



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Estimating surface CO<sub>2</sub> fluxes from space-borne CO<sub>2</sub> dry air mole fraction observations using an ensemble Kalman Filter

**Citation for published version:**

Feng, L, Palmer, P, Boesch, H & Dance, S 2009, 'Estimating surface CO<sub>2</sub> fluxes from space-borne CO<sub>2</sub> dry air mole fraction observations using an ensemble Kalman Filter', *Atmospheric Chemistry and Physics*, vol. 9, no. 8, pp. 2619-2633. <https://doi.org/10.5194/acp-9-2619-2009>

**Digital Object Identifier (DOI):**

[10.5194/acp-9-2619-2009](https://doi.org/10.5194/acp-9-2619-2009)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Atmospheric Chemistry and Physics

**Publisher Rights Statement:**

This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are properly cited.

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Estimating surface CO<sub>2</sub> fluxes from space-borne CO<sub>2</sub> dry air mole fraction observations using an ensemble Kalman Filter

L. Feng<sup>1</sup>, P. I. Palmer<sup>1</sup>, H. Bösch<sup>2</sup>, and S. Dance<sup>3</sup>

<sup>1</sup>School of GeoSciences, University of Edinburgh, King's Buildings, Edinburgh EH9 3JN, UK

<sup>2</sup>Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK

<sup>3</sup>Department of Mathematics and Department of Meteorology, University of Reading, Reading RG6 6BB, UK

Received: 8 September 2008 – Published in Atmos. Chem. Phys. Discuss.: 28 November 2008

Revised: 2 April 2009 – Accepted: 2 April 2009 – Published: 15 April 2009

**Abstract.** We have developed an ensemble Kalman Filter (EnKF) to estimate 8-day regional surface fluxes of CO<sub>2</sub> from space-borne CO<sub>2</sub> dry-air mole fraction observations ( $X_{\text{CO}_2}$ ) and evaluate the approach using a series of synthetic experiments, in preparation for data from the NASA Orbiting Carbon Observatory (OCO). The 32-day duty cycle of OCO alternates every 16 days between nadir and glint measurements of backscattered solar radiation at short-wave infrared wavelengths. The EnKF uses an ensemble of states to represent the error covariances to estimate 8-day CO<sub>2</sub> surface fluxes over 144 geographical regions. We use a 12×8-day lag window, recognising that  $X_{\text{CO}_2}$  measurements include surface flux information from prior time windows. The observation operator that relates surface CO<sub>2</sub> fluxes to atmospheric distributions of  $X_{\text{CO}_2}$  includes: a) the GEOS-Chem transport model that relates surface fluxes to global 3-D distributions of CO<sub>2</sub> concentrations, which are sampled at the time and location of OCO measurements that are cloud-free and have aerosol optical depths <0.3; and b) scene-dependent averaging kernels that relate the CO<sub>2</sub> profiles to  $X_{\text{CO}_2}$ , accounting for differences between nadir and glint measurements, and the associated scene-dependent observation errors. We show that OCO  $X_{\text{CO}_2}$  measurements significantly reduce the uncertainties of surface CO<sub>2</sub> flux estimates. Glint measurements are generally better at constraining ocean CO<sub>2</sub> flux estimates. Nadir  $X_{\text{CO}_2}$  measurements over the terrestrial tropics are sparse throughout the year because of either clouds or smoke. Glint measurements provide the most effective constraint for estimating tropical terrestrial CO<sub>2</sub> fluxes by accurately sampling fresh continental outflow over neighbouring oceans. We also present results from sensitivity experiments that investigate how flux estimates change with 1) bias and

unbiased errors, 2) alternative duty cycles, 3) measurement density and correlations, 4) the spatial resolution of estimated flux estimates, and 5) reducing the length of the lag window and the size of the ensemble. At the revision stage of this manuscript, the OCO instrument failed to reach its orbit after it was launched on 24 February 2009. The EnKF formulation presented here is also applicable to GOSAT measurements of CO<sub>2</sub> and CH<sub>4</sub>.

## 1 Introduction

CO<sub>2</sub> surface fluxes inferred from atmospheric CO<sub>2</sub> concentrations by inverting models of atmospheric transport have led to substantial improvements in our understanding of the contemporary carbon cycle (e.g., Bousquet et al., 2000). Previous studies that employ these methods to estimate surface fluxes of CO<sub>2</sub> have tended to use accurate, but spatially sparse and heterogeneous, ground-based measurements, which were not designed for the flux estimation problem, consequently limiting the extent of spatial disaggregation of fluxes that can be achieved (e.g., Houweling et al., 1999; Rödenbeck et al., 2003). Satellite measurements of CO<sub>2</sub> offer new constraints for estimating surface fluxes. The SCanning Imaging Absorption spectroMeter for Atmospheric Chartography (SCIAMACHY) satellite instrument (Bovensmann et al., 1999) has measured short-wave infrared wavelengths (SWIR), with greatest sensitivity to CO<sub>2</sub> in the lower troposphere, since its launch in 2002. Current CO<sub>2</sub> column volume mixing ratio products from SCIAMACHY have an estimated measurement accuracy of between 1 and 5% (Schneising et al., 2008; Barkley et al., 2006, 2007). Uncharacterized systematic and random errors (e.g., Houweling et al., 2005), while the subject of ongoing research (Schneising et al., 2008), limit the application of these data for surface



Correspondence to: L. Feng  
(lfeng@staffmail.ed.ac.uk)

flux estimation. Top-down studies that use satellite measurements of CO<sub>2</sub> retrieved at thermal infra-red wavelengths, with greatest vertical sensitivity in the free troposphere, have concluded that uncharacterized observation and model biases compromise resulting surface flux estimates (Chevallier et al., 2005; Tiwari et al., 2006).

The NASA Orbiting Carbon Observatory (OCO)<sup>1</sup> (Crisp et al., 2004), and Japanese Greenhouse Observing SATellite (GOSAT) (Maksyutov et al., 2008), launched in early 2009, measure SWIR wavelengths, that are sensitive to CO<sub>2</sub> in the free and lower troposphere. OCO and GOSAT will operate two modes of observation: (1) nadir, and (2) glint, where the instrument boresight is directed off-nadir to the angle of specular reflection. The glint mode increases the signal to noise of measurements over the ocean. Dry-air CO<sub>2</sub> mole fractions ( $X_{\text{CO}_2}$ ) will be retrieved from the observed spectra to a precision of 1–2 ppmv (parts per million by volume) (Crisp et al., 2004), a level of precision necessary to improve upon constraints from existing in situ measurements (Rayner et al., 2002; Patra et al., 2003; Miller et al., 2007). We focus on OCO measurements of CO<sub>2</sub>, but the general assimilation approach described here can easily be applied to GOSAT measurements of CO<sub>2</sub> or CH<sub>4</sub>. Recent studies have used variational data assimilation methods with synthetic OCO observations to show that these data have the potential to estimate weekly and daily surface CO<sub>2</sub> fluxes at model grid scales of order 3.75° in longitude and 2.5° in latitude (Baker et al., 2006; Chevallier et al., 2007a; Chevallier, 2007b). These studies (1) assumed a constant measurement error (1–2 ppmv), and (2) used a flat weighting function to convert the model vertical CO<sub>2</sub> profiles into  $X_{\text{CO}_2}$ .

We have developed an Ensemble Kalman Filter (EnKF) (Evensen, 1994, 2003; Ehrendorfer, 2007) to estimate surface CO<sub>2</sub> fluxes from space-borne measurements of  $X_{\text{CO}_2}$  (Sect. 2). The EnKF, an independent and complementary approach to variational assimilation, has been developed in the physical oceanography and meteorology communities (e.g., Evensen, 1994; Houtekamer and Mitchell, 1998; Lorenc, 2003), and recently applied to carbon cycle research (Peters et al., 2005; Bruhwiler et al., 2005). The EnKF methodology we use is outlined in Sect. 3. We use the GEOS-Chem chemistry transport model to describe the relationship between surface CO<sub>2</sub> fluxes and 3-D atmospheric CO<sub>2</sub> concentrations, which are then sampled along the proposed OCO orbits and convolved with scene-dependent instrument averaging kernels as a function of observation modes, surface types, solar zenith angles, and optical depths (Sect. 2). This new, improved description of OCO measurements and their errors (Bösch et al., 2009, Sect. 2) is expected to provide more realistic descriptions of  $X_{\text{CO}_2}$  distributions, with which to infer more realistic flux estimates. We use the EnKF to explore

<sup>1</sup>At the time of revision, the NASA OCO satellite failed to reach its orbit after it was launched from Vandenberg Air Force Base, California, USA on 24 February 2009.

the sensitivity of the surface flux inverse problem to changes in instrument configurations and the size of geographical regions over which fluxes are to be estimated (Sect. 4). We conclude the paper in Sect. 5.

## 2 Simulated OCO $X_{\text{CO}_2}$ observations and uncertainties

The OCO instrument was planned to be launched into the NASA EOS Afternoon Constellation (A-train), which is in a sun-synchronous polar orbit at an approximate altitude of 705 km. This orbit has 14.6 equator crossings per day, separated by 24.7° in longitude, resulting in a 16-day repeat cycle. OCO will have a local equatorial crossing time of 13:18. The OCO platform includes three, high-resolution grating spectrometers that measure absorptions of the reflected sunlight by using two CO<sub>2</sub> bands (1.61 and 2.06 μm) and the O<sub>2</sub> A-Band (0.765 μm) using nadir view geometry or glint view geometry in which the instrument will be pointed to the spot where solar radiation is specularly reflected from the surface (Crisp et al., 2004). The 32-day duty cycle of OCO will alternate between 16-day cycles of nadir and glint modes.

We model OCO  $X_{\text{CO}_2}$  measurements in a four-step process, which constitutes the observation operator  $H$  that relates surface CO<sub>2</sub> fluxes to global distributions of  $X_{\text{CO}_2}$ . First, we use the GEOS-Chem chemistry transport model (v7-03-06) to relate surface fluxes to global 3-D CO<sub>2</sub> concentrations. For the purpose of these calculations we use the same flux inventory as described in Palmer et al. (2008) for 2003. For the sake of brevity, here we describe the model briefly and refer the reader to Appendix A and Palmer et al. (2008) for further details. We use a horizontal resolution of 2° in latitude and 2.5° in longitude for the experiments described here, with meteorological analyses from version 4 of the GEOS model from the NASA Goddard Global Modelling and Assimilation Office.

We include CO<sub>2</sub> estimates for daily biospheric fluxes (Potter et al., 1993), monthly oceanic fluxes (Takahashi et al., 2002), monthly biomass burning fluxes from the second version of the Global Fire Emission Database (GFEDv2) for 2003, a climatological distribution of annual fossil fuel emissions that have been scaled to 2003 (Palmer et al., 2008), and climatological biofuel fluxes (Yevich and Logan, 2003).

Second, we sample the 3-D field of CO<sub>2</sub> concentrations at the time and the location of each nadir and glint measurement using the orbits of the Aqua satellite in 2006, which leads the A-train constellation with a local equatorial crossing time of 13:30. In this study, we use a distribution of measurements based on the availability of full-physics retrievals (Bösch et al., 2009), assumed to be made at every 20 s so that two consecutive observations in one orbit are separated by 1.2° in latitude.

Third, we use seasonal probability density functions (PDFs) of cloud and aerosol optical depths (AODs), derived from the MODIS and MISR instruments (Bösch et al., 2009),

to remove cloudy scenes and scenes with  $\text{AOD} > 0.3$  that will not be retrieved, at least, initially from OCO. These restrictions remove about 50–60% of the daily available nadir measurements, and 60–70% of the daily glint measurements. Hereinafter, we refer to the resulting measurements as clear.

Finally, we apply scene-dependent averaging kernels, which account for the vertical sensitivity of OCO, to map from the 1-D  $\text{CO}_2$  concentration profiles to  $X_{\text{CO}_2}$  (Connor et al., 2008):

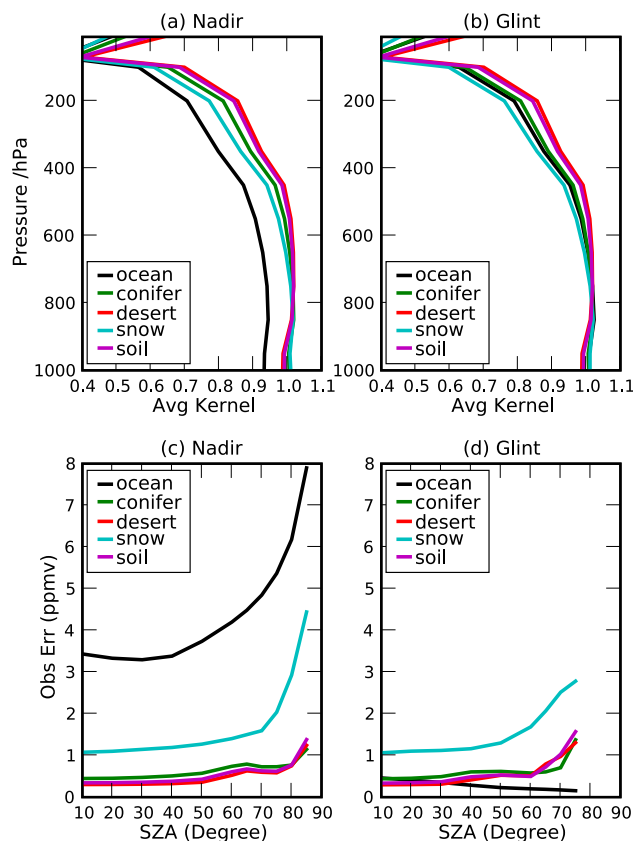
$$X_{\text{CO}_2} = X_{\text{CO}_2,a} + \mathbf{a} \left( (1 - \mathbf{w})^{-1} (M(\mathbf{x}^t) - \mathbf{f}_a) \right). \quad (1)$$

Bold lower case variables denote vectors and bold upper case variables denote matrices. The subscript  $a$  denotes a priori;  $M(\mathbf{x}^t)$  is the GEOS-Chem chemistry transport model driven by “true” surface fluxes of  $\text{CO}_2$  ( $\mathbf{x}^t$ );  $\mathbf{w}$  denotes GEOS-4 water mole fractions that are used to map from  $\text{CO}_2$  concentrations to dry mole fraction; and  $\mathbf{f}_a$  is climatological dry  $\text{CO}_2$  mole fractions that will be used to retrieve  $\text{CO}_2$  profile information from OCO, and  $X_{\text{CO}_2,a}$  is the associated column amount. We use an annual zonal mean for  $\mathbf{f}_a$ . The column averaging kernel  $\mathbf{a}$  is given by  $\mathbf{t}^T \mathbf{A}$ , where  $\mathbf{A}$  is the averaging kernel,  $\mathbf{t}$  is the column integration operator that integrates a vertical profile to a column and superscript T denotes the matrix transpose operation.

We use averaging kernels as a function of two view modes (nadir and glint), five surface types (snow, ocean, soil, conifer, and desert), ten solar zenith angles (SZA) (from  $10^\circ$  to  $85^\circ$  for nadir measurements, and from  $10^\circ$  to  $72^\circ$  for glint measurements), and seven AODs from 0 to 0.3 (Bösch et al., 2009).

Figure 1a and b shows averaging kernels for five different surface types at a SZA of  $10^\circ$  under a clear-sky with an AOD of 0.1. In general, OCO averaging kernels peak in the mid and lower troposphere. The instrument sensitivity to changes in  $\text{CO}_2$  near the surface is particularly important for flux estimation. Using nadir view geometry, the oceans are relatively dark at the SWIR wavelengths measured by OCO, and the resulting averaging kernels below 400 hPa are lower than over other surface types. In contrast, glint view measurements over the oceans take advantage of specular reflection, resulting in a large signal to noise and an averaging kernel close to unity below 400 hPa.

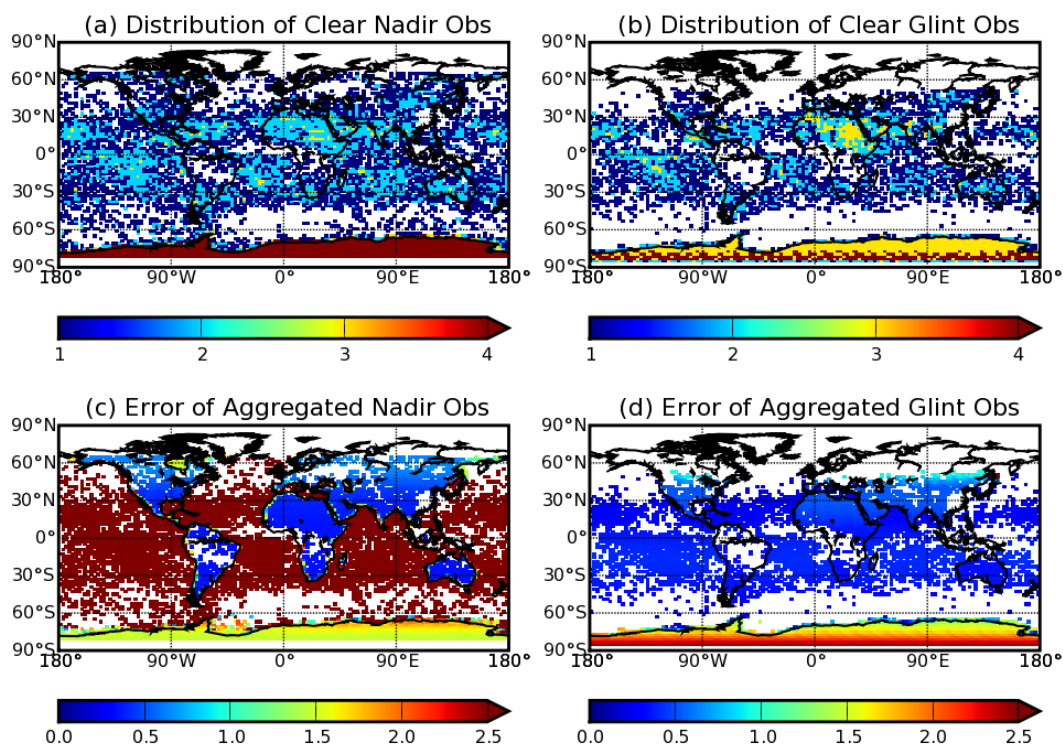
The uncertainty associated with the simulated  $X_{\text{CO}_2}$  also depends on the scene characterization. Figure 1c and d shows observation errors over 5 different surface types as a function of SZA. The error over land is usually  $< 0.5$  ppmv for a single nadir measurement at SZAs  $< 40^\circ$ , but increases with SZA, eventually reaching 1.2 ppmv at a SZA of  $85^\circ$ . Observation errors for nadir measurements over ocean are typically  $> 3.0$  ppmv for all SZAs. In contrast, the error for a single glint measurement over ocean is typically  $< 0.4$  ppmv, smaller than nadir errors over land. These errors for measurements over lands are smaller than the assumed model transport and representation errors (2.5 ppmv), which, as we



**Fig. 1.** Orbiting Carbon Observatory (OCO) instrument averaging kernels (dimensionless) associated with (a) nadir and (b) glint SWIR  $X_{\text{CO}_2}$  measurements as a function of pressure (hPa) for different land types, at a solar zenith angle (SZA) of  $10^\circ$  and an aerosol optical depth (AOD) of 0.1. Observation errors (ppmv) associated with (c) nadir and (d) glint  $X_{\text{CO}_2}$  measurements as function of SZA for different land types and an AOD of 0.1.

show later, has important implications for flux estimation using these data.

Figure 2 summarises the resulting  $2^\circ \times 2.5^\circ$  distribution and uncertainties of nadir and glint  $X_{\text{CO}_2}$  measurements over 16 days between 17 January and 1 February 2003. Continental regions at mid and low latitudes typically have uncertainties of  $< 0.2$  ppmv (Fig. 2b), well within the target precision of OCO (Crisp et al., 2004), while oceans have an observation error  $> 3$  ppmv. Over tropical regions, scenes are frequently obscured by clouds during the wet seasons, and frequently obscured by smoke aerosol from biomass burning during the dry seasons. Glint measurements are generally restricted to a smaller latitude range ( $85^\circ \text{S}$  to  $55^\circ \text{N}$  for the months shown) than nadir measurements because they are used over a small range of SZAs ( $< 72^\circ$  for glint versus  $< 85^\circ$  for nadir). There are typically less glint observations than nadir over the tropics due to their larger view spot ( $\approx 25 \text{ km}^2$  at the extreme view angles versus  $\approx 3 \text{ km}^2$  for the nadir view). The larger view spot increases the probability



**Fig. 2.** Number of clear observations (aerosol optical depth  $<0.3$  and cloud-free) for (a) nadir and (b) glint  $X_{\text{CO}_2}$  measurements averaged over 16 days from 17 January to 1 February 2003, on a horizontal grid of  $2^\circ \times 2.5^\circ$ . Associated aggregated errors (ppmv) for the (c) nadir and (d) glint  $X_{\text{CO}_2}$  measurements.

of cloud obscuration. A similar method is used to define “model”  $X_{\text{CO}_2}$  distributions for the observation system simulation experiment (OSSE) in Sect. 4.

### 3 The Ensemble Kalman Filter

#### 3.1 Basic formulation

We have developed an ensemble data assimilation system based on the Ensemble Transform Kalman Filter (ETKF) technique (Bishop et al., 2001) to simultaneously assimilate consecutive  $X_{\text{CO}_2}$  observations. At each assimilation cycle, we assimilate 8-day OCO observations  $\mathbf{y}_{\text{obs}}$  to improve the prior estimation of regional surface  $\text{CO}_2$  fluxes via:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}[\mathbf{y}_{\text{obs}} - H(\mathbf{x}^f)] \quad (2)$$

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T [\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R}]^{-1}, \quad (3)$$

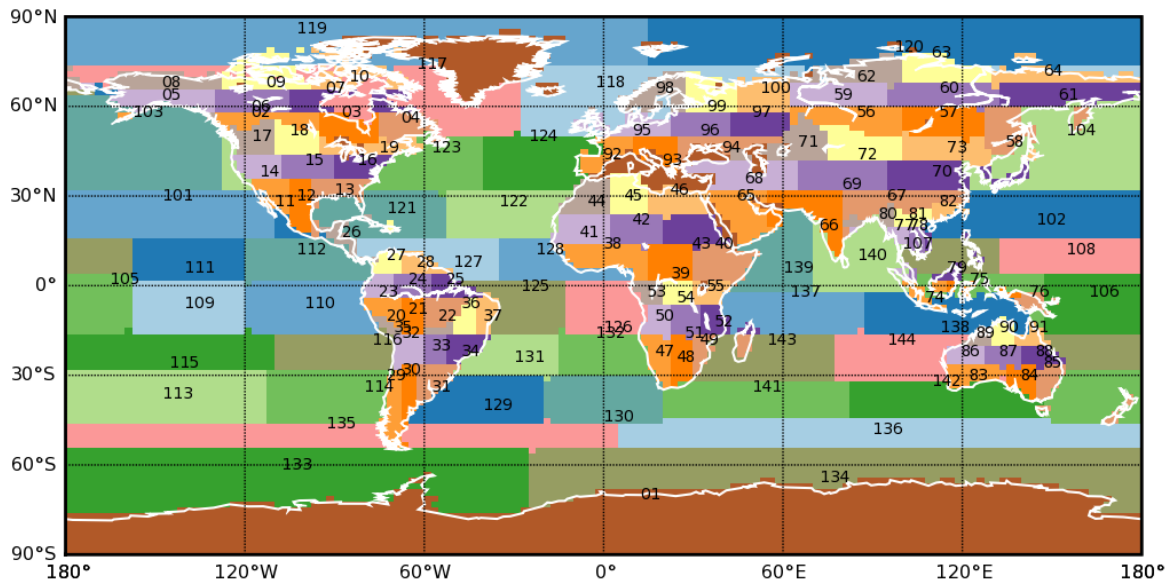
where  $\mathbf{x}^f$  is the a priori state vector;  $\mathbf{x}^a$  is the a posteriori;  $H$  is the observation operator that describes the relationship between the state vector and the observations (Sect. 2); and  $\mathbf{K}$  is the Kalman gain matrix that determines the adjustment to the a priori based on the difference between model and

observations and their uncertainties.  $\mathbf{R}$  is the observation error covariance matrix, and  $\mathbf{P}^f$  is the a priori error covariance matrix.  $\mathbf{H}$ , the Jacobian of the observation operator  $H$  (Sect. 2), maps  $\mathbf{P}^f$  into observation space. As mentioned above in Sect. 2, the observation operator  $H$  includes the GEOS-Chem model to describe the atmospheric transport of  $\text{CO}_2$ , which uses prescribed meteorological analysis; consequently, there is no model feedback between  $\text{CO}_2$  and atmospheric dynamics and the transport of  $\text{CO}_2$  can be considered as a linear process.

The observation error covariance  $\mathbf{R}$  includes measurement (instrument + retrieval) error, model (transport) error and representation error (Peylin et al., 2002). Quantifying model error is non-trivial, and for simplicity we have assumed uniform model and representation errors: 2.5 ppmv over land regions and 1.5 ppmv over oceans (Rödenbeck et al., 2003), both of which are uncorrelated with the measurement errors. Also, we assume that  $\mathbf{R}$  is either diagonal (i.e., no observation correlation, Sect. 4.1), or has a simple block structure for correlations between successive observations (Sect. 4.4).

Below we show that our model  $X_{\text{CO}_2}$  fields are able to independently estimate 8-day mean surface fluxes at a spatial resolution of  $1000 \times 1000 \text{ km}^2$ ; estimating at much finer spatial resolution introduces strong negative error correlations





**Fig. 3.** The continental and ocean regions used to estimate CO<sub>2</sub> source and sinks, based on a coarser distribution from the TransCom-3 experiment (see Table 1 in the text and Gurney et al., 2002).

between neighbouring flux estimates (Sect. 4.5). Figure 3 shows our geographical regions that span the globe, which are based on previous work by TransCom-3 (T3) (Law et al., 2003; Gurney et al., 2002). We divide each T3 land region into 9 near-equal areas (resulting in  $9 \times 11$  land regions), and divide each T3 ocean region into 4 near-equal areas (resulting in  $4 \times 11$  ocean regions). We have included one region (region 1) to represent all other low-emission regions, typically covered by snow. Recognising that observed CO<sub>2</sub> concentration variations contain a history of source/sink signatures, we use a lag window of  $12 \times 8$  days in the control run so that the state vector  $\mathbf{x}$  consists of these regional fluxes at the current assimilation timestep and those from the previous 11 timesteps of 8 days. This corresponds to solving the regional fluxes over a  $3 \times 32$ -day OCO duty cycles, resulting in  $12 \times 144$  control variables. For simplicity, we do not account for uncertainties in the initial CO<sub>2</sub> concentration distribution at the beginning of the whole experiment, which progressively become less important than the uncertainties in the recent surface emissions as more OCO observations are digested (Chevallier et al., 2007a).

For the control variables corresponding to emissions within  $11 \times 8$  days prior to the current 8-day observations, both the prior estimates and the associated uncertainties are the results from the previous assimilation cycles. We use climatological flux estimates as the current 8-day regional flux forecasts (after being enlarged by 80% with respect to the model run used to generate the observations, Fig. 4), with the associated error  $\varepsilon_i$  for an individual geographical region  $i$  estimated by rescaling the annual-mean T3 regional a priori error  $\varepsilon_{T3}$  (e.g., Patra et al., 2003) to an 8-day mean flux error

over an area of  $1/9$  (for lands) or  $1/4$  (for oceans) size of the parent T3 region:

$$\varepsilon_i = \varepsilon_{T3} \left( \sqrt{\frac{365}{8}} \right) \left( \sqrt{\frac{A_l}{A_{T3}}} \right), \quad (4)$$

where the first bracketed term represents the scaling from the TransCom annual mean error to the 8-day period, and the second bracketed term represents the scaling from T3 area  $A_{T3}$  to the regional area  $A_l$ . Table 1 summarises the forecast errors for the regions 2 to 144. We assume that the snow region (region 1) has zero emissions with an uncertainty of  $0.1 \text{ GtC yr}^{-1}$ . We assume that the a priori errors, given by Eq. (4), have no temporal or spatial correlation, so that their (sub) error covariance matrix is diagonal. The other  $11 \times 144$  variables (from previous 8-day periods), having passed through a number of assimilation cycles, include spatial and temporal correlations.

Using the ensemble approach, we approximate the a priori error covariance by introducing an ensemble of perturbation states  $\Delta \mathbf{X}^f = [\Delta \mathbf{x}_1, \Delta \mathbf{x}_2, \dots, \Delta \mathbf{x}_{N_e}]$  (Evensen, 1994), so that

$$\mathbf{P}^f = \Delta \mathbf{X}^f (\Delta \mathbf{X}^f)^T, \quad (5)$$

where we have absorbed the normalization factor  $1/(N_e - 1)$  into  $\Delta \mathbf{X}^f$  (Zupanski, 2005). As a result,  $\mathbf{K}$  can now be approximated by the ensemble gain matrix  $\mathbf{K}_e$ :

$$\mathbf{K}_e = \Delta \mathbf{X}^f (\Delta \mathbf{Y})^T [\Delta \mathbf{Y} (\Delta \mathbf{Y})^T + \mathbf{R}]^{-1}, \quad (6)$$

$$\Delta \mathbf{Y} = H(\mathbf{x}^f + \Delta \mathbf{X}^f) - H(\mathbf{x}^f). \quad (7)$$

**Table 1.** Uncertainty (GtC yr<sup>-1</sup>) associated with original TransCom-3 (T3) continental and ocean regions that have been subdivided for our EnKF inversion. We assume the uncertainty of region 1 (the snow region) to be 0.1 GtC yr<sup>-1</sup>.

T3 Region	Err	EnKF Region	Err
North American Boreal	0.73	Reg (002–010)	1.64
North American Temperate	1.50	Reg (011–019)	3.38
South American Tropical	1.41	Reg (020–028)	3.18
South American Temperate	1.23	Reg (029–037)	2.76
North Africa	1.33	Reg (038–046)	3.00
South Africa	1.41	Reg (047–055)	3.18
Eurasia Boreal	1.51	Reg (056–064)	3.41
Eurasia Temperate	1.73	Reg (065–073)	3.89
Tropical Asia	0.87	Reg (074–082)	1.95
Australia	0.59	Reg (083–091)	1.34
Europe	1.42	Reg (092–100)	3.20
North Pacific Temperate	0.27	Reg (101–104)	0.61
West Pacific Tropics	0.39	Reg (105–108)	0.88
East Pacific Tropics	0.37	Reg (109–112)	0.83
South Pacific Temperate	0.63	Reg (113–116)	1.42
Northern Ocean	0.35	Reg (117–120)	0.79
Northern Atlantic Temperate	0.27	Reg (121–124)	0.61
Atlantic Tropics	0.41	Reg (125–128)	0.92
South Atlantic Temperate	0.55	Reg (129–132)	1.24
South Ocean	0.72	Reg (133–136)	1.62
Indian Tropical	0.48	Reg (137–140)	1.08
South Indian Temperate	0.41	Reg (141–144)	0.92

Using the EnKF approach we do not need the Jacobian matrix  $\mathbf{H}$  explicitly to calculate the gain matrix  $\mathbf{K}_e$ .

The EnKF is able to provide a direct estimation of the analysis error covariance. We use the revised, unbiased Ensemble Transform Kalman Filter (ETKF) algorithm (Wang et al., 2004; Livings et al., 2008) to determine the analysis ensemble  $\Delta\mathbf{X}^a$  and the a posteriori error covariance,  $\mathbf{P}^a$ :

$$\Delta\mathbf{X}^a = \Delta\mathbf{X}^f \mathbf{T}, \quad (8)$$

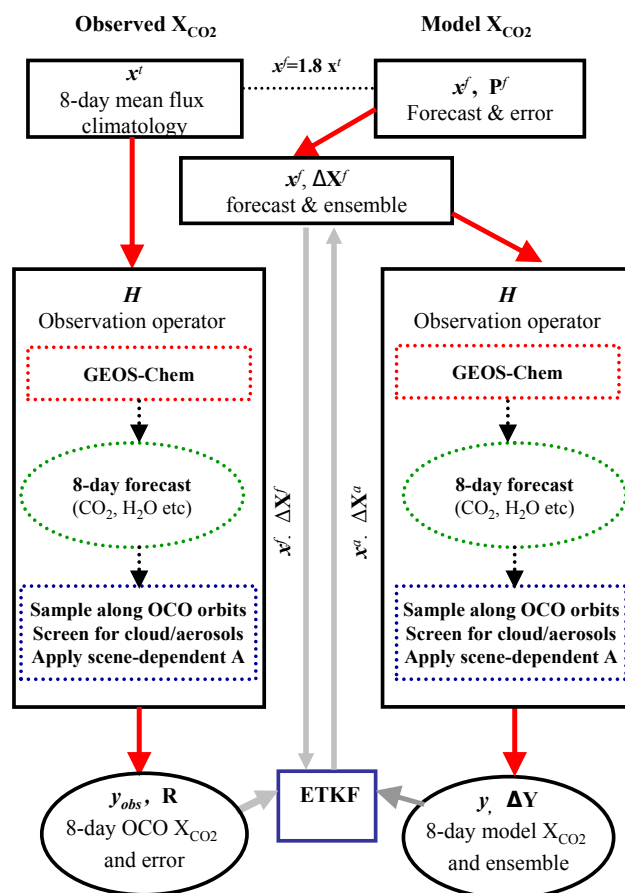
and

$$\mathbf{P}^a = \Delta\mathbf{X}^f \mathbf{T} (\Delta\mathbf{X}^f \mathbf{T})^T. \quad (9)$$

The transform matrix  $\mathbf{T}$  is given by

$$\mathbf{T}(\mathbf{T})^T = \mathbf{I} - (\Delta\mathbf{Y})^T [\Delta\mathbf{Y}(\Delta\mathbf{Y})^T + \mathbf{R}]^{-1} \Delta\mathbf{Y}. \quad (10)$$

We simplify the calculation of  $\mathbf{T}$ ,  $\mathbf{K}_e$ , and  $[\Delta\mathbf{Y}(\Delta\mathbf{Y})^T + \mathbf{R}]$ , which is large due to the dense OCO observations, by using singular value decomposition (SVD) of the scaled model observation ensemble  $\Delta\mathbf{Y}^T \mathbf{R}^{-1/2}$  (Livings, 2005).



**Fig. 4.** Schematic diagram of the OCO  $X_{\text{CO}_2}$  Observing System Simulation Experiment (OSSE). The left column describes the simulation of OCO  $X_{\text{CO}_2}$  measurements:  $\mathbf{x}^f$  denotes the “true” fluxes; and  $H$  is the observation operator for mapping surface fluxes to  $X_{\text{CO}_2}$  observations  $\mathbf{y}_{\text{obs}}$ . It includes the GEOS-Chem global 3-D transport model that relates surface fluxes to global 3-D  $\text{CO}_2$  distributions, which are then sampled along OCO orbits. Scenes with cloud or aerosol optical depths  $>0.3$  are removed. The resulting profiles are mapped to  $X_{\text{CO}_2}$  using scene-specific averaging kernels, with associated scene-specific error  $\mathbf{R}$ . The right column describes the simulation of model  $X_{\text{CO}_2}$  measurements using prior fluxes  $\mathbf{x}^f$  (80% larger than  $\mathbf{x}^f$ ) and the associated error covariance  $\mathbf{P}^f$ , which is approximated by the perturbation state vector ensemble  $\Delta\mathbf{X}^f$ . Mapping  $\mathbf{x}^f$  and  $\Delta\mathbf{X}^f$  to the observation space by observation operator  $H$  results in the model observation  $\mathbf{y}$ , and the associated variations  $\Delta\mathbf{Y}$ . The middle column shows that the Ensemble Transform Kalman Filter (ETKF) algorithm generates the optimal estimate  $\mathbf{x}^a$ , and the a posteriori error covariance  $\mathbf{P}^a$  by comparing the model forecasts with observations.

### 3.2 A priori error and its representation

We construct an ensemble of perturbation states to reflect the a priori error covariance matrix, using eigenvalue decomposition:

$$\mathbf{P}^f = \mathbf{V}_x \mathbf{p}^{1/2} \left( \mathbf{V}_x \mathbf{p}^{1/2} \right)^T, \quad (11)$$

where  $\mathbf{V}_x$  and  $\mathbf{p}$  are the eigenvector matrix and the eigenvalue diagonal matrix of the error covariance, respectively.

At the limit of using the full-rank matrix, as we do here, the ensemble of perturbation states is defined as:

$$\Delta \mathbf{X}^f = \mathbf{V}_x \mathbf{p}^{1/2}, \quad (12)$$

where the matrix  $\Delta \mathbf{X}^f$  has a size of  $N_x \times N_e$ , with  $N_x = 12 \times 144$ , and the ensemble size  $N_e$  being equal to  $N_x$ . When a full-rank representation is used, the Kalman gain matrix and the a posteriori error covariance determined from Eq. (6) and Eq. (9) are fully consistent with the ordinary Kalman filter approach (Zupanski, 2005). The most time-consuming part of our EnKF is the projection of the flux perturbations to the observation space, using the observation operator that includes running a global transport model (Sect. 2).

In our sensitivity study, we do not need to re-run the transport model for inversions using different observation configurations. Instead, we define one diagonal matrix  $\Delta \mathbf{X}_0^f$  of the same size as  $\Delta \mathbf{X}^f$ , with each column only specifying an emission occurring in one of the twelve 8-day periods over one of the 144 regions. We then calculate the variations in the observed  $X_{\text{CO}_2}$  caused by these emissions through the observation operator  $H$

$$\Delta \mathbf{Y}_0 = \mathbf{H}(\Delta \mathbf{X}_0^f) = H(\mathbf{x}^f + \Delta \mathbf{X}_0^f) - H(\mathbf{x}^f). \quad (13)$$

We can calculate  $\Delta \mathbf{Y}$  for any given a priori ensemble  $\Delta \mathbf{X}^f$  by:

$$\Delta \mathbf{Y} = \mathbf{H}(\Delta \mathbf{X}^f) = \Delta \mathbf{Y}_0 \left( [\Delta \mathbf{X}_0^f]^{-1} \Delta \mathbf{X}^f \right). \quad (14)$$

In practice, we retain only a subset of the column vectors given by Eq. (12), ignoring those associated with small amplitudes, sufficient to provide a good approximation of the error covariances (Appendix B). In such a reduced-rank representation, Eq. (14) becomes invalid.

## 4 Results

We evaluate our EnKF approach using an observing system simulation experiment (OSSE) framework, which is illustrated in Fig. 4. OSSEs have been used extensively to study the impacts of new observations on data assimilation systems (see for example, Lahoz et al., 2005), but it is widely recognised that they can lead to over-optimistic results (Atlas, 1997). In our case, we use the same GEOS-Chem transport model to generate and assimilate  $X_{\text{CO}_2}$  measurements.

Such an OSSE framework is not suitable to study the effects of systematic model errors on inversions. Instead, we focus our OSSE on quantifying the science capabilities of realistic distributions of  $X_{\text{CO}_2}$  measurements from space-borne sensors.

Observed  $X_{\text{CO}_2}$  distributions are described in Sect. 2, which we regard as the “truth”. Model  $X_{\text{CO}_2}$  distributions are defined similarly but we assume the prior flux estimates to be 80% higher than the “true” values.

First, we present results from a control experiment for a 7-month period from 1 January to 31 July 2003, during which the OCO instrument is assumed to operate at the nominal 32-day duty cycle with alternating 16-day nadir and glint measurements. We then assess the sensitivity of the a posteriori flux estimates to 1) systematic (bias) and random (unbiased) errors; 2) observation error, density and correlations; 3) alternative duty cycles; 4) the spatial resolution of the state vectors; and 5) the length of the lag window and the size of the ensemble.

We evaluate the performance of the EnKF by using an error reduction  $\gamma$

$$\gamma = 1 - \sigma^a / \sigma^f, \quad (15)$$

where  $\sigma^f$  and  $\sigma^a$  denote the a priori and a posteriori variance uncertainties, respectively. For each 8-day mean regional flux, we calculate its  $\sigma^f$  from the a priori error covariance at the time when it first enters the lag window, and calculate  $\sigma^a$  from the a posteriori error covariance at the time when it leaves the lag window.

The error reduction  $\gamma$  is insensitive to our assumptions about the “true” surface fluxes, as well as the values of the simulated  $X_{\text{CO}_2}$  observations. However, approximations in the EnKF approach may lead to underestimation of the a posteriori uncertainties (Livings et al., 2008) when a reduced-rank representation of the error covariance is used (see Sect. 4.6).

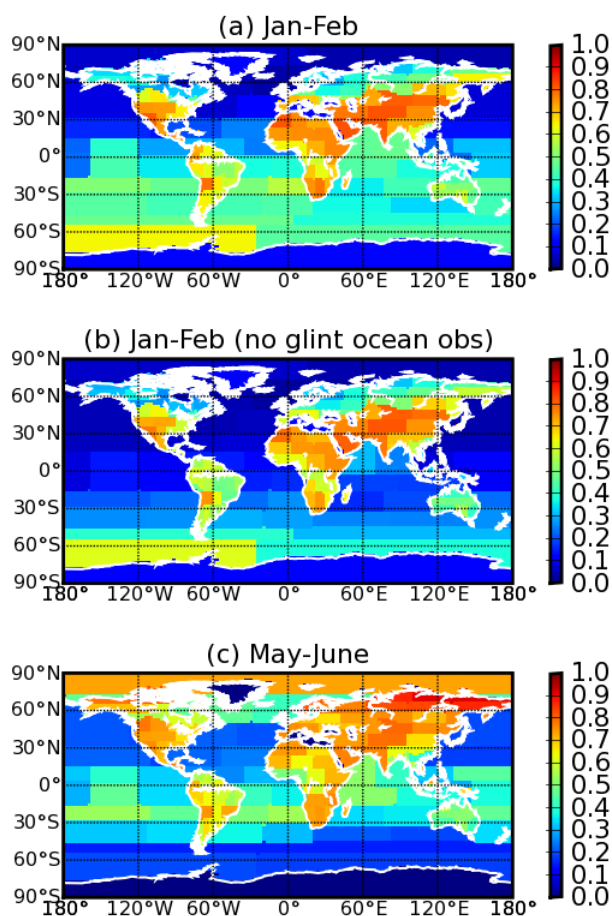
### 4.1 Control experiment

Figure 5 presents the error reduction in the estimates for 8-day mean fluxes over 144 regions. The results have been averaged over a 32-day period from 17 January to 17 February 2003.

During the northern winter, nadir measurements cover the latitudes between 90° S and 60° N, while glint measurements only reach 55° N. As a result, over most land regions between 30° S and 50° N, OCO measurements reduce uncertainties in the flux estimates by more than 70% (Fig. 5a), while errors over the boreal latitudes decrease by 20–65%. The widespread error reduction reflects the coverage and the precision of nadir and glint measurements. These significant reductions are also related to the large uncertainties in the prior estimates.

We find that most of the error reduction occurs when the continental signal is younger than 3 weeks and still distinct





**Fig. 5.** (a) A posteriori error reduction,  $\gamma = 1 - \sigma^a / \sigma^f$ , associated with CO<sub>2</sub> flux estimation using OCO X<sub>CO<sub>2</sub></sub> observations over one duty cycle, including alternate 16-day period of nadir and glint measurements. (a) Over 17 January–17 February 2003; (b) over 17 January–17 February 2003 without glint measurements over the ocean; and (c) over 1 May–1 June 2003. Warm colours (high values of  $\gamma$ ) denote large error reductions and cold colours (low values of  $\gamma$ ) denote small error reductions.

from the slowly varying background. Over regions with a dense distribution of observations, the error reduction can reach saturation well within three (8-day) assimilation cycles. Conversely, at the northern high latitudes during winter, when there is a low observation density, saturation of error reduction requires more time but is still within 5–6 assimilation cycles (<2 months). These results suggest that our 3-month lag window is more than sufficient (Sect. 4.6).

We find that OCO measurements reduce the uncertainties in oceanic CO<sub>2</sub> flux estimates by 10–60%, despite their a priori errors being much smaller than the typical values of the land regions. Most of these reductions are attributed to the accurate glint measurements over the oceans (Fig. 1). To highlight this point, Fig. 5b shows the error reduction when the glint measurements over the oceans are excluded from the assimilation. Without these measurements, the error re-

duction over ocean reaches 10–30%. We also find that omitting glint measurements over ocean also leads to lower error reduction over tropical continents. For example, the error reduction for region 27 over tropical South America is 50%, compared to 70% with all clear glint measurements (Fig. 5a). The results reflect the additional constraints on continental surface CO<sub>2</sub> flux from accurate glint measurements of continental outflow over the surrounding oceans.

The results for the time period from 1 May to 1 June 2003 (Fig. 5c) show similar significant error reductions over land. In particular, the northern high latitudes are now fully covered by OCO measurements, and the corresponding error reduction reaches 40–90%.

#### 4.2 Sensitivity to bias and unbiased error

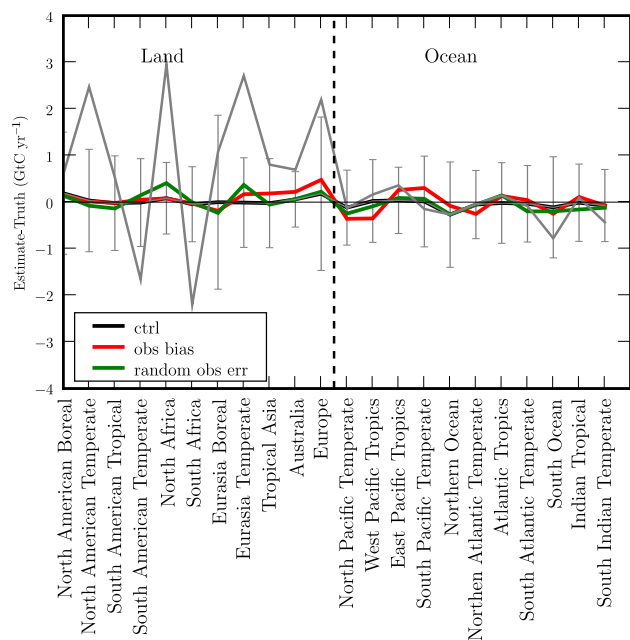
We generate a random observation error for each OCO measurement by sampling a Gaussian probability distribution function with the variance equal to the measurement uncertainty. Figure 6 shows that including random errors to the “true” OCO observations (see Fig. 4) leads to departures from the control flux estimates. These departures are well within the a posteriori errors, and usually become even smaller when averaged over a longer period (not shown).

The impacts from small-scale or scene-dependent measurement biases on the source/sink estimation are of interest. Previous work has implemented spatially coherent bias by relating it to sub-micron aerosols (Chevallier et al., 2007a). We use a similar approach to include an observation bias (in ppmv) twice that of the obscuring AODs. This results in a maximum bias for clear observations of about 0.6 ppmv, a magnitude similar to the typical observation error. Using this approach, we find significant positive biases over Eurasia, and over Europe, accompanied by negative and positive biases over Pacific, similar to Chevallier et al. (2007a). However, these systematic differences in the estimated fluxes are usually smaller than the a posteriori errors.

#### 4.3 Sensitivity to measurement duty cycle

The OCO satellite repeats its sun-synchronous orbit every 16 days and the current instrument configuration is to switch between nadir and glint mode at the same frequency (Crisp et al., 2004), which can be reprogrammed within orbit, if necessary. Consequently, there are a number of nadir–glint measurement combinations that could form a 32-day duty cycle over the nominal two-year OCO mission. Here we assess the impact of two alternative duty cycles on estimating surface CO<sub>2</sub> fluxes: the nadir-only cycle and the glint-only cycle.

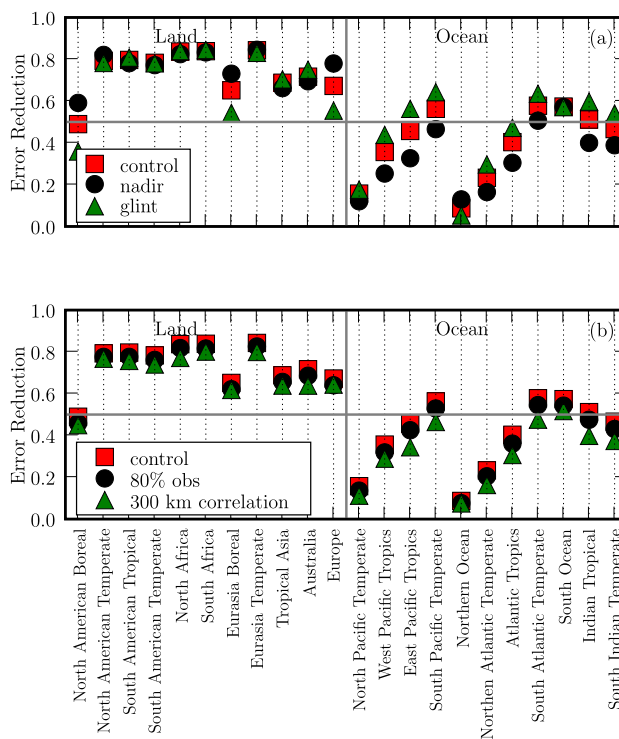
Figure 7a compares the results for the nadir-only and glint-only duty cycles with the control experiment. We have averaged the results over a 32-day cycle from 17 January to 17 February 2003. For all the three duty cycles, the geographical pattern of the resulting error reduction is similar, showing significant reductions (40–85%) over land, and



**Fig. 6.** The deviations ( $\text{GtC yr}^{-1}$ ) of the estimated  $\text{CO}_2$  fluxes from the “truth” for one duty cycle from 17 January to 17 February 2003. The results have been aggregated from 144 regions (Fig. 3) to the 22 TransCom-3 regions (see Table 1 in the text and Gurney et al., 2002). Grey line denotes the difference between the a priori and the “truth”, and the black line is the results for the a posteriori in the control run. Red and green lines denote the departures of the a priori from the “truth” in the experiments with systematic or random observation errors, respectively. For clarity, the a posteriori errors in the control run are given as the vertical solid lines. The vertical dashed line demarcates land ocean flux estimates.

moderate reductions (10–65%) over oceans. Because of the wider observation coverage, the nadir-only cycle has better performance over northern high latitudes than the other duty cycles. However, glint-only measurements lead to slightly larger error reductions over the terrestrial tropics, although nadir measurements theoretically represent better constraints for terrestrial sources and sinks by sampling overhead. As mentioned previously, we generally find that tropical land masses are typically characterized by extensive and persistent cloud cover during the wet season and by smoke aerosol during the dry season so the observation density of nadir measurements is low. High-precision glint measurements, sampling continental outflow over the oceans, provide important constraints for estimating land flux estimates.

Nadir measurements provide little constraint on ocean  $\text{CO}_2$  flux estimates, as expected. Glint measurements lead to significant reductions of flux errors over the oceans, reaching 40–60% over the tropics. The 16-day nadir/glint switch leads to a moderate performance between the glint-only and nadir-only duty cycles.



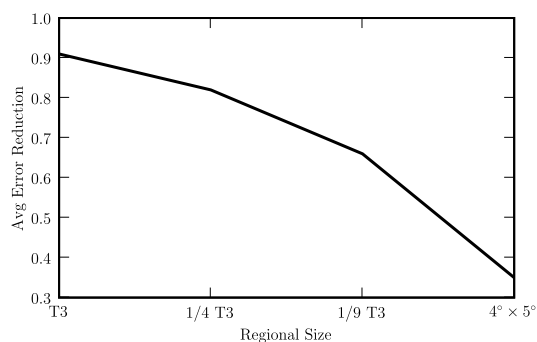
**Fig. 7.** Regional a posteriori error reduction,  $\gamma = 1 - \sigma^a / \sigma^f$ , associated with  $\text{CO}_2$  flux estimation using OCO  $X_{\text{CO}_2}$  observations over one 32-day duty cycle (17 January–17 February 2003). Results have been aggregated from 144 regions (Fig. 3) to the 22 TransCom-3 regions (see Table 1 in the text and Gurney et al., 2002). The vertical solid line demarcates land and ocean flux estimates. Significant error reduction ( $\gamma > 0.5$ ) is marked by the horizontal solid line. **(a)** Squares denote results from the control run, circles denote results from using only nadir measurements, and triangles denote results from using only glint measurements; **(b)** circles denote results from using 80% of available measurements, and triangles denote results from including spatial correlations in the measurement error covariance  $\mathbf{R}$  with an e-folding length scale of 300 km.

#### 4.4 Sensitivity to observation density and correlation

Figure 7b shows that because of the high observation density, reducing the clear observation number by 20% only slightly increases the uncertainties of the estimated fluxes over 144 regions.

OCO observations are made during daylight, and two consecutive orbits are separated by about  $24^\circ$  in longitude. To investigate the impact of measurement correlations, we assume a distance-dependent spatial correlation between observations from the same satellite orbits so that the off-diagonal term  $R(m_1, m_2)$  for two measurements  $m_1$  and  $m_2$  in one orbit is given as

$$R(m_1, m_2) = \sqrt{R(m_1, m_1) R(m_2, m_2)} \exp(-l(m_1, m_2)/l_{\text{cor}}), \quad (16)$$



**Fig. 8.** The sensitivity of error reduction,  $\gamma = 1 - \sigma^a / \sigma^f$ , associated with CO<sub>2</sub> flux estimation using OCO X<sub>CO<sub>2</sub></sub> observations over one 32-day duty cycle (17 January–17 February 2003), to changes in the spatial resolution of the state vector. T3 denotes TransCom-3 regions that are approximately 9 500 000 km<sup>2</sup>.

where  $l_{\text{cor}} = 300$  km is the characteristic spatial correlation length scale, and  $l(m_1, m_2)$  is the distance between the two measurements  $m_1$  and  $m_2$ . Here we assume that these correlations between successive X<sub>CO<sub>2</sub></sub> arise from both the model and observation errors. Figure 7b shows that imposing a spatial correlation weakens the measurement constraint on flux estimations, as expected. We find the largest impacts from including observation correlations are over the oceans where there is a greater density of cloud and aerosol-free measurements, in agreement with Chevallier (2007b). Successive clear measurements over most land regions are sparse and consequently strong correlations are rare. The associated smaller reduction in error reflects a weaker but possibly more realistic measurement constraint than used in the control run, but does not suggest that it is a beneficial practice to ignore the existing observation correlations in data assimilation (Stewart et al., 2008).

#### 4.5 Sensitivity of state vector resolution

To investigate the sensitivity of our results to the spatial resolution of the state vector, we estimate 8-day surface fluxes over the South American tropical region (Table 1) during 17 January to 17 February 2003 at 4 different spatial resolutions: 1) the standard T3 region (Table 1) (c. 9 500 000 km<sup>2</sup>); 2) the 1/4 T3 region (c. 2 300 000 km<sup>2</sup>); 3) the 1/9 T3 region (c. 1 100 000 km<sup>2</sup>); and 4) the 4° × 5° grid box resolution (c. 220 000 km<sup>2</sup>). To reduce the computational costs, we represent the rest of world using the other standard T3 regions (plus one low emission region, see Sect. 3).

Figure 8 shows that the mean error reduction (i.e., the averaged error reduction over tropical South America) decreases rapidly as the the spatial resolution increases. We also find strong negative error correlations for neighbouring grid box in the inversions at the 4° × 5° resolution (not shown). We obtain similar results over other T3 land regions. Over most land regions, there are only 1–2 clear observations within

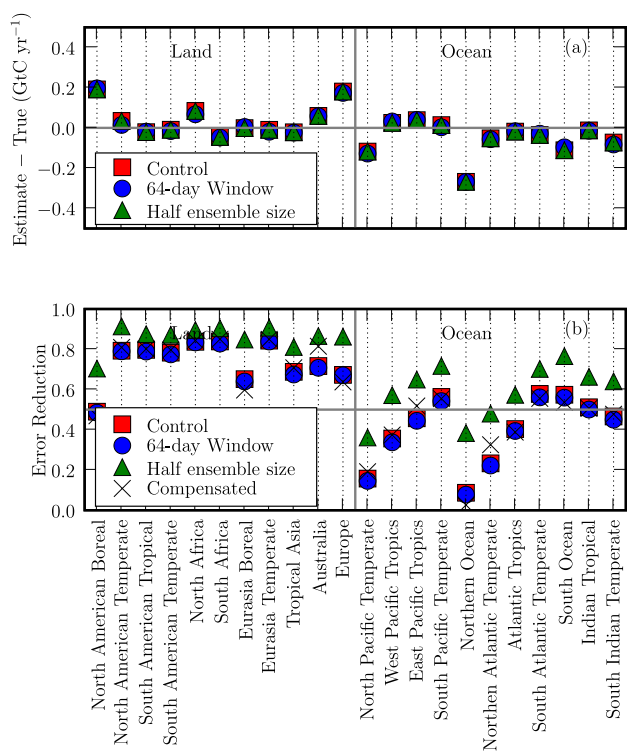
each 2° × 2.5° grid box for a 16-day period (Fig. 2). After we account for model transport and representation errors (assumed to be 2.5 ppmv over lands for the OCO instrument with a field of view < 25 km<sup>2</sup>), our simulated X<sub>CO<sub>2</sub></sub> measurements alone will be insufficient to determine the 8-day mean fluxes on spatial scales 4° × 5°. We acknowledge that by using a relatively coarse resolution model to simulate OCO data, we may over or under-estimate the number of scenes obscured by clouds or aerosols. The distribution of measurements also assumes that only a certain percentage of the available measurements are processed using the computationally expensive full-physics retrievals (Bösch et al., 2009); in practice, OCO data will also be processed using a highly-parameterised retrieval scheme, resulting in a larger number of X<sub>CO<sub>2</sub></sub> measurements. We anticipate that using more dense observations (once their spatial correlations are quantified), we should be able to independently estimate fluxes on smaller scales.

#### 4.6 Sensitivity to lag window and ensemble size

Figure 9a compares the estimated fluxes in the control experiment with the results for the inversions with a shorter lag window (8 × 8 days vs. 12 × 8), or with a smaller ensemble size (6 × 144 vs. 12 × 144, see Appendix B). The differences in the resulting fluxes are typically smaller than 0.05 GtC yr<sup>-1</sup>. The good agreement is partially because our approach uses a full-rank representation for the error covariance of the current forecasts, while approximating the error covariances of the surface flux estimates for the past 8-day time periods, which have been constrained by observations in the previous one assimilation cycles.

Figure 9b shows that the error reductions associated with a shorter lag window (8 × 8 days) are close to those of the control run, indicating our results are insensitive to significantly reducing the original length of the lag time window.

When the size of the a priori ensemble is halved from 12 × 144 to 6 × 144, the error reduction is generally overestimated (i.e., the posterior errors are underestimated), due to the loss of the posterior variations by progressively removing perturbation states associated with small flux uncertainties. Covariance localisation and inflation are commonly used to deal with under-sampling and underestimation of error variances (e.g., Hamill et al., 2001; Ehrendorfer, 2007). Here, we have developed a new simple approach to compensate for this underestimation by accumulating the uncertainties represented by those removed perturbation states back into the final calculation of the a posteriori error covariance (see Appendix B). It requires almost no extra computational cost, but improves the agreement of the a posteriori errors with the full-rank approach (Fig. 9b).



**Fig. 9.** (a) CO<sub>2</sub> flux errors (GtC yr<sup>-1</sup>) over one duty cycle from 17 January to 17 February 2003. The results have been aggregated from 144 regions (Fig. 3) to the 22 TransCom-3 regions (see Table 1 in the text and Gurney et al., 2002). The squares denote the results for the control experiment, the circles denote the experiment using a shorter (8×8 days vs. 12×8 days) lag window, and the triangles denote the experiment using half the ensemble size (6×144 vs. 12×144). (b) A posteriori error reduction,  $\gamma = 1 - \sigma^a / \sigma^f$ , associated with regional fluxes shown in (a). Crosses denote the error reductions when we have compensated for the underestimated posterior uncertainties due to the reduced-rank representation of the EnKF.

## 5 Conclusions

We developed an ensemble Kalman Filter (EnKF) to estimate 8-day regional surface fluxes of CO<sub>2</sub> from space-borne CO<sub>2</sub> dry-air mole fraction observations ( $X_{\text{CO}_2}$ ) and evaluated the approach using a series of synthetic experiments, in preparation for data from the NASA Orbiting Carbon Observatory (OCO). The 32-day duty cycle of OCO alternates between nadir and glint (specular reflection) measurements of backscattered solar radiation at short-wave infrared wavelengths. Our EnKF represents a complementary approach to the variational techniques that have already been developed for interpreting the space-borne  $X_{\text{CO}_2}$  data (e.g., Chevallier et al., 2007a). The main advantages of the EnKF is that it does not require an adjoint model for the forecast and observation operators, and provides a direct estimation of the uncertainty of a posteriori fluxes. We use the ensemble

transform Kalman Filter algorithm to determine the ensemble analysis and its error covariance.

For this work, we estimate 8-day CO<sub>2</sub> surface fluxes over 144 geographical regions (corresponding to 1 100 000 km<sup>2</sup> over land), based on the TransCom-3 experiments (Gurney et al., 2002). We use a 12×8-day lag window, taking into account that  $X_{\text{CO}_2}$  measurements include surface flux information from prior time windows. The observation operator relates surface CO<sub>2</sub> fluxes to the global distributions of the “observed”  $X_{\text{CO}_2}$ . First, we use the GEOS-Chem transport model to relate surface fluxes to global 3-D distributions of CO<sub>2</sub> concentrations. Second, these distributions are sampled at the time and location of OCO measurements to remove cloudy scenes and scenes with aerosol optical depth (AOD) > 0.3. Finally, we use scene-dependent averaging kernels to relate the CO<sub>2</sub> profiles to  $X_{\text{CO}_2}$ . We use the scene-dependent measurement errors that correspond to the averaging kernels. These scene-dependent calculations provide us with the most realistic simulation of  $X_{\text{CO}_2}$  distributions to date, with which to understand potential of OCO to estimate surface CO<sub>2</sub> fluxes. We use the same observation operator to model atmospheric distributions of  $X_{\text{CO}_2}$ , but with an 80% bias in the prior surface emissions.

We show that OCO  $X_{\text{CO}_2}$  measurements significantly reduce the uncertainties of surface CO<sub>2</sub> flux estimates, consistent with previous studies (Baker et al., 2006; Chevallier et al., 2007a; Chevallier, 2007b). We find that nadir measurements are better at estimating land-based fluxes and glint measurements are generally better at constraining ocean fluxes. Nadir  $X_{\text{CO}_2}$  measurements over the terrestrial tropics are typically sparse throughout the year because of either widespread and persistent cloud cover during the wet season or smoke aerosol associated with extensive biomass burning during the dry season. We find that glint measurements over the oceans provide the most effective constraint for estimating terrestrial CO<sub>2</sub> fluxes by accurately sampling fresh continental outflow over neighbouring oceans.

We also presented the results from sensitivity experiments to investigate how flux estimates change with 1) bias and unbiased errors, 2) alternative duty cycles, 3) measurement density and correlations, 4) the spatial resolution of estimated flux estimates, and 5) reducing the length of the lag window and the size of the ensemble. We find that biases in the observations, which we introduce by scaling the error using AOD (by a factor of two), cause large perturbations to some of the posterior fluxes but they are still within the posterior uncertainties of the control experiment. In real observations, there may be larger systematic errors than we discuss here, and their effects will require further investigation. We find that either the current 32-day duty cycle (alternating 16-day cycle between glint and nadir measurements) or one that uses only glint view measurements will address the primary science objectives of the OCO mission, a reflection of the importance of glint measurements in constraining tropical terrestrial fluxes. A modest 20% reduction in the number of

available clear observations does not affect a posteriori regional flux estimates, reflecting the high measurement density. Introducing a spatial correlation between successive measurements effectively reduces the number of independent observations. We find that spatial correlations mainly affect glint measurements over the oceans where there is a greater number of neighbouring scenes that are cloud-free and have AODs  $<0.3$ . We find that reducing the size of the geographical regions over which to estimate surface fluxes much below 1 million km<sup>2</sup> introduces large correlations between neighbouring regional estimates. In the control experiment, we simultaneously estimate surface fluxes at the time of the assimilation and at times up to 3 months prior. We find that surface flux estimates for a particular 8-day period typically converge after ingesting 4–6 weeks of data. To improve the speed of the EnKF we halved the number of ensemble states used to determine the a priori error covariance and showed that the flux estimates were close to the control experiment but using a reduced number of ensemble states, we generally underestimated the associated error. Constructing efficient reduced-rank representations of the EnKF, and the methods to compensate for associated underestimation of the posterior uncertainties, necessary to reduce computational costs related to estimating fluxes on finer spatial resolutions, is the subject of ongoing work.

In light of the failed OCO launch (Palmer and Rayner, 2009), we will focus our developed EnKF on GOSAT measurements of CO<sub>2</sub> and CH<sub>4</sub>. This will require information about the GOSAT orbit, sampling strategy, and retrieval error diagnostics that will soon become available. We anticipate that application of the EnKF to GOSAT data will differ only slightly to the application to OCO data shown in this paper, e.g., length of the assimilation window, which will be the subject of further work.

## Appendix A

### Description of the GEOS-Chem Model of Atmospheric CO<sub>2</sub>

We use the GEOS-Chem global 3-D chemistry transport model (v7-03-06) to calculate X<sub>CO<sub>2</sub></sub> concentrations from prescribed surface CO<sub>2</sub> fluxes described below. We used the model with a horizontal resolution of 2° × 2.5°, and 30 vertical levels (derived from the native 48 levels) ranging from the surface to the mesosphere, 20 of which are below 12 km. The model is driven by GEOS-4 assimilated meteorology data from the Global Modeling and Assimilation Office Global Circulation Model based at NASA Goddard. The 3-D meteorological data is updated every six hours, and the mixing depths and surface fields are updated every three hours. The CO<sub>2</sub> simulation is based on Suntharalingam et al. (2005) and Palmer et al. (2006, 2008).

We use gridded fossil fuel emission distributions, representative of 1995 (Suntharalingam et al., 2005), which we have scaled to 2003 values using regional budget estimates for the top 20 emitting countries in 2003 from the Carbon Dioxide Information Analysis Center (Marland et al., 2007). Biofuel emission estimates are taken from Yevich and Logan (2003) and represent climatological values. Monthly mean biomass burning emission estimates are taken from the second version of the Global Fire Emission Database (GFEDv2) for 2003 (van der Werf et al., 2006), which are derived from ground-based and satellite observations. Daily mean land biosphere fluxes are taken from the CASA model for 2001 (Randerson et al., 1997), in the absence of corresponding fluxes for 2003. We do not explicitly account for the contribution of fuel combustion CO<sub>2</sub> from the oxidation of reduced carbon species (Suntharalingam et al., 2005) as they make only a small contribution to the CO<sub>2</sub> column. Monthly mean air-sea fluxes of CO<sub>2</sub> are taken from (Takahashi et al., 1999).

CO<sub>2</sub> concentrations for January 2002 were initialized from a previously evaluated model run (Palmer et al., 2006), which we integrate forward to January 2003. We include an additional initialization to correction for the model bias introduced by not accounting for the net uptake of CO<sub>2</sub> from the terrestrial biosphere. We make this downward correct by comparing the difference between GLOBALVIEW CO<sub>2</sub> data (GLOBALVIEW-CO<sub>2</sub>) and model concentrations over the Pacific during January 2003. Differences range from 1 to 4 ppmv with a median of 3.5 ppmv, and we subtract this value globally, following Suntharalingam et al. (2005). From January 2003 the total CO<sub>2</sub> tracer becomes the “background” CO<sub>2</sub> concentration and is only subject to atmospheric transport. At that time, we also introduce additional model tracers, initialized with a uniform value (for numerical reasons and which is subtracted in subsequent analyses), that account for the monthly production and loss of CO<sub>2</sub> originating from specific geographical regions and surface processes (“tagged” tracers). The linear sum of these monthly tagged tracers (and the “background”) is equivalent to the total CO<sub>2</sub>.

## Appendix B

### Description of the reduced representation of a priori error covariance

To reduce the computational costs of our EnKF approach, we use a reduced-rank representation to approximate the a priori error covariances, so that fewer ensemble states need to be projected to the observation space using the observation operator  $H$  that includes a global 3-D transport model of CO<sub>2</sub>.

As mentioned in Sect. 3, during the  $j$ -th cycle of assimilating X<sub>CO<sub>2</sub></sub> observations from day  $d$  to  $d+8$ , our state vector consists of 1) the current forecast of the regional surface CO<sub>2</sub>



fluxes from day  $d$  to  $d+8$ , and 2) the 8-day mean regional flux estimates from time periods prior to day  $d$ .

We assume no error correlation between the current forecast and the previous analysis  $\mathbf{x}_{j-1}^a$ , and hence the prior error covariance at the assimilation cycle  $j$  consists of two blocks: 1) the error covariance of the current forecast; and 2) the error covariance of  $\mathbf{x}_{j-1}^a$ .

The error covariance of the current forecast is a diagonal matrix of size  $N_r \times N_r$  (Sect. 3), where  $N_r$  is the number of the global regions ( $N_r=144$  in the control run). In the reduced-rank approach discussed here, we still use  $N_r$  new perturbation states  $\Delta\mathbf{X}_j^{P1}$  to represent this diagonal error covariance matrix.

But we only choose a subset of the a posteriori ensemble of the previous cycle to approximate the error covariance of the estimates over the past time periods in the steps described below.

For clarity, we start by calculating the a posteriori ensemble at the end of the previous assimilation cycle  $j-1$  via Eq. (8), where the subscript denotes the assimilation cycle number:

$$\Delta\mathbf{X}_{j-1}^a = \Delta\mathbf{X}_{j-1}^f \mathbf{T}_{j-1}. \quad (\text{B1})$$

Matrix  $\Delta\mathbf{X}_{j-1}^a$  is the same size of matrix  $\Delta\mathbf{X}_{j-1}^f$ , which consists of  $N_e$  columns, each with  $N_x$  elements representing perturbations in the regional 8-day mean surface CO<sub>2</sub> fluxes prior to the current day  $d$ .

We then use SVD to decompose  $\Delta\mathbf{X}_{j-1}^a$ :

$$\Delta\mathbf{X}_{j-1}^a = \mathbf{U}_{j-1}^a \Sigma_{j-1}^a (\mathbf{V}_{j-1}^a)^T, \quad (\text{B2})$$

where  $\mathbf{U}_{j-1}^a$ , and  $\mathbf{V}_{j-1}^a$  are two orthogonal matrices of size  $N_x \times N_x$  and  $N_e \times N_e$ , respectively, and  $\Sigma_{j-1}^a$  is a diagonal matrix of size  $N_x \times N_e$ , with its non-zero diagonal elements presenting the singular values of matrix  $\Delta\mathbf{X}_{j-1}^a$  in descending order of magnitude.

By applying  $\mathbf{V}_{j-1}^a$  to Eq. (B2), we obtain

$$\Delta\mathbf{X}_j^c = \Delta\mathbf{X}_{j-1}^a \mathbf{V}_{j-1} = \mathbf{U}_{j-1}^a \Sigma_{j-1}^a. \quad (\text{B3})$$

Matrix  $\Delta\mathbf{X}_j^c$  has the same size as  $\Delta\mathbf{X}_{j-1}^a$ , and satisfies

$$\mathbf{P}_{j-1}^a = \Delta\mathbf{X}_{j-1}^a (\Delta\mathbf{X}_{j-1}^a)^T = \Delta\mathbf{X}_{j-1}^c (\Delta\mathbf{X}_{j-1}^c)^T \quad (\text{B4})$$

We divide the ensemble  $\Delta\mathbf{X}_j^c = [\Delta\mathbf{x}_1^c, \Delta\mathbf{x}_2^c, \dots, \Delta\mathbf{x}_{N_e}^c]$  into two subsets: 1) the major subset  $\Delta\mathbf{X}_j^{P2}$  that consists of its first  $N_b$  columns; and 2) the minor subset  $\Delta\mathbf{X}_j^s$  for its last  $N_e - N_b$  columns. The two subsets together satisfy

$$\mathbf{P}_{j-1}^a = \Delta\mathbf{X}_j^{P2} (\Delta\mathbf{X}_j^{P2})^T + \Delta\mathbf{X}_j^s (\Delta\mathbf{X}_j^s)^T. \quad (\text{B5})$$

The major subset  $\Delta\mathbf{X}_j^b$  contains the columns with the largest amplitudes, and in principle, a suitable choice of  $N_b$  ensures that  $\mathbf{P}_j^{P2} = \Delta\mathbf{X}_j^{P2} (\Delta\mathbf{X}_j^{P2})^T$  provides a good approximation of the error covariance  $\mathbf{P}_{j-1}^a$ , while the uncertainties represented by  $\mathbf{P}_j^s = \Delta\mathbf{X}_j^s (\Delta\mathbf{X}_j^s)^T$  are small enough to be ignored.

In this study, we choose  $N_b$  to be equal to  $N_e - N_r$  (in the spin-up period where the lag window is shorter than  $12 \times 8$  days,  $N_b$  is chosen to increase with the size of the state vector till it reaches the pre-defined ensemble size  $N_e$ ). Combining the resulting ensemble  $\Delta\mathbf{X}_j^{P2}$  and  $\Delta\mathbf{X}_j^{P1}$  provides a reduced-rank representation of the prior uncertainties of the current assimilation cycle  $j$ .

During the previous assimilation cycle, the GEOS-Chem model run, forced by its prior surface flux estimates, also generates an estimate of the 3-D CO<sub>2</sub> concentrations  $\mathbf{c}_{j-1}^f$  at the beginning of day  $d$ . Similar simulations for the a priori ensemble  $\Delta\mathbf{X}_{j-1}^f$  provide an ensemble of 3-D CO<sub>2</sub> concentrations  $\Delta\mathbf{C}_{j-1}^f$  for the variations caused by the perturbations in the surface CO<sub>2</sub> fluxes. We calculate the ‘‘analysis’’ of the 3-D CO<sub>2</sub> concentrations from  $\mathbf{c}_{j-1}^f$  and  $\Delta\mathbf{C}_{j-1}^f$  using an equation similar to Eq. (2), and generate the variations corresponding to  $\Delta\mathbf{X}_j^{P2}$  by selecting the first  $N_b$  fields from matrix  $\Delta\mathbf{C}_j^c = \Delta\mathbf{C}_{j-1}^f \mathbf{T}_{j-1} \mathbf{V}_{j-1}$ .

In the current assimilation cycle over day  $d$  to day  $d+8$ , we use the GEOS-Chem transport model to propagate these resulting CO<sub>2</sub> fields to the current observation space to account the contributions from the surface CO<sub>2</sub> fluxes prior to day  $d$ .

The reduced-rank representation tends to underestimate the a posteriori error covariance. To compensate for this underestimation, we accumulate the previously ignored small uncertainties  $\mathbf{P}_j^s$  into the calculation of the final a posteriori error covariance at the time the regional fluxes estimates over day  $d$  to day  $d+8$  leave the lag window of  $12 \times 8$  days, after having been constrained by observations in 12 consecutive assimilation cycles. This compensation is consistent with the assumption that these previously ignored uncertainties are too small to be significantly reduced by ingesting observations.

*Acknowledgements.* This study is funded by the UK Natural Environment Research Council under NE/F000014/1. Work at the Jet Propulsion Laboratory (JPL), California Institute of Technology was carried out under a contract with the National Aeronautics and Space Administration.

Edited by: W. Lahoz

## References

- Atlas, R.: Atmospheric observations and experiments to assess their usefulness in data assimilation, *J. Meteorol. Soc. Japan*, 75 (1B), 111–130, 1997.
- Baker, D. F., Doney, S. C., and Schimel, D. S.: Variational data assimilation for atmospheric CO<sub>2</sub>, *Tellus Ser. B*, 58, 359–365, 2006.
- Barkley, M. P., Monks, P. S., Frieß, U., Mittermeier, R. L., Fast, H., Körner, S., and Heimann, M.: Comparisons between SCIAMACHY atmospheric CO<sub>2</sub> retrieved using (FSI) WFM-DOAS to ground based FTIR data and the TM3 chemistry transport model,

- Atmos. Chem. Phys., 6, 4483–4498, 2006,  
<http://www.atmos-chem-phys.net/6/4483/2006/>.
- Barkley, M. P., Monks, P. S., Hewitt, A. J., Machida, T., Desai, A., Vinnichenko, N., Nakazawa, T., Yu Arshinov, M., Fedoseev, N., and Watai, T.: Assessing the near surface sensitivity of SCIAMACHY atmospheric CO<sub>2</sub> retrieved using (FSI) WFM-DOAS, Atmos. Chem. Phys., 7, 3597–3619, 2007,  
<http://www.atmos-chem-phys.net/7/3597/2007/>.
- Bishop, C. H., Etherton, B. J., and Majumdar, S. J.: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects, Mon. Weather Rev., 129, 420–436, 2001.
- Bösch, H., Baker, D., Connor, B., O'Brien, D., Crisp, D., and Miller, C.: Global Characterization of XCO<sub>2</sub> Retrievals from OCO Observations, in preparation, 2009.
- Bousquet, P., Peylin, P., Ciais, P., Quere, C. L., Friedlingstein, P., and Tans, P. P.: Regional changes in carbon dioxide fluxes of land and oceans since 1980, Science, 290, 1342–1346, 2000.
- Bovensmann, H., Burrows, J., Buchwitz, M., Frerick, J., Noël, S., Rozanov, V., Chance, K. V., and Goede, A.: SCIAMACHY: Mission objectives and measurement modes, J. Atmos. Sci., 56, 127–150, 1999.
- Bruhwyler, 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, 2005,  
<http://www.atmos-chem-phys.net/5/2691/2005/>.
- Chevallier, F.: Impact of correlated observation errors on inverted CO<sub>2</sub> surface fluxes from OCO measurements, Geophys. Res. Lett., 34, L24804, doi:10.1029/2007GL030463, 2007b.
- Chevallier, F., Bréon, F.-M., and Rayner, P. J.: Contribution of the Orbiting Carbon Observatory to the estimation of CO<sub>2</sub> sources and sinks: Theoretical study in a variational data assimilation framework, J. Geophys. Res., 112, D09307, doi:10.1029/2006JD007375, 2007a.
- Chevallier, F. M. F., Peylin, P., Bousquet, S. S. P., Bréon, F.-M., Chédin, A., and Ciais, P.: Inferring CO<sub>2</sub> sources and sinks from satellite observations: Method and application to TOVS data, J. Geophys. Res., 110, D24309, doi:10.1029/2005JD006390, 2005.
- Connor, B. J., Bösch, H., Toon, G., Sen, B., Miller, C., and Crisp, D.: Orbiting Carbon Observatory: Inverse method and prospective error analysis, J. Geophys. Res., 113, D05305, doi:10.1029/2006JD008336, 2008.
- Crisp, D., Atlas, R. M., Bréon, F.-M., et al.: The Orbiting Carbon Observatory (OCO) Mission, Adv. Space. Res., 34(4), 700–709, 2004.
- Ehrendorfer, M.: A review of issues in ensemble-based Kalman filtering, Meteorol. Z., 16, 795–818, 2007.
- Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99(C5), 10143–10162, 1994.
- Evensen, G.: The Ensemble Kalman Filter: Theoretical formulation and practical implementation, Ocean Dynam., 53, 343–367, 2003.
- GLOBALVIEW-CO<sub>2</sub>: Cooperative Atmospheric Data Project Carbon Dioxide, CD-ROM, NOAA GMD, Boulder, Colorado, USA (also available via anonymous FTP to ftp.cmdl.noaa.gov, path://ccg/co2/GLOBALVIEW, 2006.
- Gurney, K. R., Law, R. L., Denning, A. S., et al.: Towards robust regional estimates of CO<sub>2</sub> sources and sinks using atmospheric transport models, Nature, 415, 626–630, 2002.
- Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter, Mon. Weather Rev., 129, 2776–2790, 2001.
- Houtekamer, P. L. and Mitchell, H. L.: Data assimilation using an ensemble Kalman filter technique, Mon. Weather Rev., 126, 796–811, 1998.
- Houweling, S., Kaminski, T., Dentener, F., Lelieveld, J., and Heimann, M.: Inverse modeling of methane sources and sinks using the adjoint of a global transport model, J. Geophys. Res., 104, 26137–26160, 1999.
- Houweling, S., Hartmann, W., Aben, I., Schrijver, H., Skidmore, J., Roelofs, G.-J., and Breon, F.-M.: Evidence of systematic errors in SCIAMACHY-observed CO<sub>2</sub> due to aerosols, Atmos. Chem. Phys., 5, 3003–3013, 2005,  
<http://www.atmos-chem-phys.net/5/3003/2005/>.
- Lahoz, W. A., Brugge, R., Jackson, D. R., Migliorini, S., Swinbank, R., Lary, D., and Lee, A.: An observing system simulation experiment to evaluate the scientific merit of wind and ozone measurements from the future SWIFT instrument, Q. J. Roy. Meteor. Soc., 131, 503–523, doi:10.1256/qj.03.109, 2005.
- Law, R. M., Chen, Y. H., and Gurney, K. R.: TransCom-3 CO<sub>2</sub> inversion intercomparison: 2. Sensitivity of annual mean results to data choices, Tellus, Ser. B, 55, 580–595, 2003.
- Livingston, D. M.: Aspects of the Ensemble Kalman Filter, M.Sc. thesis, Department of Mathematics, University of Reading, UK, 34–37, 2005.
- Livingston, D. M., Dance, S. L., and Nichols, N. K.: Unbiased ensemble square root filters, Physica D., 237/8, 1021–1028, 2008.
- Lorenc, A. C.: The potential of the ensemble Kalman filter for NWP—A comparison with 4D-Var, Q. J. Roy. Meteor. Soc., 129, 3183–3203, 2003.
- Maksyutov, S., Kadyrov, N., Nakatsuka, Y., Patra, P. K., Nakazawa, T., Yokota, T., and Inoue, G.: Projected impact of the GOSAT observations on regional CO<sub>2</sub> flux estimations as a function of total retrieval error, Journal of Remote Sensing Society of Japan, 28, 190–197, 2008.
- Marland, G., Boden, T. A., and Andres, R. J.: Global, Regional, And National CO<sub>2</sub> Emissions, in Trends: A Compendium of Data on Global Change, Tech. Rep. 2007, 7346, Carbon Dioxide Information Analysis Center Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tenn., USA, 2007.
- Miller, C. E., Crisp, D., DeCola, P. L., et al.: Precision requirements for space-based XCO<sub>2</sub> data, J. Geophys. Res., D10314, doi:10.1029/2006JD007659, 2007.
- Palmer, P. I., Suntharalingam, P., Jones, D. B. A., Jacob, D. J., Streets, D. G., Fu, Q., Vay, S. A., and Sachse, G. W.: Using CO<sub>2</sub>:CO correlations to improve inverse analyses of carbon fluxes, J. Geophys. Res., 111, D12318, doi:10.1029/2005JD006697, 2006.
- Palmer, P. I., Barkley, M. P., and Monks, P. S.: Interpreting the variability of space-borne CO<sub>2</sub> column-averaged volume mixing ratios over North America using a chemistry transport model, Atmos. Chem. Phys., 8, 5855–5868, 2008,  
<http://www.atmos-chem-phys.net/8/5855/2008/>.
- Palmer, P. I. and Rayner, P.: Failure to launch, Nature Geosci., 2, 247, doi:10.1038/ngeo495, 2009.
- Patra, P. K., Maksyutov, S., Sasano, Y., Nakajima, H., Inoue, G., and Nakazawa, T.: An evaluation of CO<sub>2</sub> observations with So-

- lar Occultation FTS for Inclined-Orbit Satellite sensor for surface source inversion, *J. Geophys. Res.*, 108(D24), 4759, doi:10.1029/2003JD003661, 2003.
- 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<sub>2</sub> surface fluxes from atmospheric trace gas observations, *J. Geophys. Res.*, 110, D24304, doi:10.1029/2005JD006157, 2005.
- Peylin, P., Baker, D., Sarmiento, J., Ciais, P., and Bousquet, P.: Influence of transport uncertainty on annual mean and seasonal inversions of atmospheric CO<sub>2</sub> data, *J. Geophys. Res.*, 107(D19), 4385, doi:10.1029/2001JD000857, 2002.
- Potter, C. S., Randerson, J. T., Field, C. B., Matson, P. A., Vitousek, P. M., Mooney, H. A., and Klooster, S. A.: Terrestrial ecosystem production: A process model based on global satellite and surface data, *Global Biogeochem. Cy.*, 7(4), 811–842, 1993.
- Randerson, J. T., Thompson, M. V., Conway, T. J., Fung, I. Y., and Field, C. B.: The contribution of terrestrial sources and sinks to trends in the seasonal cycle of atmospheric carbon dioxide, *Global Biogeochem. Cy.*, 11(4), 535–560, 1997.
- Rayner, P. J., Law, R. M., O'Brien, D. M., Butler, T. M., and Dilley, A. C.: Global observations of the carbon budget: 3. Initial assessment of the impact of satellite orbit, scan geometry, and cloud on measuring CO<sub>2</sub> from space, *J. Geophys. Res.*, 107(D21), 4557, doi:10.1029/2001JD000618, 2002.
- Rödenbeck, C., Houweling, S., Gloor, M., and Heimann, M.: CO<sub>2</sub> flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport, *Atmos. Chem. Phys.*, 3, 1919–1964, 2003, <http://www.atmos-chem-phys.net/3/1919/2003/>.
- Schneising, O., Buchwitz, M., Burrows, J. P., Bovensmann, H., Reuter, M., Notholt, J., Macatangay, R., and Warneke, T.: Three years of greenhouse gas column-averaged dry air mole fractions retrieved from satellite – Part 1: Carbon dioxide, *Atmos. Chem. Phys.*, 8, 3827–3853, 2008, <http://www.atmos-chem-phys.net/8/3827/2008/>.
- Stewart, L. M., Dance, S., and Nichols, N.: Information content and correlated observation errors, *Int. J. Numer. Meth. Fl.*, 56, 1521–1527, 2008.
- Suntharalingam, P., Randerson, J. T., Krakauer, N., Logan, J. A., and Jacob, D. J.: The influence of reduced carbon emissions and oxidation on the distribution of atmospheric CO<sub>2</sub>: implications for inversion analysis, *Global Biogeochem. Cy.*, 19, GB4003, doi:10.1029/2005GB002466, 2005.
- Takahashi, T., Wanninkhof, R. T., Feely, R. A., Weiss, R. F., Chapman, D. W., Bates, N. R., Olafsson, J., Sabine, C. L., and Sutherland, C. S.: Net sea-air CO<sub>2</sub> flux over the global oceans, proceedings of the 2nd international symposium CO<sub>2</sub> in the oceans: CGER 1037, National Institute for Environmental Studies, Tsukuba, Japan, 915 pp., 1999.
- Takahashi, T., Sutherland, S. C., Sweeney, C., et al.: Global sea-air CO<sub>2</sub> flux based on climatological surface ocean pCO<sub>2</sub>, and seasonal biological and temperature effects, *Deep Sea Res., Part II*, 49, 1601–1622, 2002.
- Tiwari, Y. K., Gloor, M., Engelen, R. J., Chevallier, F., Rödenbeck, C., Körner, S., Peylin, P., Braswell, B. H., and Heimann, M.: Comparing CO<sub>2</sub> retrieved from Atmospheric Infrared Sounder with model predictions: Implications for constraining surface fluxes and lower-to-upper troposphere transport, *J. Geophys. Res.*, 111, D17106, doi:10.1029/2005JD006681, 2006.
- 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, 2006, <http://www.atmos-chem-phys.net/6/3423/2006/>.
- Wang, X., Bishop, C. H., and Julier, S. J.: Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble, *Mon. Weather Rev.*, 132, 1590–1605, 2004.
- Yevich, R. and Logan, J. A.: An assessment of biofuel use and burning of agricultural waste in the developing world, *Global Biogeochem. Cy.*, 17, 1095, doi:10.1029/2002GB001952, 2003.
- Zupanski, M.: Maximum likelihood Ensemble Filter: Theoretical Aspects, *Mon. Weather Rev.*, 133, 1710–1726, 2005.