Australia	19.28° S	147.06° E	184 m	ESRL	2.5
CGO_01D0	Cape Grim, Tasmania, Australia	40.68° S	144.69° E	94 m	ESRL	0.75
CGO_02D0	Cape Grim, Tasmania, Australia	40.68° S	144.69° E	94 m	CSIRO	0.75
CHR_01D0	Christmas Island, Republic of Kiribati	1.70° N	157.17° W	3 m	ESRL	0.75
CRZ_01D0	Crozet Island, France	46.45° S	51.85° E	120 m	ESRL	0.75
cya_02D0	Casey, Antarctica, Australia	66.28° S	110.5° E	51 m	CSIRO	0.75
EGB_06C0	Egbert, Ontario, Canada	44.23° N	79.78° W	251 m	EC	3
EIC_01D0	Easter Island, Chile	27.15° S	109.45° W	50 m	ESRL	7.5
ESP_06C0	Estevan Point, British Columbia, Canada	49.38° N	126.54° W	7 m	EC	3
ETL_06C0	East Trout Lake, Saskatchewan, Canada	54.35° N	104.98° W	492 m	EC	3
FEF_03C0	Fraser, Colorado, US	39.91° N	105.88° W	2745 m	NCAR	3
FSD_06C0	Fraserdale, Canada	49.88° N	81.57° W	210 m	EC	3
GMI_01D0	Mariana Islands, Guam	13.43° N	144.78° E	2 m	ESRL	1.5
HBA_01D0	Halley Station, Antarctica, UK	75.58° S	26.50° W	30 m	ESRL	0.75
HDP_03C0	Hidden Peak (Snowbird), Utah, US	40.56° N	111.65° W	3351 m	NCAR	3
HUN_01D0	Hegyhatsal, Hungary	46.95° N	16.65° E	248 m	ESRL	7.5
ICE_01D0	Storhofdi, Vestmannaeyjar, Iceland	63.40° N	20.29° W	118 m	ESRL	1.5
KEY_01D0	Key Biscayne, Florida, US	25.67° N	80.16° W	3 m	ESRL	2.5
KUM_01D0	Cape Kumukahi, Hawaii, US	19.52° N	154.82° W	3 m	ESRL	1.5
KZD_01D0	SaryTaukum, Kazakhstan	44.06° N	76.82° E	601 m	ESRL	2.5
KZM_01D0	Plateau Assy, Kazakhstan	43.25° N	77.88° E	2519 m	ESRL	2.5
LEF_01C3	Park Falls, Wisconsin, US	45.95° N	90.27° W	472 m	ESRL	3
LEF_01P0	Park Falls, Wisconsin, US	45.95° N	90.27° W	472 m	ESRL	3
LLB_06C0	Lac La Biche, Alberta, Canada	54.95° N	112.45° W	540 m	EC	3
MAA_02D0	Mawson Station, Antarctica, Australia	67.62° S	62.87° E	32 m	CSIRO	0.75
MHD_01D0	Mace Head, County Galway, Ireland	53.33° N	9.90° W	5 m	ESRL	2.5
MID_01D0	Sand Island, Midway, US	28.21° N	177.38° W	4 m	ESRL	1.5

ACPD
14, 13561–13602, 2014

Influence of CO_2 observations on the optimized CO_2 flux in an ensemble Kalman filter

J. Kim et al.

Table 1. Continued.

Site code	Location	Latitude	Longitude	Height	Laboratory	MDM
MKN_01D0	MT. Kenya, Kenya	0.05° S	37.30° E	3897 m	ESRL	2.5
MLO_01C0	Mauna Loa, Hawaii, US	19.54° N	155.58° W	3397 m	ESRL	0.75
MLO_01D0	Mauna Loa, Hawaii, US	19.54° N	155.58° W	3397 m	ESRL	1.5
MNM_19C0	Minamitorishima, Japan	24.29° N	153.98° E	8 m	JMA	3
MQA_02D0	Macquarie Island, Australia	54.48° S	158.97° E	12 m	CSIRO	0.75
NMB_01D0	Gobabeb, Namibia	23.58° S	15.03° E	456 m	ESRL	2.5
NWR_01D0	Niwot Ridge, Colorado, US	40.05° N	105.58° W	3523 m	ESRL	1.5
NWR_03C0	Niwot Ridge, Colorado, US	40.05° N	105.58° W	3523 m	NCAR	3
OBN_01D0	Obninsk, Russia	55.11° N	36.60° E	183 m	ESRL	7.5
OXK_01D0	Ochsenkopf, Germany	50.03° N	11.80° E	1022 m	ESRL	2.5
PAL_01D0	Pallas-Sammaltunturi, GAW Station, Germany	67.97° N	24.12° E	560 m	ESRL	2.5
POC_01D1	Pacific Ocean, N/A	0.39° S	132.43° W	10 m	ESRL	0.75
PSA_01D0	Palmer Station, Antarctica, US	64.92° S	64.00° W	10 m	ESRL	0.75
PTA_01D0	Point Arena, California, US	38.95° N	123.74° W	17 m	ESRL	7.5
RPB_01D0	Ragged Point, Barbados	13.17° N	59.43° W	45 m	ESRL	1.5
RYO_19C0	Ryori, Japan	39.03° N	141.82° E	260 m	JMA	3
SCT_01C3	Beech Island, South Carolina, US	33.41° N	81.83° W	115 m	ESRL	3
SEY_01D0	Mahe Island, Seychelles	4.67° S	55.17° E	3 m	ESRL	0.75
SGP_01D0	Southern Great Plains, Oklahoma, US	36.80° N	97.50° W	314 m	ESRL	2.5
SGP_64C3	Southern Great Plains, Oklahoma, US	36.80° N	97.50° W	314 m	ESRL	3
SHM_01D0	Shemya Island, Alaska, US	52.72° N	174.10° E	40 m	ESRL	2.5
SMO_01C0	Tutuila, American Samoa	14.25° S	170.56° W	42 m	ESRL	0.75
SMO_01D0	Tutuila, American Samoa	14.25° S	170.56° W	42 m	ESRL	1.5
SNP_01C3	Shenandoah National Park, US	38.62° N	78.35° W	1008 m	ESRL	3
SPL_01C3	Storm Peak Laboratory (Desert Research Institute), US	40.45° N	106.73° W	3210 m	NCAR	3
SPO_01C0	South Pole, Antarctica, US	89.98° S	24.80° W	2810 m	ESRL	0.75
SPO_01D0	South Pole, Antarctica, US	89.98° S	24.80° W	2810 m	ESRL	1.5
STM_01D0	Ocean Station M, Norway	66.00° N	2.00° E	0 m	ESRL	1.5
STR_01P0	Sutro Tower, San Francisco, California, US	37.76° N	122.45° W	254 m	ESRL	3
SUM_01D0	Summit, Greenland	72.57° N	38.48° W	3238 m	ESRL	1.5
SYO_01D0	Syowa Station, Antarctica, Japan	69.00° S	39.58° E	11 m	ESRL	0.75
TAP_01D0	Tae-ahn Peninsula, Republic of Korea	36.73° N	126.13° E	20 m	ESRL	5
TDF_01D0	Tierra Del Fuego, Ushuaia, Argentina	54.87° S	68.48° W	20 m	ESRL	0.75
THD_01D0	Trinidad head, California, US	41.73° N	91.35° W	107 m	ESRL	2.5
UTA_01D0	Wendover, Utah, US	39.90° N	113.72° W	1320 m	ESRL	2.5
UUM_01D0	Ulaan Uul, Mongolia	44.45° N	111.10° E	914 m	ESRL	2.5
WBI_01C3	West Branch, Iowa, US	41.73° N	91.35° W	242 m	ESRL	3
WBI_01P0	West Branch, Iowa, US	41.73° N	91.35° W	242 m	ESRL	3
WGC_01C3	Walnut Grove, California, US	38.27° N	121.49° W	0 m	ESRL	3
WGC_01P0	Walnut Grove, California, US	38.27° N	121.49° W	0 m	ESRL	3
WIS_01D0	WIS Station, Negev Desert, Israel	31.13° N	34.88° E	400 m	ESRL	2.5
WKT_01C3	Moody, Texas, US	31.32° N	97.33° W	251 m	ESRL	3
WKT_01C3	Moody, Texas, US	31.32° N	97.33° W	251 m	ESRL	3
WLG_01D0	Mt. Waliguan, Peoples Republic of China	36.29° N	100.90° E	3810 m	ESRL	1.5
WSA_06C0	Sable Island, Nova Scotia, Canada	49.93° N	60.02° E	5 m	EC	3
YON_19C0	Yonagunijima, Japan	24.47° N	123.02° E	30 m	JMA	3
ZEP_01D0	Ny-Alesund, Svalbard, Norway and Sweden	78.90° N	11.88° E	475 m	ESRL	1.5

ACPD
14, 13561–13602, 2014

Influence of CO_2 observations on the optimized CO_2 flux in an ensemble Kalman filter

J. Kim et al.

Table 2. Observation site categories and corresponding model–data mismatch values [ppm].

Observation category	Description	Observation frequency	Model–data mismatch [ppm]
Marine Boundary Layer (MBL)	Observation site close to Marine boundary layer	Once a week	0.75
Mixed land/ocean and mountain (Mixed)	Observation site located in mixed land, ocean, and mountain	Once a week	1.5
Continental	Observation site located in the continent	Once a week	2.5
Continuous	Observation site with continuous observations	Once a day	3
Difficult	Difficult	Once a week	7.5 (5.0)

ACPD
14, 13561–13602, 2014

Table 3. Information on the observation sites located in Asia, including the number of observations, number of rejected observations, MDM values, innovation χ^2 statistics, and the average bias of the model CO_2 concentrations calculated by optimized fluxes. For the TAP_01D0 site, the numbers in parentheses are values used in previous studies, and the numbers without parentheses are the modified values based on the innovation χ^2 statistics in this study.

Site name	Number of observations	Number of rejected observations	MDM	Innovation χ^2	Bias of model CO_2 concentration
BKT_01D0	207	0	7.5	0.57	−4.01
KZD_01D0	430	11	2.5	1.25	−0.4
KZM_01D0	384	9	2.5	1.22	−0.67
MNM_19C0	3304	0	3	0.16	−0.45
RYO_19C0	3149	108	3	0.53	−0.9
TAP_01D0	339	10	5	0.59	0.01
	(269)	(3)	(7.5)	(0.37)	(−0.26)
UUM_01D0	454	10	2.5	1.03	0.26
WIS_01D0	489	3	2.5	0.72	−0.15
WLG_01D0	347	10	1.5	1.14	0.04
YON_19C0	2947	8	3	0.53	−0.9

ACPD
14, 13561–13602, 2014

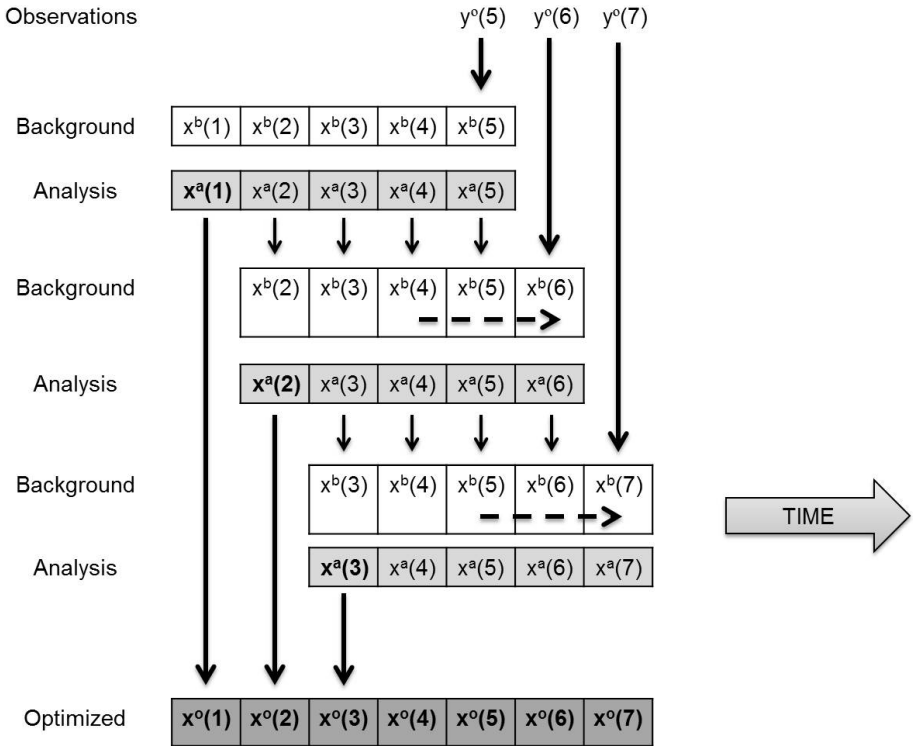

Figure 1. Schematic diagram of the assimilation process employed in CarbonTracker. In each analysis cycle, observations made within one week are used to update the state vectors with a five-week lag. The dashed line indicates how the simple dynamic model uses analysis state vectors from the previous one 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.

ACPD
14, 13561–13602, 2014

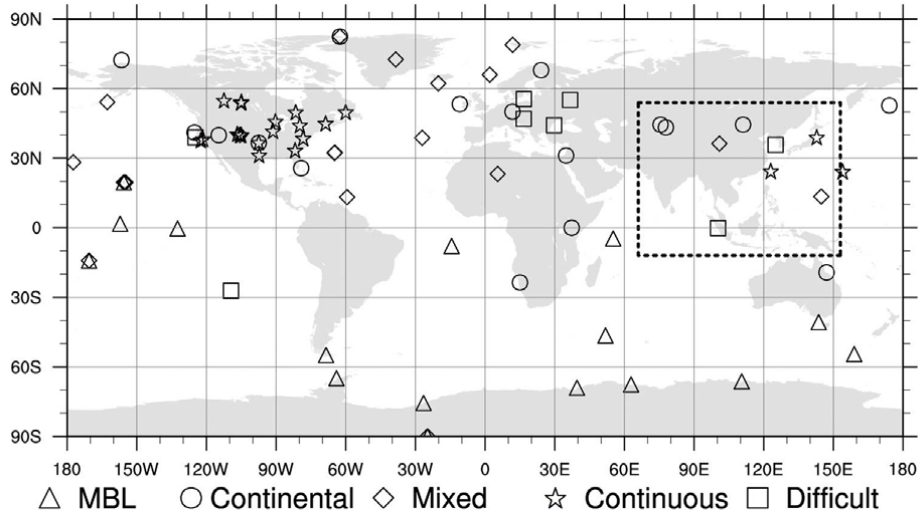

Figure 2. Observation network of CO_2 concentrations around the globe and the nested domain of the TM5 transport model over Asia (dashed box). Each observation site is assigned to different categories (△: MBL; ○: continental; ◇: mixed land/ocean and mountain; ☆: continuous; □: difficult).

ACPD
14, 13561–13602, 2014

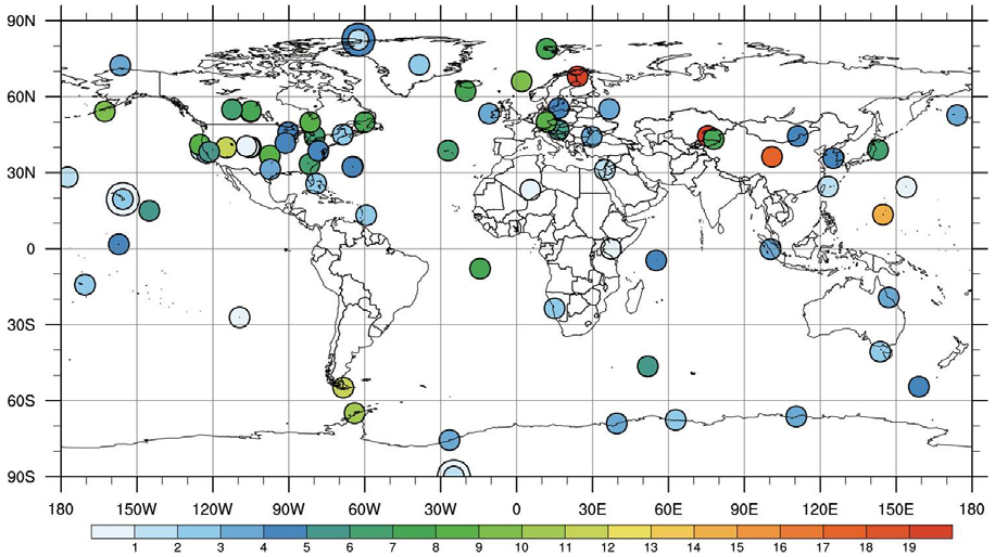

Figure 3. Average self-sensitivity at each observation site from 2000 to 2009. The overlapping observation sites at the same locations or at close locations are distinguished by different sizes of circles.

ACPD
14, 13561–13602, 2014

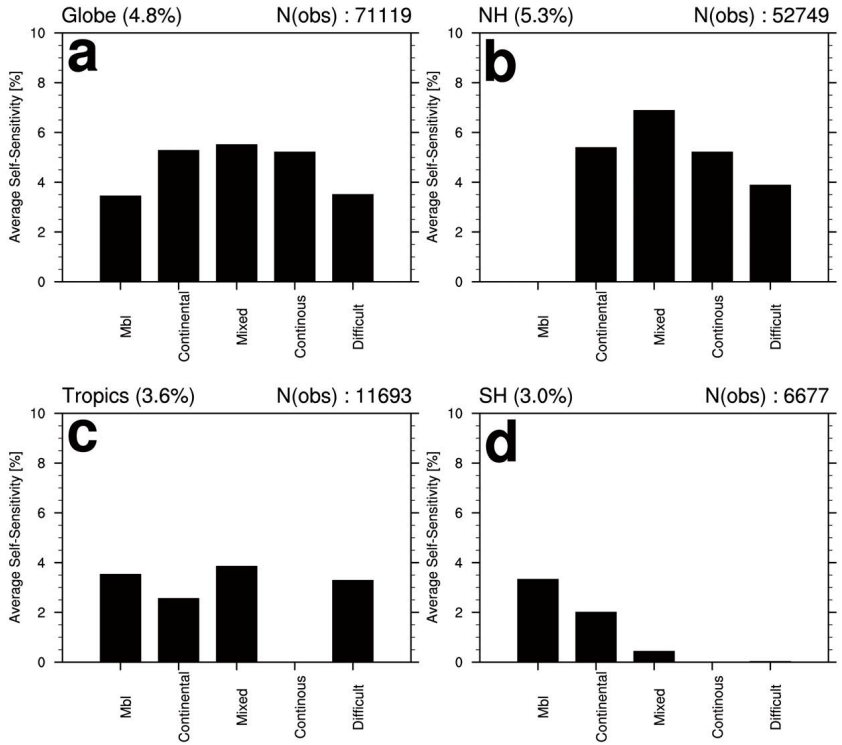

Figure 4. Histograms of the average self-sensitivity for each observation site category from 2000 to 2009 **(a)** around the globe and in the **(b)** Northern Hemisphere, **(c)** Tropics, and **(d)** Southern Hemisphere. N(obs) in the upper right corner represents the number of observations used in data assimilation.

ACPD
14, 13561–13602, 2014

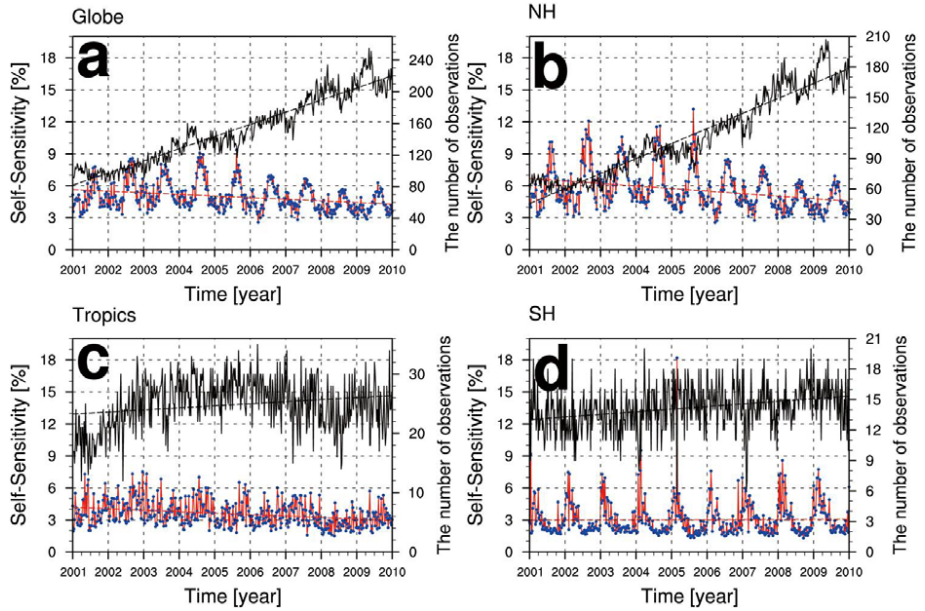

Figure 5. Time series of the average self-sensitivity (red solid line with blue dots) and the number of observations (black solid line) with a weekly temporal resolution **(a)** around the globe and in the **(b)** Northern Hemisphere, **(c)** Tropics, and **(d)** Southern Hemisphere from 2000 to 2009. The dashed lines represent the regression lines for the average self-sensitivity (red dashed line) and the number of observations (black dashed line).

ACPD
14, 13561–13602, 2014

Influence of CO_2 observations on the optimized CO_2 flux in an ensemble Kalman filter

J. Kim et al.

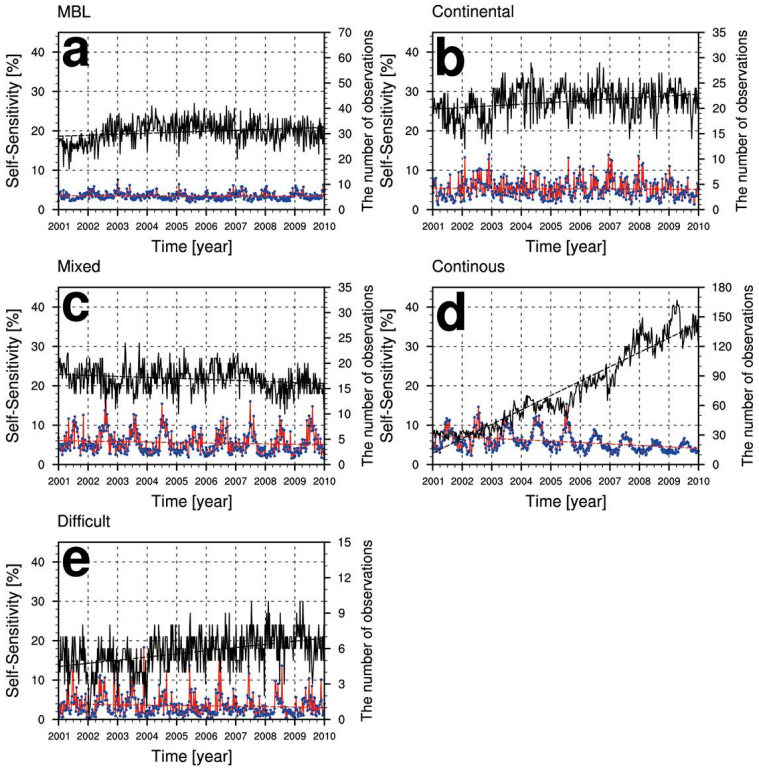

Figure 6. Time series of the average self-sensitivity (red solid line with blue dots) and the number of observations (black solid line) with a weekly temporal resolution for the **(a)** MBL, **(b)** Continental, **(c)** Mixed, **(d)** Continuous, and **(e)** Difficult observation site categories from 2000 to 2009. The dashed lines represent the regression lines for the average self-sensitivity (red dashed line) and the number of observations (black dashed line).

ACPD
14, 13561–13602, 2014

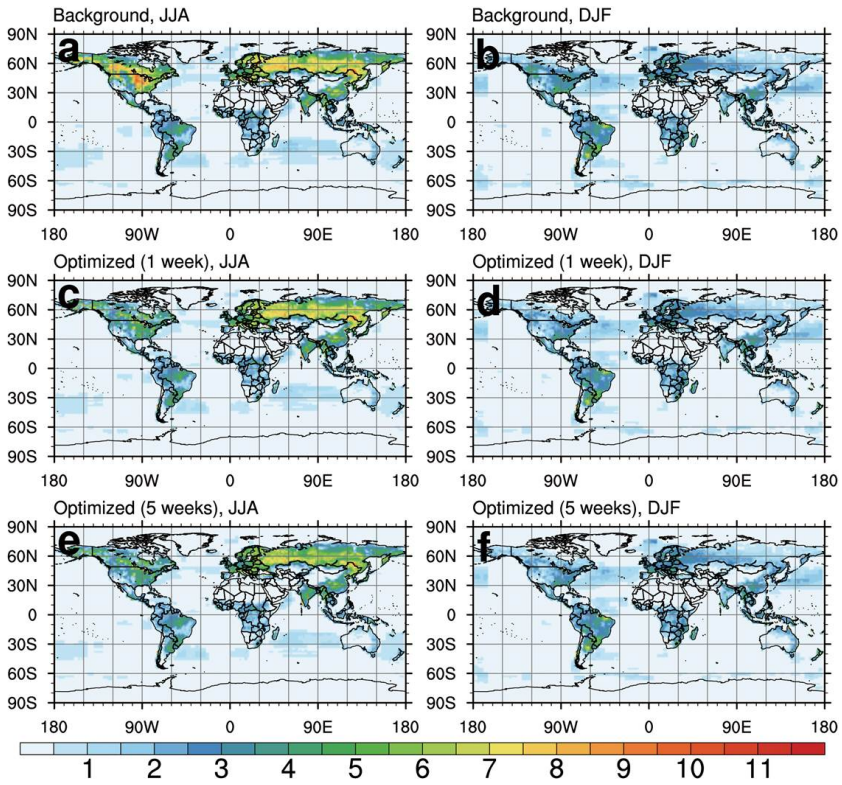

Figure 7. Average standard deviation of background biosphere and ocean fluxes in **(a)** JJA and **(b)** DJF; the posterior biosphere and ocean fluxes optimized by one-week observations in **(c)** JJA and **(d)** DJF; and the posterior biosphere and ocean fluxes optimized by five weeks of observations in **(e)** JJA and **(f)** DJF. The units are g C m^{-2} $week^{-1}$.

ACPD
14, 13561–13602, 2014

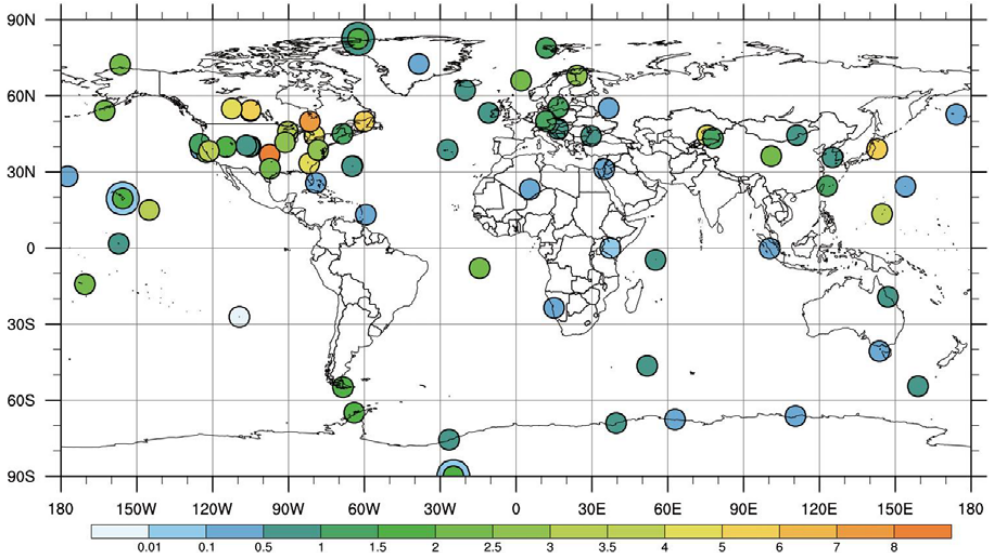

Figure 8. Average normalized information content for each observation site from 2000 to 2009. The overlapping observation sites at the same locations or at close locations are distinguished using different sizes of circles.

ACPD
14, 13561–13602, 2014

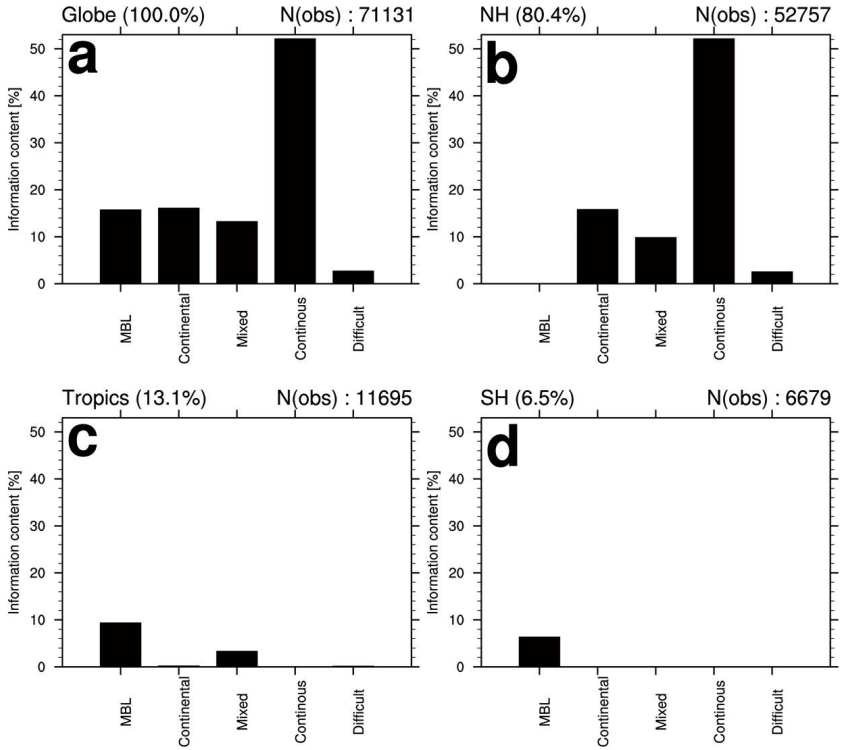

Figure 9. Histograms of the average information content for each observation site category **(a)** around the globe and in the **(b)** Northern Hemisphere, **(c)** Tropics, and **(d)** Southern Hemisphere from 2000 to 2009. N(obs) in the upper right corner represents the number of observations used in data assimilation.

ACPD
14, 13561–13602, 2014

Influence of CO_2 observations on the optimized CO_2 flux in an ensemble Kalman filter

J. Kim et al.

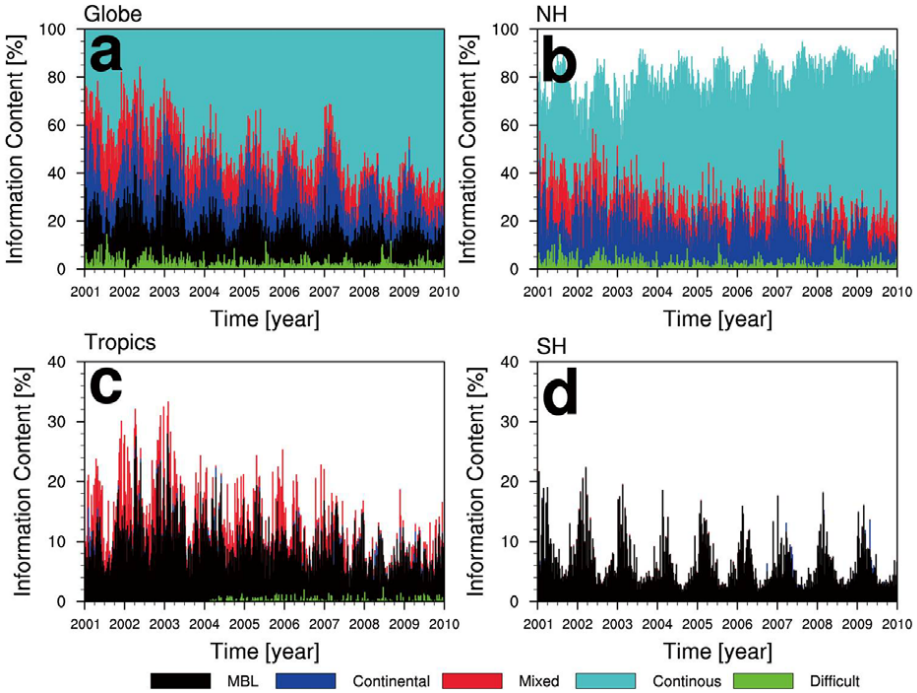

Figure 10. Time series of the average information content for each observation site category **(a)** around the globe and in the **(b)** Northern Hemisphere, **(c)** Tropics, and **(d)** Southern Hemisphere from 2000 to 2009.

ACPD
14, 13561–13602, 2014

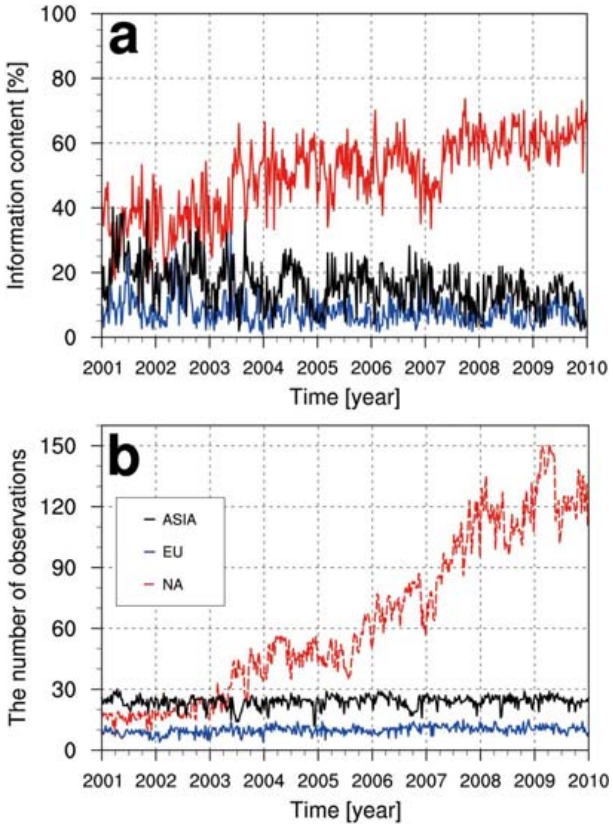

Figure 11. Times series of the **(a)** weekly averaged information content and **(b)** number of observations in Asia (black line), Europe (blue line), and North America (red line) from 2000 to 2009.

ACPD
14, 13561–13602, 2014

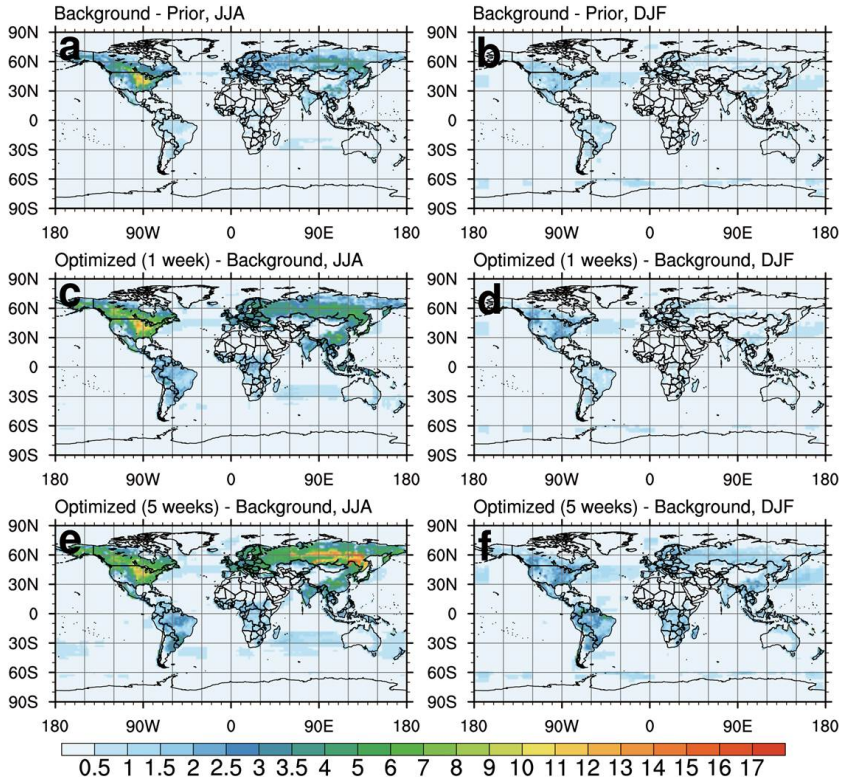

Figure 12. Root mean square difference (RMSD) between the background flux and prior flux in **(a)** JJA and **(b)** DJF; RMSD between the background flux and posterior flux optimized by one-week observations in **(c)** JJA and **(d)** DJF; and RMSD between the background flux and posterior flux optimized by five weeks of observations in **(e)** JJA and **(f)** DJF. The units are g C m^{-2} $week^{-1}$.

ACPD
14, 13561–13602, 2014